

# Malawi - Karonga HDSS INDEPTH Core Dataset 2003-2016 (Release 2018)

**Albert Dube - Karonga Prevention Study**  
**Amelia C. Crampin - London School of Hygiene and Tropical Medicine**

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.MW011.CMD2016.v1

## Version

---

VERSION DESCRIPTION  
CMD2016.v1: Edited dataset for public distribution

PRODUCTION DATE  
2018-05-25

NOTES  
Developed in iSHARE2 workshop in Dubai May 2018

## Overview

---

### ABSTRACT

The Karonga Health and Demographic Surveillance System (Karonga HDSS) in northern Malawi currently has a population of more than 42 000 individuals under continuous demographic surveillance since completion of a baseline census (2002-2004). The surveillance system collects data on vital events and migration for individuals and for households. It also provides data on cause-specific mortality obtained by verbal autopsy for all age groups, and estimates rates of disease for specific presentations via linkage to clinical facility data. The Karonga HDSS provides a structure for surveys of socio-economic status, HIV-prevalence and incidence, sexual behavior, fertility intentions and a sampling frame for other studies, as well as evaluating the impact of interventions, such as antiretroviral therapy and vaccination programs. Uniquely, it relies on a network of village informants to report vital events and household moves, and furthermore is linked to an archive of biological samples and data from population surveys and other studies dating back three decades.

KIND OF DATA  
Event history data

UNITS OF ANALYSIS  
Individual

## Scope

---

NOTES  
The scope of this dataset includes the following:

All resident population of the HDSS

Enumeration

Birth

Death

Migration

## TOPICS

Topic	Vocabulary	URI
Demography [N01.224]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Emigration and Immigration [N01.224.625.350]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Mortality [N01.224.935.698]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>
Birth Rate [N01.224.935.849.500]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>

## KEYWORDS

Demography, Census, Vital statistics, Mortality, Population dynamics, Life expectancy, Malawi, Sub-Saharan Africa

## Coverage

## GEOGRAPHIC COVERAGE

The Karonga HDSS surrounds the port village of Chilumba, on Lake Malawi, in the southern part of Karonga District. The surveillance area is located between latitudes 10.388 and 10.508S and longitudes 34.088 and 34.278E, covering an area of approximately 135 km<sup>2</sup>. The lake shore and the Nyika National Park form the boundaries to the east and west, respectively, whereas the boundaries to the north and south are delineated by village boundaries

## UNIVERSE

The DSS covers all the individuals of all ages who are members of households and are residents within the geographical boundaries of the study area. Visitors too were captured in the baseline survey.

Of the 36,783 individuals under observation at the end of 2012, 48% (17,715/36,783) were male and 52% were female (19,068/36,783); 20% (7542/36,783) were aged under 5 years and 50% (18 396/ 36,783) were under 15 years of age. In all, 44% (16 277/36 783) of the population were aged between 15 and 59 years, and 6% (2160/36,783) were aged 60 years and above. The majority of the population under surveillance are rural dwellers that depend on subsistence farming and fish from the lake as a source of their livelihood.

## Producers and Sponsors

## PRIMARY INVESTIGATOR(S)

Name	Affiliation
Albert Dube	Karonga Prevention Study
Amelia C. Crampin	London School of Hygiene and Tropical Medicine

## OTHER PRODUCER(S)

Name	Affiliation	Role
Malawi Epidemiology & Intervention Research Unit	London School of Hygiene and Tropical Medicine	Original producer

## FUNDING

Name	Abbreviation	Role
Wellcome Trust		All stages of data production process

## OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Keith Branson	London School of Hygiene and Tropical Medicine	Data Manager
Estelle McLean	London School of Hygiene and Tropical Medicine	Data Scientist

## Metadata Production

## METADATA PRODUCED BY

<b>Name</b>	<b>Abbreviation</b>	<b>Affiliation</b>	<b>Role</b>
INDEPTH Network	INDEPTH	INDEPTH Network	agency
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical Support
Estelle McLean	EM	LSHTM	Data Scientist
Malawi Epidemiology & Intervention Research Unit	MEIRU	LSHTM	agency

## DATE OF METADATA PRODUCTION

2018-05-25

## DDI DOCUMENT VERSION

MW011.CMD2016.v1

## DDI DOCUMENT ID

DDI.INDEPTH.MW011.CMD2016.v1

## Sampling

### **Sampling Procedure**

---

The whole population

### **Deviations from Sample Design**

---

None

### **Response Rate**

---

The average response rate for household visits over all the rounds is around 99%

### **Weighting**

---

Not Applicable

# Questionnaires

## Overview

---

The questionnaires are designed to capture the core HDSS information which includes the baseline, birth, immigration, outmigration and death along with the other questionnaires.

## Data Collection

### Data Collection Dates

---

Start	End	Cycle
2003-01-01	2016-12-31	Release coverage

### Time Periods

---

Start	End	Cycle
2002-09-01		Baseline
2004-07-01		Round 1
2005-07-01		Round 2
2006-07-01		Round 3
2007-07-01		Round 4
2008-07-01		Round 5
2009-07-01		Round 6
2010-07-01		Round 7
2011-07-01		Round 8
2012-07-01		Round 9
2013-07-01		Round 10
2014-07-01		Round 11
2015-07-01		Round 12

### Data Collection Mode

---

Proxy Respondent [proxy]

### Questionnaires

---

The questionnaires are designed to capture the core HDSS information which includes the baseline, birth, immigration, outmigration and death along with the other questionnaires.

### Data Collectors

---

Name	Abbreviation	Affiliation
Field Workers		Karonga Prevention Study
Key Informants		Karonga Community Members

## Data Processing

### Data Editing

---

Data was left censored to 1 Jan 2003 to account for the start-up phase of the surveillance.

Data were edited to allow people to appear in only one household at a time (for surveillance purposes men with multiple wives can be recorded as living in more than one household).

### Other Processing

---

Data are entered manually from paper records onto an Access database. Data tables are extracted and processed in Stata prior to production of the core micro dataset.

# Data Appraisal

## Estimates of Sampling Error

Not Applicable

## Other forms of Data Appraisal

CentreId MetricTable QMetric Illegal Legal Total Metric RunDate

MW011 MicroDataCleaned Starts 72935 2018-05-25 13:59

MW011 MicroDataCleaned Transitions 0 250770 250770 0. 2018-05-25 13:59

MW011 MicroDataCleaned Ends 72935 2018-05-25 13:59

MW011 MicroDataCleaned SexValues 250770 2018-05-25 13:59

MW011 MicroDataCleaned DoBValues 250770 2018-05-25 13:59

# File Description

# Variable List

**MW011.CMD2016.v1**

Content	Event History Dataset of Karonga HDSS
Cases	281505
Variable(s)	14
Structure	Type: Keys: ()
Version	MW011.CMD2016.v1
Producer	Karonga HDSS
Missing Data	

**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	contin	numeric	
V8	EventNr	EventNr	contin	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: MW011.CMD2016.v1

### Overview

Type: Continuous	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-277338	Maximum: 281505
	Mean: 140753
	Standard deviation: 81263.6

### Description

A sequential number uniquely identifying each record in the data file

## CountryId (CountryId)

File: MW011.CMD2016.v1

### Overview

Type: Discrete	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	
Range: 454-454	

### Description

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

## CentreId (CentreId)

File: MW011.CMD2016.v1

### Overview

Type: Discrete	Valid cases: 281505
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: MW011.CMD2016.v1

### Overview

Type: Continuous	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-68653	Maximum: 81830
	Mean: 40869.3
	Standard deviation: 23594.7

### 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 the contributing centre to identify the individual.

## Sex (Sex)

File: MW011.CMD2016.v1

### Overview

Type: Discrete	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

### Description

Sex of individual

## DoB (DoB)

File: MW011.CMD2016.v1

### Overview

Type: Discrete	Valid cases: 281505
Format: character	Minimum: NaN
	Maximum: NaN

### Description

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

## EventCount (EventCount)

File: MW011.CMD2016.v1

### Overview

Type: Continuous	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 2
Range: 2-29	Maximum: 26
	Mean: 5.3
	Standard deviation: 3.2

### Description

The total number of events associated with the individual in this dataset

## EventNr (EventNr)

File: MW011.CMD2016.v1

### Overview

Type: Continuous	Valid cases: 281505
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-29	Maximum: 26
	Mean: 3.1
	Standard deviation: 2.4

### Description

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

## EventCode (EventCode)

File: MW011.CMD2016.v1

### Overview

## EventCode (EventCode)

File: MW011.CMD2016.v1

Type: Discrete  
Format: character  
Width: 3

Valid cases: 281505  
Invalid: 0

### Description

A code identifying the type of event that has occurred

## EventDate (EventDate)

File: MW011.CMD2016.v1

### Overview

Type: Discrete  
Format: character

Valid cases: 281505  
Minimum: NaN  
Maximum: NaN

### Description

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

## ObservationDate (ObservationDate)

File: MW011.CMD2016.v1

### Overview

Type: Discrete  
Format: character

Valid cases: 0

### Description

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

## LocationId (LocationId)

File: MW011.CMD2016.v1

### Overview

Type: Continuous  
Format: numeric  
Decimals: 0  
Range: 1-19276

Valid cases: 0  
Invalid: 281505

### Description

Unique identifier associated with a residential unit within the site and is the location where the individual was or became resident when the event occurred. For data anonymization purposes, this identifier should not be the same as the identifier used internally by the contributing centre, but the contributing centre should retain a mapping of this identifier to their internal location identifier

## MotherId (MotherId)

File: MW011.CMD2016.v1

### Overview

## MotherId (MotherId)

File: MW011.CMD2016.v1

Type: Continuous  
Format: numeric  
Decimals: 0  
Range: 9-68641

Valid cases: 18554  
Invalid: 262951  
Minimum: 2  
Maximum: 81819  
Mean: 40566  
Standard deviation: 23590.1

### Description

The IndividualId of the mother. Only provided for BTH event

## DeliveryId (DeliveryId)

File: MW011.CMD2016.v1

### Overview

Type: Continuous  
Format: numeric  
Decimals: 0  
Range: 1-16416

Valid cases: 18554  
Invalid: 262951  
Minimum: 1  
Maximum: 17604  
Mean: 8891.5  
Standard deviation: 5095.7

### Description

The RecNr of the delivery event associated with this birth

