

Kenya - Mbita HDSS INDEPTH Core Dataset 2009-2015 (Release 2018)

**Prof. Satoshi Kaneko - Nagasaki University Institute of Tropical Medicine - Kenya
Medical Research Institute**

**Prof Mohamed Karama - Nagasaki University Institute of Tropical Medicine - Kenya
Medica**

Report generated on: June 30, 2018

Visit our data catalog at:

<http://indepth-ishare.org.md-in-71.hostgatorwebservers.com/index.php>

Overview

Identification

ID NUMBER
INDEPTH.KE041.CMD2015.v1

Version

VERSION DESCRIPTION
KE041.CMD2015.v1: Edited dataset for public distribution.

PRODUCTION DATE
2018-06-16

NOTES
Developed during May 2018 iSHARE2 workshop in Dubai and refined later in June 2018

Overview

ABSTRACT

The Mbita Health and Demographic Surveillance System (Mbita HDSS), located on the shores of Lake Victoria in Kenya, was established in 2006. The main objective of the HDSS is to provide a platform for population-based research on relationships between diseases and socio-economic and environmental factors, and for the evaluation of disease control interventions.

The Mbita HDSS had a population of approximately 54 014 inhabitants from 11 576 households in June 2013. Regular data are collected using personal digital assistants (PDAs) every 3 months, which includes births, pregnancies, migration events and deaths. Coordinates are taken using geographical positioning system (GPS) units to map all dwelling units during data collection. Cause of death is inferred from verbal autopsy questionnaires. In addition, other health-related data such as vaccination status, socio-economic status, water sources, acute illness and bed net distribution are collected.

The HDSS has also provided a platform for conducting various other research activities such as entomology studies including malaria research on neglected tropical diseases, and environmental health projects which have benefited the organization as well as the HDSS community residents. Data collected are shared with the community members, health officials, local administration and other relevant organizations. Opportunities for collaboration and data sharing with the wider research community are available and those interested should contact shimadam@nagasaki-u.ac.jp <mailto:shimadam@nagasaki-u.ac.jp> or mhmdkarama@yahoo.com <mailto:mhmdkarama@yahoo.com>.

KIND OF DATA
Event history data

UNITS OF ANALYSIS
Individual

Scope

NOTES
Individual Identifiers, memid

Individuals gender, sex

Individuals dates of birth

Events for inhabitants: migrations, deaths, pregnancies and pregnancy outcomes

Individuals mothers, mother id for children born within the HDSS site

Location id, Houseid

Family identifiers, famid

Observation date, edittime on all individual tables, edittime on death tables, edittime in the migration tables

TOPICS

Topic	Vocabulary	URI
Demography [N01.224]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Rural Population [N01.600.725]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Health [N01.400]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Socioeconomic Factors [N01.824]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Health Facilities [N02.278]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Environment and Public Health [N06]	MeSH	http://www.ncbi.nlm.nih.gov/mesh

KEYWORDS

Migration, Fertility, Mortality

Coverage

GEOGRAPHIC COVERAGE

Mbita HDSS covers part of the western Kenya near the shores of the Lake Victoria including Rusinga Island(North Part) and mainland labeled as Gembe in Gembe location(South Part).

Area coverage is about 168KM square.

The area is accessible by a ferry through the Lake victoria to the mainland of Lwanda Kotieno in Kisumu area of Kisumu County. There is also a tarmac road connection that goes through the mainland from the Homabay County to Mbita HDSS area and it is linked to the man-made bridge that links the the mainland to Rusinga Island.

UNIVERSE

All individuals in the HDSS survey area.

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Prof. Satoshi Kaneko	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute
Prof Mohamed Karama	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute

OTHER PRODUCER(S)

Name	Affiliation	Role
Prof. M. Shimada	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute	Team - Leader
Ms. S. Wanyua	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute	Field manager

Name	Affiliation	Role
Mr. Morris Ndemwa	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute	Research Scientist - Lead Data Manager/Analyst
Mr. Abdulatif Kisule Mohamed	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute	Data Manager

FUNDING

Name	Abbreviation	Role
Government of Japan		Financial Support
Nagasaki University Institute of Tropical Medicine	NUITM	Project Leadership and Technical Support
Kenya Medical Research Institute	KEMRI	Partnership affiliation and Technical Support

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Morris Ndemwa	NUITM	Data Manager
Mbita Community		Partnership and cooperation in data collection
County Leadership	Devolved government system (Homabay county)	Partnership and cooperation in data collection

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical support
INDEPTH Network	int.indepth	INDEPTH	agency
Abdulatif Kisule Mohamed	AKM	Nagasaki University Institute of Tropical Medicine - Kenya Medical Research Institute	DDI Author

DATE OF METADATA PRODUCTION

2018-06-16

DDI DOCUMENT VERSION

KE041.CMD2015.v1

DDI DOCUMENT ID

DDI.INDEPTH.KE041.CMD2015.v1

Sampling

Sampling Procedure

N/A

Deviations from Sample Design

N/A

Response Rate

Response rate is about 95%.

Weighting

N/A

Questionnaires

Overview

The questionnaire is divided into several parts. The 1st part consists of the baseline census dealing mainly with demographic information, such as compounds, household and members. The second part is made up of questions on migrations, health, diseases, causes of death, and other health and hygiene related information.

These processes are repeated to accumulate demographic and health related information for several years.

Data collection in the field uses electronic devices pda.

These devices support field interviewers' data collection activities and execute data consistency checks with regard to questionnaire items on site.

Assumed variables of questionnaire are:

- ∅ information for individual identification
- ∅ demographic information
- ∅ household information
- ∅ migration information
- ∅ pregnancy information
- ∅ health related information
- ∅ vital event information

Collected data is accumulated in the field stations and sent to data management centre in the central office in nairobi, and are subjected to detail-verification check by data managers.

If inconsistent data or unverified data are detected, those data and error codes (or explanation) are sent back to the field.

Verbal Autopsy Questionnaires design was made with some modifications but adapted from the recommended WHO questionnaire

Data Collection

Data Collection Dates

Start	End	Cycle
2009-01-01	2015-12-31	Release Coverage

Time Periods

Start	End	Cycle
2008-10-14		1
2009-01-08		2
2009-02-17		3
2009-03-11		4
2009-04-09		5
2009-05-10		6
2009-06-05		7
2009-07-02		8
2009-07-30		9
2009-08-21		10
2009-09-18		11
2009-10-23		12
2009-12-02		13
2010-01-08		14
2010-03-05		15
2010-04-23		16
2010-06-10		17
2010-07-20		18
2010-09-22		19
2010-11-23		20
2011-01-12		21
2011-03-11		22
2011-06-02		23
2011-09-28		24
2011-11-01		25

Data Collection Mode

Proxy Respondent [proxy]

Data Collection Notes

Training of enumerators (known as filed interviewres in Mibita HDSS site) is routinely done before the start of any new round. This is usually done to familiarize with the system updates on the programme or any additional changes that may seem necessary and necessitate smooth running of the data collection criteria or inform the enumerators of the anychanges that they may have suggested for the smooth running of data collection process.

- Usually interview may take about 30mins to about 45mins depending on whether there were any additional questionnaires added to their routine data collection method.
- Community mobilization and sensitization was first done before data collection at the baseline.
- In addition, more involvement was done to inform community leaders especially the chiefs, leaders and community heads (including policy makers) of the objectives of the HDSS programe.
- The data collection team has usually been conducted by field interviewers (FI) 17 in number guided by immediate supervisors that are named as field interviewer manager (FIM) who are 3 in number.
- Supervisory team comprising of a data manager at the field station and an asdministrator would oversee the whole process and they may seek intervention by community leaders in areas where the community may not seem intrested or refusing to accept enumeration.

- Interview is normally done in local language (Luo) or in Swahili where there find in-migrants.

Questionnaires

The questionnaire is divided into several parts. The 1st part consists of the baseline census dealing mainly with demographic information, such as compounds, household and members. The second part is made up of questions on migrations, health, diseases, causes of death, and other health and hygiene related information.

These processes are repeated to accumulate demographic and health related information for several years.

Data collection in the field uses electronic devices pda.

These devices support field interviewers' data collection activities and execute data consistency checks with regard to questionnaire items on site.

Assumed variables of questionnaire are:

- ∅ information for individual identification
- ∅ demographic information
- ∅ household information
- ∅ migration information
- ∅ pregnancy information
- ∅ health related information
- ∅ vital event information

Collected data is accumulated in the field stations and sent to data management centre in the central office in Nairobi, and are subjected to detail-verification check by data managers.

If inconsistent data or unverified data are detected, those data and error codes (or explanation) are sent back to the field.

Verbal Autopsy Questionnaires design was made with some modifications but adapted from the recommended WHO questionnaire

Data Collectors

Name	Abbreviation	Affiliation
Mbita HDSS	Mbita, HDSS	NUITM, KEMRI

Supervision

Enumerators (field interviewers) were organized in teams per working areas and regions which they are assigned to all interviewers!

All enumerators report to a supervisor (field interviewer manager) who is assigned about 6 enumerators and he/she is able to manage them by using an assigned motorbike to move from one working area to the other.

The field interviewer manager (supervisor) roles include the following: coordinate data collection, supplies of equipment and gadgets, coordinating with local authorities, quality assurance report including checking empty houses/structures, conducting verbal autopsy after the mourning period has passed and following up on skipped houses which may have been skipped by enumerators if for some reasons owners were absent or only minors were available during the visit.

The data manager at the field station was responsible for the transfer of collected data to our remote server for cleaning before synchronization of the data and backup.

The data manager would go through the quality check reports generated automatically by the system and assigns duties to field interviewer manager based on the results for updates and corrections.

Upper management from the (Nairobi) managing the core data set visited the field once every month and after every round had been completed.

Data Processing

Data Editing

Data management is conducted on PDA, Station PC and main server database.

For quality control, several error check programmes were installed, for the following purposes: range checks, entry missing checks, inter-variable checks, consistency check between currently entered data and past data, etc.

Data was left censored to 1 Jan 2009 to account for the start-up phase of the surveillance.

Other Processing

PDA's, the verification programmes were executed on developed programmes in the project.

Station PCs, the verification programmes were executed by Microsoft Access.

For the main server DB, the programmes were run on SQL.

After consistency check between previous corrected data, which are summarized and confirmed data about families, households and their members, the latest correct upto date data were transferred to station PC from the main server DB. The data was overwritten on the PDA data and used in the next round interview.

Although data on station PC and PDA are overwritten on the previous data, data on the main server DB are not overwritten, but appended and stored for future use and verification.

Data Appraisal

Estimates of Sampling Error

N/A

Other forms of Data Appraisal

CentreId MetricTable QMetric Illegal Legal Total Metric RunDate

KE041 MicroDataCleaned Starts 33074 2018-06-16 10:51
KE041 MicroDataCleaned Transitions 0 66148 66148 0. 2018-06-16 10:51
KE041 MicroDataCleaned Ends 33074 2018-06-16 10:51
KE041 MicroDataCleaned SexValues 66148 2018-06-16 10:51
KE041 MicroDataCleaned DoBValues 66148 2018-06-16 10:51

File Description

Variable List

KE041.CMD2015.v1

Content	This file contains KE041 INDEPTH Core microdataset in event history format. This file was generated using ETL through Pentaho Kettle.
Cases	78480
Variable(s)	14
Structure	Type: Keys: ()
Version	KE041.CMD2015.v1
Producer	Mbita HDSS (KE041)
Missing Data	The year 2013 to 2014 we were doing some program upgrade that should explain the drops in mortality and fertility rates as well as peaks for the life expectancy.

Variables

ID	Name	Label	Type	Format	Question
V1	RecNr	RecNr	contin	numeric	
V2	CountryId	CountryId	discrete	numeric	
V3	CentrelId	CentrelId	discrete	character	
V4	IndividualId	IndividualId	contin	numeric	
V5	Sex	Sex	discrete	numeric	
V6	DoB	DoB	discrete	character	
V7	EventCount	EventCount	discrete	numeric	
V8	EventNr	EventNr	discrete	numeric	
V9	EventCode	EventCode	discrete	character	
V10	EventDate	EventDate	discrete	character	
V11	ObservationDate	ObservationDate	discrete	character	
V12	LocationId	LocationId	contin	numeric	
V13	MotherId	MotherId	contin	numeric	
V14	DeliveryId	DeliveryId	contin	numeric	

RecNr (RecNr)

File: KE041.CMD2015.v1

Overview

Type: Continuous	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-172247	Maximum: 78480
	Mean: 39240.5
	Standard deviation: 22655.4

Description

A sequential number uniquely identifying each record in the data file

CountryId (CountryId)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	
Range: 404-404	

Description

ISO 3166-1 numeric code of the country where the surveillance site is situated

CentreId (CentreId)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: character	Invalid: 0
Width: 5	

Description

An identifier issued by INDEPTH to each member centre of the format CCNNNS, Where CCNNNS is the INDEPTH Member site code, constructed as follows:

- CC the ISO 3166-1 alpha-2 code of the country where the site is situated
- NN is a sequential number uniquely identifying an INDEPTH member centre within the country. Starting with 01 to 99, for a maximum of 99 centres per country
- S is a sequential character uniquely identifying the geographical surveillance site. Starting with 1 to 9, thereafter A to Z, for a maximum of 35 sites per centre.

IndividualId (IndividualId)

File: KE041.CMD2015.v1

Overview

Type: Continuous	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-79480	Maximum: 36438
	Mean: 18212.2
	Standard deviation: 10531.8

Description

A number uniquely identifying all the records belonging to a specific individual in the data file. This number is not the same as the identifier used by a contributing centre to be identify the individual

Sex (Sex)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

Description

Sex of the individual

- 0 - Unknown
- 1 - Male
- 2 - Female

DoB (DoB)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: character	Minimum: NaN
	Maximum: NaN

Description

The date of birth of the individual Format YYYY-MM-DD

EventCount (EventCount)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	
Range: 2-5	

Description

The total number of events associated with this individual in this data set

EventNr (EventNr)

File: KE041.CMD2015.v1

Overview

Type: Discrete	Valid cases: 78480
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-5	

Description

A number increasing from 1 to EventCount for each event record in order of event occurred

EventCode (EventCode)

File: KE041.CMD2015.v1

Overview

EventCode (EventCode)

File: KE041.CMD2015.v1

Type: Discrete
Format: character
Width: 3

Valid cases: 78480
Invalid: 0

Description

A code which identifies the type of event that has occurred

EventDate (EventDate)

File: KE041.CMD2015.v1

Overview

Type: Discrete
Format: character

Valid cases: 78480
Minimum: NaN
Maximum: NaN

Description

The date on which the event occurred Format: YYYY-MM-DD

ObservationDate (ObservationDate)

File: KE041.CMD2015.v1

Overview

Type: Discrete
Format: character

Valid cases: 78480
Minimum: NaN
Maximum: NaN

Description

Date on which the event was observed (recorded), also known as surveillance visit date Format YYYY-MM-DD

LocationId (LocationId)

File: KE041.CMD2015.v1

Overview

Type: Continuous
Format: numeric
Decimals: 0
Range: 1-17889

Valid cases: 27119
Invalid: 51361
Minimum: 1
Maximum: 4570
Mean: 2353.6
Standard deviation: 1344.8

Description

Unique identifier associated with a residence unit within the site and is the location where the individual was or became resident, when the event occurred. this identifier is not the same as the identifier used internally by the contributing centre

MotherId (MotherId)

File: KE041.CMD2015.v1

Overview

MotherId (MotherId)

File: KE041.CMD2015.v1

Type: Continuous
Format: numeric
Decimals: 0
Range: 7-79457

Valid cases: 12017
Invalid: 66463
Minimum: 24
Maximum: 36439
Mean: 17906.2
Standard deviation: 10504.5

Description

The individual of the mother, only provided for BTH events

DeliveryId (DeliveryId)

File: KE041.CMD2015.v1

Overview

Type: Continuous
Format: numeric
Decimals: 0
Range: 3-6059

Valid cases: 12017
Invalid: 66463
Minimum: 1
Maximum: 12029
Mean: 5988
Standard deviation: 3459.4

Description

The RecNr of the delivery event associated with this birth

