

Malawi - Karonga HDSS INDEPTH Core Dataset 2003 - 2014 (Release 2016)

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

Report generated on: June 30, 2016

Visit our data catalog at: <http://indepth-ishare.org/index.php>

Overview

Identification

ID NUMBER

INDEPTH.MW011.CMD2014.v1

Version

VERSION DESCRIPTION

CMD2014.v1: First version for public distribution

PRODUCTION DATE

2016-05-17

Overview

ABSTRACT

The Karonga Health and Demographic Surveillance System (Karonga HDSS) in northern Malawi currently has a population of more than 38 000 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	http://www.ncbi.nlm.nih.gov/mesh

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 km². 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
Karonga Prevention Study	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

Name	Abbreviation	Affiliation	Role
Estelle McLean	EM	LSHTM	Data Scientist
Karonga Prevention Study	KPS	LSHTM	agency
INDEPTH Network	INDEPTH	INDEPTH Network	agency
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical Support

DATE OF METADATA PRODUCTION

2016-05-18

DDI DOCUMENT VERSION

CMD2014.v1: 2016-05-18

DDI DOCUMENT ID

DDI.INDEPTH.MW011.CMD2014.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	2014-12-31	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

Data Collection Mode

Proxy Respondent [proxy]

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	68081	2016-05-17	13:21		
MW011	MicroDataCleaned	Transitions	0	227156	227156	0.	2016-05-17 13:21
MW011	MicroDataCleaned	Ends	68081	2016-05-17	13:21		

File Description

Variable List

MW011.CMD2014.v1

Content	Event HistoryDataset
Cases	261044
Variable(s)	14
Structure	Type: Keys: ()
Version	CMD2014.v1
Producer	
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	discrete	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	discrete	numeric	

RecNr (RecNr)

File: MW011.CMD2014.v1

Overview

Type: Continuous	Valid cases: 261044
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-192239	Maximum: 261044
	Mean: 130522.5
	Standard deviation: 75357.1

Description

A sequential number uniquely identifying each record in the data file

CountryId (CountryId)

File: MW011.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
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.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
Format: character	Invalid: 0
Width: 5	

Description

An identifier issued by INDEPTH to each member centre of the format CCNNS, Where CCNNS 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.CMD2014.v1

Overview

Type: Continuous	Valid cases: 261044
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 54
Range: 1-60235	Maximum: 129931
	Mean: 95508
	Standard deviation: 20866.2

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.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

Description

Sex of individual

DoB (DoB)

File: MW011.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
Format: character	Minimum: NaN
Width: 23	Maximum: NaN

Description

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

EventCount (EventCount)

File: MW011.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
Format: numeric	Invalid: 0
Decimals: 0	
Range: 2-22	

Description

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

EventNr (EventNr)

File: MW011.CMD2014.v1

Overview

Type: Continuous	Valid cases: 261044
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-22	Maximum: 27
	Mean: 3.2
	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.CMD2014.v1

Overview

Type: Discrete	Valid cases: 261044
Format: character	Invalid: 0
Width: 3	

EventCode (EventCode)

File: MW011.CMD2014.v1

Description

A code identifying the type of event that has occurred

EventDate (EventDate)

File: MW011.CMD2014.v1

Overview

Type: Discrete

Format: character

Width: 23

Valid cases: 261044

Minimum: NaN

Maximum: NaN

Description

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

ObservationDate (ObservationDate)

File: MW011.CMD2014.v1

Overview

Type: Discrete

Format: character

Width: 23

Valid cases: 261044

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: MW011.CMD2014.v1

Overview

Type: Continuous

Format: numeric

Decimals: 0

Range: 1-14463

Valid cases: 261044

Invalid: 0

Minimum: 1

Maximum: 19053

Mean: 7734.2

Standard deviation: 4836.7

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.CMD2014.v1

Overview

Type: Continuous

Format: numeric

Decimals: 0

Range: 13-25530

Valid cases: 30719

Invalid: 230325

Minimum: 1084

Maximum: 129781

Mean: 95355.7

Standard deviation: 20277.2

MotherId (MotherId)

File: MW011.CMD2014.v1

Description

The IndividualId of the mother. Only provided for BTH event

DeliveryId (DeliveryId)

File: MW011.CMD2014.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 1-6

Valid cases: 30719
Invalid: 230325

Description

The RecNr of the delivery event associated with this birth

