

Kenya - Nairobi HDSS INDEPTH Core Dataset 2003-2014 (Release 2017)

Dr.Alex Ezeh - Site Leader, KE031, Dr.Donatien Beguy - Site Representative, KE031

Report generated on: June 30, 2017

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

Overview

Identification

ID NUMBER

INDEPTH.KE031.CMD2014.v1

Version

VERSION DESCRIPTION

CMD2014.v1: Cleaned and anonymised dataset of the Core Microdata 2002-2014 for public distribution

PRODUCTION DATE

2017-05-17

Overview

ABSTRACT

The places we live affect our health status and the choices and opportunities we have (or do not have) to lead fulfilling lives. Over the past ten years, the African Population & Health Research Centre (APHRC) has led pioneering work in highlighting some of the major health and livelihood challenges associated with rapid urbanization in sub-Saharan Africa (SSA). In 2002, the Centre established the first longitudinal platform in urban Africa in the city of Nairobi in Kenya. The platform known as the Nairobi Urban Health and Demographic Surveillance System collects data on two informal settlements - Korogocho and Viwandani - in Nairobi City every four months on issues ranging from household dynamics to fertility and mortality, migration and livelihood as well as on causes of death, using a verbal autopsy technique. The dataset provided here contains key demographic and health indicators extracted from the longitudinal database. Researchers interested in accessing the micro-data can look at our data access policy and contact us.

KIND OF DATA

Event history data

UNITS OF ANALYSIS

Individual

Scope

NOTES

- CORE DATA: Enumeration, Birth, Migration and Death

TOPICS

Topic	Vocabulary	URI
Demography [N01.224]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Age Distribution [N01.224.033]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Censuses [N01.224.175]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Family Characteristics [N01.224.361]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Health Status Disparities [N06.850.505.400.425.675]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Population Dynamics [N01.224.625]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Residential Mobility [N01.224.791.700]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Residence Characteristics [N01.224.791]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Sex Ratio [N01.224.803.815]	MeSH	http://www.ncbi.nlm.nih.gov/mesh

Topic	Vocabulary	URI
Vital Statistics [N01.224.935]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Urban Health [N01.400.800]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Urban Population [N01.600.900]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Educational Status [N01.824.196]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Employment [N01.824.245]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Occupations [N01.824.547]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Poverty [N01.824.600]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Social Conditions [N01.824.827]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Policy [N03.623]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Health Services Research [N05.425]	MeSH	http://www.ncbi.nlm.nih.gov/mesh
Epidemiologic Factors [N05.715.350]	MeSH	http://www.ncbi.nlm.nih.gov/mesh

KEYWORDS

Migration, Fertility, Mortality, Population

Coverage**GEOGRAPHIC COVERAGE**

The Demographic Surveillance Area (combining Viwandani and Korogocho slum settlements) covers a land area of about 0.97 km², with the two informal settlements located about 7 km from each other. Korogocho is located 12 km from the Nairobi city center; in Kasarani division (now Kasarani district), while Viwandani is about 7 km from Nairobi city center in Makadara division (now Madaraka district). The DSA covers about seven villages each in Korogocho and Viwandani.

UNIVERSE

Between 1st January and 31st December, 2015 the Nairobi HDSS covered 86,304 individuals living in 30,219 households distributed across two informal settlements (Korogocho and Viwandani) were observed. All persons who sleep in the household prior to the day of the survey are included in the survey, while non-resident household members are excluded from the survey.

The present universe started out through an initial census carried out on 1st August, 2002 of the population living in the two informal settlements (Korogocho and Viwandani). Regular visits have since then been made (3 times a year) to update information on births, deaths and migration that have occurred in the households observed at the initial census. New members join the population through a birth to a registered member, or an in-migration, while existing members leave through a death or out-migration. The DSS adopts the concept of an open cohort that allows new members to join and regular members to leave and return to the system.

Producers and Sponsors**PRIMARY INVESTIGATOR(S)**

Name	Affiliation
Dr. Alex Ezeh	Site Leader, KE031
Dr. Donatien Beguy	Site Representative, KE031

OTHER PRODUCER(S)

Name	Affiliation	Role
Clement Oduor	Field Coordinator	Data Collection
Kanyiva Muindi	Research Officer	Questionnaire Design
Nelson Mbaya	Database Programmer	Data Processing

Name	Affiliation	Role
Marylene Wamukoya	Data Manager	Data Cleaning and Analysis

FUNDING

Name	Abbreviation	Role
Bill and Melinda Gates Foundation, USA		Current Funder
William and Flora Hewlett Foundation, USA		Current Funder
Swedish International Development Cooperation Agency		Current Funder
Wellcome Trust, UK		Previous Funder
Rockefeller Foundation, USA		Previous Funder

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Residents of Korogocho and Viwandani Slums	APHRC	Study Subjects

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical Support
INDEPTH Network	int.indepth	INDEPTH	Agency
Data & Measurement Unit of KE031	KE031	Nairobi Urban HDSS, APHRC	Documentation of the Study
Nelson Mbaya	KE031	Nairobi Urban HDSS, APHRC	DDI Author

DATE OF METADATA PRODUCTION

2017-05-17

DDI DOCUMENT VERSION

CMD2014.V1 (17 May 2017)

DDI DOCUMENT ID

DDI.INDEPTH.KE031.CMD2014.v1

Sampling

Sampling Procedure

This dataset is related to the whole demographic surveillance area population. The number of respondents has varied over the last 13 years (2002-2015), with variations being observed at both household level and at Individual level. As at 31st December 2015, 66,848 were being observed under the Nairobi HDSS living in 25,812 households distributed across two informal settlements (Korogocho and Viwandani). The variable IndividualId uniquely identifies every respondent observed while the variable LocationId uniquely identifies the room in which the individual was living at any point in time. To identify individuals who were living together at any one point in time (a household) the data can be split on location and observation dates.

Deviations from Sample Design

None

Response Rate

Over the years the response rate at household level has varied between 95% and 97% with response rate at Individual Level varying between 92% and 95%. Challenges to achieving a 100% response rate have included:

1. High Population Mobility within the study area
2. High Population Attrition
3. Respondent Fatigue
4. Security in some areas

Weighting

Not applicable

Questionnaires

Overview

Questionnaires are printed and administered in Swahili (The country's national language)

The questionnaires for the Nairobi HDSS were structured questionnaires based on the INDEPTH Model Questionnaire and were translated into Swahili with some modifications and additions. After an initial review the questionnaires were translated back into English by an independent translator with no prior knowledge of the survey. The back translation from the Swahili version was independently reviewed and compared to the English original. Differences in translation were reviewed and resolved in collaboration with the original translators.

The English and Swahili questionnaires were both piloted as part of the survey pretest.

At baseline a household questionnaire was administered in each household, which collected various information on household members including sex, age, relationship, and orphanhood status. In later rounds questionnaires to track the migration of the population observed at baseline, and additional questionnaires to capture demographic and health events happening to the population have been introduced.

Data Collection

Data Collection Dates

Start	End	Cycle
2003-01-01	2014-12-31	Release Coverage

Time Periods

Start	End	Cycle
2002-08-01		Round 0
2002-10-16		Round 1
2003-01-21		Round 2
2003-05-04		Round 3
2003-09-02		Round 4
2004-01-23		Round 5
2004-05-28		Round 6
2004-09-25		Round 7
2005-01-23		Round 8
2005-06-14		Round 9
2005-09-27		Round 10
2006-01-13		Round 11
2006-05-01		Round 12
2006-09-01		Round 13
2007-02-24		Round 14
2007-05-29		Round 15
2007-09-25		Round 16
2008-01-01		Round 17
2008-09-01		Round 18
2009-01-01		Round 19
2009-06-23		Round 20
2009-09-26		Round 22
2010-01-20		Round 21
2010-05-04		Round 23
2010-09-01		Round 24
2011-01-01		Round 25
2011-05-05		Round 26
2011-09-09		Round 27
2012-01-01		Round 28
2012-05-02		Round 29
2012-09-04		Round 30
2013-01-27		Round 31
2013-06-01		Round 32
2013-10-23		Round 33
2014-02-03		Round 34
2014-05-16		Round 35
2014-09-15		Round 36
2014-12-25		Round 37
2015-06-18		Round 38
2016-01-21		Round 39
2016-06-08		Round 40
2016-09-28		Round 41

Data Collection Mode

Proxy Respondent [proxy]

DATA COLLECTION NOTES

Interviews are conducted in Swahili (the country's national language), with translation into the local dialects in specific

situations where the respondent is unable to understand Swahili.

A detailed 4-week training of enumerators was conducted at baseline. Currently, 3-5 days refresher trainings are conducted at the start of each new round of data collection

The field team constitutes 32 staff: 25 interviewers, 4 team leaders, 2 supervisors, and 1 field coordinator. On average a complete household interview takes between 30 minutes and 45 minutes to complete. Interviews took place everyday throughout the field work period, with teams being permitted to take only one day off per week.

Prior to the baseline detailed community sensitization was undertaken, in latter rounds however community liaison has constituted, feedback barazas, and intervention activities targeted to address community health-related problems identified during the data collection. Intervention activities are carried out through piggy-back studies with targeted interventions, for example a maternal and child health study had a component of interventions to address mother and child health-complications, or through bi-annual health camps at which free medical treatment is provided for conditions identified as prevalent in the community during data collection.

Data Collectors

Name	Abbreviation	Affiliation
Data and Measurement Unit	D&M	African Population and Health Research Center

SUPERVISION

Interviewing is conducted by teams of interviewers. Each interviewing team comprised 3-4 interviewers, a field editor and a site supervisors. The role of the supervisor is to coordinate field data collection activities, including management of the field teams, supplies and equipment, finances, maps and listings, coordinate with local authorities concerning the survey plan and make arrangements for accommodation and travel. Additionally, the field supervisor assigned the work to the interviewers, spot checked work, maintained field control documents, and sent completed questionnaires and progress reports to the central office. The field editor is responsible for reviewing each questionnaire at the end of the day, checking for missed questionnaires, skip errors, fields incorrectly completed, and checking for inconsistencies in the data. The field editor also observed interviews and conducted review sessions with interviewers. Responsibilities of the supervisors and field editors are described in the Instructions for Supervisors and Field Editors, together with the different field controls that were in place to control the quality of the fieldwork.

Data Processing

Data Editing

Data editing took place at a number of stages throughout the processing, including:

- a) Office editing and coding
- b) During data entry
- c) Structure checking and completeness
- d) Secondary editing
- e) Structural checking of STATA data files

Detailed documentation of the editing of data can be found in the "Standard Procedures Manual" document provided as an external resource.

Where changes were made by the program, a cold deck imputation is preferred; where incorrect values were imputed using existing data from another dataset. If cold deck imputation was found to be insufficient, hot deck imputation was used. In this case, a missing value was imputed from a randomly selected similar record in the same dataset.

Some corrections are made automatically by the program(80%) and the rest by visual control of the questionnaires (20%)

1. 100% forms filled in by FRAs are rechecked for completeness, ensured that all the necessary event forms are filled in.
2. Spot checks are done on field over data collection by FRAs for reliability of data.
3. FRS instructs revisits wherever required.
4. Forms are checked on sample basis
5. Checks if all the necessary event forms are filled in.
6. Forms with inconsistencies identified at the time of entry are sent back to the field.
7. Creating and managing data entry checks for picking up inconsistencies
8. Monitoring field work: balancing work target and quality.
9. Dealing with data inconsistencies at data level and giving feedbacks to field staff.
10. Conducting training and refresher training wherever required.
11. Data cleaning

Other Processing

Data entry was performed manually at APHRC's headquarters on desktop computers and using an in-house built software with a Visual Basic.Net front-end and Microsoft SQL Server back-end. Double data entry was carried out on 10% of the questionnaires.

Data were processed in clusters, with each cluster being processed as a complete unit through each stage of data processing. Each cluster went through the following steps:

1. Questionnaire Reception
2. Office editing and coding
3. Data entry
4. Structure and completeness checking
5. Verification entry
6. Comparison of verification data
7. Backup of raw data
8. Secondary editing
9. Edited data back up

After all clusters are processed, all data was concatenated together and then the following steps are completed for all data files:

- 10) Export to STATA 13 in 2 files (migration and employment history, migration & employment calendar)
- 11) Recoding of variables needed for analysis
- 12) Structural checking of STATA 13 files
- 14) Data quality tabulations
- 15) Production of analysis tabulations

Details of each of these steps can be found in the "Standard Procedures Manual"

Data Appraisal

Estimates of Sampling Error

Not applicable for Surveillance Data

Other forms of Data Appraisal

CentreId	MetricTable	QMetric	Illegal	Legal	Total	Metric	RunDate	
KE031	MicroDataCleaned	Starts	219285	2017-05-16	18:25			
KE031	MicroDataCleaned	Transitions	825036	825036	0	2017-05-16	18:25	
KE031	MicroDataCleaned	Ends	219285	2017-05-16	18:25			
KE031	MicroDataCleaned	SexValues	825036	2017-05-16	18:25			
KE031	MicroDataCleaned	DoBValues	42	824994	825036	0	2017-05-16	18:25

File Description

Variable List

KE031.CMD2014.v1

Content	This file contains the INDEPTH Core Microdataset of Nairobi Urban HDSS. The file was generated using ETL through Pentaho Kettle.
Cases	964531
Variable(s)	14
Structure	Type: Keys: ()
Version	CMD2014.v1
Producer	Nairobi Urban HDSS
Missing Data	Missing Data is coded as follows: -Missing Data: No data or a missing data is assigned the code(95,995,etcetera). All missing values should be coded as 5,95,995, etcetera, depending on the value of the largest valid code in that variable. - Response Not Within the Pre-Defined Range/Domain: A data code(96,996,etcetera) provided where response was outside the range/domain pre-defined during study design. A variable that contains this code is often succeeded by a variable that contains the specific response. Check the succeeding variable to ensure no missing values exist where a response was expected. -Refusals: A data code (97,997,etcetera) is used to indicate that the respondent refused to respond to this question. -"Don't Know" Responses: A data code(98,998,etcetera) is used to indicate that the respondent did not know the answer to the question. -Skipped Questions: A data code(99,999,etcetera) is used to indicate that the respondent was not eligible to answer the particular question.

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	discrete	numeric	

RecNr (RecNr)

File: KE031.CMD2014.v1

Overview

Type: Continuous	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-637510	Maximum: 964531
	Mean: 482266
	Standard deviation: 278436.3

Description

A sequential number uniquely identifying each record in the data file

CountryId (CountryId)

File: KE031.CMD2014.v1

Overview

Type: Discrete	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	
Range: 404-404	

Description

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

CentreId (CentreId)

File: KE031.CMD2014.v1

Overview

Type: Discrete	Valid cases: 964531
Format: character	Invalid: 0
Width: 5	

Description

An identifier issued by INDEPTH to each member centre of the format CCCSS, where CCC is a sequential centre identifier and SS is a sequential identifier of the site within the centre in the case of multiple site centres

IndividualId (IndividualId)

File: KE031.CMD2014.v1

Overview

Type: Continuous	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-177721	Maximum: 211602
	Mean: 98554.2
	Standard deviation: 57807.9

Description

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

Sex (Sex)

File: KE031.CMD2014.v1

Sex (Sex)

File: KE031.CMD2014.v1

Overview

Type: Discrete	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

Description

Sex of the individual.

DoB (DoB)

File: KE031.CMD2014.v1

Overview

Type: Discrete	Valid cases: 964531
Format: character	Minimum: NaN
Width: 10	Maximum: NaN

Description

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

EventCount (EventCount)

File: KE031.CMD2014.v1

Overview

Type: Continuous	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 2
Range: 2-32	Maximum: 36
	Mean: 6.4
	Standard deviation: 4.1

Description

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

EventNr (EventNr)

File: KE031.CMD2014.v1

Overview

Type: Continuous	Valid cases: 964531
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-32	Maximum: 36
	Mean: 3.7
	Standard deviation: 3

Description

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

EventCode (EventCode)

File: KE031.CMD2014.v1

Overview

EventCode (EventCode)

File: KE031.CMD2014.v1

Type: Discrete
 Format: character
 Width: 3

Valid cases: 964530
 Invalid: 0

Description

A code identifying the type of event that has occurred.

EventDate (EventDate)

File: KE031.CMD2014.v1

Overview

Type: Discrete
 Format: character
 Width: 10

Valid cases: 964531
 Minimum: NaN
 Maximum: NaN

Description

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

ObservationDate (ObservationDate)

File: KE031.CMD2014.v1

Overview

Type: Discrete
 Format: character
 Width: 10

Valid cases: 69754
 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: KE031.CMD2014.v1

Overview

Type: Continuous
 Format: numeric
 Decimals: 0
 Range: 1-43728

Valid cases: 964529
 Invalid: 2
 Minimum: 1
 Maximum: 48237
 Mean: 22374.2
 Standard deviation: 12900.1

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. This identifier is not be the same as the identifier used internally by the contributing centre.

MotherId (MotherId)

File: KE031.CMD2014.v1

Overview

MotherId (MotherId)

File: KE031.CMD2014.v1

Type: Continuous
Format: numeric
Decimals: 0
Range: 7-177720

Valid cases: 46424
Invalid: 918107
Minimum: 3
Maximum: 211595
Mean: 95372.2
Standard deviation: 56181.1

Description

The IndividualId of the mother. Only provided for BTH events.

DeliveryId (DeliveryId)

File: KE031.CMD2014.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 1-7

Valid cases: 46424
Invalid: 918107

Description

The RecNr of the delivery event associated with this birth

