South Africa - Dikgale HDSS INDEPTH Core Dataset 1996-2016 (Release 2018)

Marianne Alberts - Dikgale HDSS , University of Limpopo (ZA021)
Sandra Burger - Dikgale HDSS , University of Limpopo (ZA021)

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.ZA021.CMD2016.v1

Version

VERSION DESCRIPTION
CMD2016.v1: Edited Core Microdata 2016, for public distribution, extracted on 23 May 2018.

PRODUCTION DATE
2018-05-23

NOTES
CMD2016.v1

Overview

ABSTRACT
Changes in socio-economic status and lifestyle behaviors among the adult population in South Africa not only in urban areas but also in rural South Africa, have led to increased prevalence of chronic diseases and associated risk factors, together with an epidemic of some infectious diseases. Researchers at the University of the North (now University of Limpopo Turfloop campus) established The Dikgale centre for Health and Demographic surveillance system in 1996 funded by a core grant from NUFU, Norway.

The broad aim of the Dikgale HDSS is to provide information to improve health of the people in Limpopo province and to assist the local government in making effective health care policy. As few data are available on the prevalence of diseases in rural and peri-urban areas of Limpopo province, the initial objective of the HDSS was to establish a field site where the incidence and prevalence of diseases could be assessed. It was therefore necessary to collect longitudinal demographic data (e.g. mortality, fertility, migration) on the population. To this end, three primary subjects are observed longitudinally in Dikgale HDSS: physical structures (e.g. homesteads, clinics and schools), households and individuals. The information about these subjects, and all related information, was at first stored in Access and later in a single MSSQL Server database, in a truly longitudinal way-i.e. not as a series of cross-sections.

The surveillance area is located in the Capricorn district, Limpopo province approximately 40 km from Polokwane, the capital city of Limpopo province and 15-50 km from the University of Limpopo (Turfloop campus). The site covers an area of approximately 200 square kilometers. The initially the total population was 8000 but the field site was expanded in 2010 and now includes approximately 40,000 people who are members of approximately 7000 households. The households are present in 15 villages of varying sizes. The population is predominantly Northern Sotho speaking. All households have electricity. Some households have piped water either inside the house or in their yards, but most fetch water from taps situated at strategic points in the villages. Most households have a pit latrine in their yards. A large proportion of adults are migrant workers, while others work as farm laborers on neighboring farms, or as domestic workers in nearby towns. Many are pensioners. The unemployment rate in the area is high.

To fulfil the eligibility criteria for the Dikgale HDSS cohort, individuals must be a member of a household within the surveillance area but not necessarily resident within it. Crucially, this means that Dikgale HDSS collects information on resident and non-resident members of households and makes a distinction between membership (self-defined on the basis of links to other household members) and residency (residing at a physical structure within the surveillance area at a particular point in time). Individuals can be members of more than one household at any point in time (e.g. polygamous married men whose wives maintain separate households). To be consistent with similar datasets from other INDEPTH Member centres, this data set contains data from resident members only.

During data collection, households are visited by fieldworkers and information supplied by a single key informant. All births, deaths and migrations of household members are recorded. If household members have moved internally within the

surveillance area, such moves are reconciled and the internal migrant retains the original identifier associated with him/her.

KIND OF DATA
Event history data

UNITS OF ANALYSIS
Individual

Scope

NOTES
This study represents only a portion of the total data associated with the complete Dikgale demographic surveillance as described in the study abstract. It specifically only includes the events defining the resident exposure of individuals under surveillance as well as the delivery events of resident women. Each type of event contains minimal attributes describing the event:

Attributes common to each event: Event Type, Event Date, Observation date

Migration: Origin & Destination

Death: Cause

Delivery: Live born and Still born counts, Parity

TOPICS

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

KEYWORDS
Fertility, Mortality, Migration

Coverage

GEOGRAPHIC COVERAGE
Demographic surveillance area situated in Capricorn District 40 km north- east of Polokwane the capital city of Limpopo province. And 20 -30 km from the University of Limpopo. The area is approximately 310 square kilometers.

UNIVERSE
Resident household members of households within the demographic surveillance area. Immigrants are defined by intention to become resident, but actual residence episodes of less than 180 days are censored. Outmigrants are defined by intention to become resident elsewhere, but actual periods of non-residence less than 180 days are censored. Children born to resident women are considered resident by default, irrespective of actual place of birth. The dataset contains the events of all individuals ever resident during the study period (only covering the original field site of 8000 people.(1 January,1996 to 31 Dec 2016)

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Marianne Alberts	Dikgale HDSS , University of Limpopo (ZA021)
Sandra Burger	Dikgale HDSS , University of Limpopo (ZA021)

OTHER PRODUCER(S)

Name	Affiliation	Role
Tlouyamma Joseph	Dikgale HDSS - University of Limpopo	Data Manager
Letsoalo Peter	Dikgale HDSS - University of Limpopo	Data Manager

FUNDING

Name	Abbreviation	Role
Vlaamse Interuniversitaire Raad	VLIR	Current funder
Norwegian Universities Committee for Development Research	NUFU	Prior Funder

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Timotheus Darikwa	Dikgale HDSS,University of Limpopo (ZA021)	Data Specialist
Ian Cook	Dikgale HDSS,University of Limpopo (ZA021)	Data Manager
Dikgale Community members		

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
iSHARE2 Technical Team	iS2TT	INDEPTH Network	Technical Support
INDEPTH Network	int.INDEPTH	INDEPTH Network	Agency
Dikgale Health and Demographic Surveillance Site	ZA021	University of Limpopo	Documentation of the study
Tlouyamma Joseph	TJ	University of Limpopo	DDI Author

DATE OF METADATA PRODUCTION
2018-05-23

DDI DOCUMENT VERSION
Version CMD2016.v1 (May 2018)

DDI DOCUMENT ID
DDI.INDEPTH.ZA021.CMD2016.v1

Sampling

Sampling Procedure

This dataset is not based on a sample but contains information from the complete demographic surveillance area.

Deviations from Sample Design

Not Applicable.

Response Rate

On an average, it is 99% over the years in all rounds

Weighting

Not applicable.

Questionnaires

Overview

Overview

List of questionnaires

Bounded structure registration (BSR) or update (BSU) form

• used to register characteristics of the BS

• Updates characteristics of the BS

• Information as at previous round is preprinted

Household registration (HHR) or update (HHU) form

• used to register characteristics of the HH

• Used to update information about the composition of the household

• Information pre-printed of composition and all registered households as at previous.

Household Membership Registration (HMR) or update (HMU)

• used to link individuals to households.

• Used to update information about the household memberships and member status observations

• Information preprinted of member status observations as at previous.

Individual registration form (IDR)

• Used to uniquely identify each individual

• Mainly to ensure members with multiple household memberships are appropriately

captured

Migration notification form (MGN)

• Used to record change in the BS of residency of individuals or households

• Migrants are tracked and updated in the database

Pregnancy history form (PGH) & pregnancy outcome notification form (PON)

• Records details of pregnancies and their outcomes

• Only if woman is a new member

• Only if woman has never completed WHL or WGH

Death notification form (DTN)

• Records all deaths that have recently occurred

• includes information about time, place, circumstances and possible cause of death

Data Collection

Data Collection Dates

Start	End	Cycle
1996-01-01	2016-12-31	Release Coverage

Time Periods

Start	End	Cycle
1995-08-01		Round -15
1996-08-01		Round -14
1997-08-01		Round -13
1998-08-01		Round -12
1999-08-01		Round -11
2000-08-01		Round -10
2001-08-01		Round -9
2002-08-01		Round -8
2003-08-01		Round -7
2004-08-01		Round -6
2005-08-01		Round -5
2006-08-01		Round -4
2007-08-01		Round -3
2008-08-01		Round -2
2009-08-01		Round -1
2010-08-01		Round 1
2011-08-01		Round 2
2012-08-01		Round 3
2013-08-01		Round 4
2014-08-01		Round 5
2015-08-01		Round 6
2016-08-01		Round 7

Data Collection Mode

Proxy Respondent [proxy]

Data Collection Notes

Enumerators were trained immediately prior to the baseline data collection. as data collectors. Data entry staff received fieldwork training in addition to training in the use of the data entry programs.

Questionnaires

Overview

List of questionnaires

Bounded structure registration (BSR) or update (BSU) form

• used to register characteristics of the BS

• Updates characteristics of the BS

• Information as at previous round is preprinted

Household registration (HHR) or update (HHU) form

• used to register characteristics of the HH

• Used to update information about the composition of the household

• Information pre-printed of composition and all registered households as at previous.

Household Membership Registration (HMR) or update (HMU)

• used to link individuals to households.

• Used to update information about the household memberships and member status observations

• Information preprinted of member status observations as at previous.

Individual registration form (IDR)

• Used to uniquely identify each individual

• Mainly to ensure members with multiple household memberships are appropriately

captured

Migration notification form (MGN)

• Used to record change in the BS of residency of individuals or households

• Migrants are tracked and updated in the database

Pregnancy history form (PGH) & pregnancy outcome notification form (PON)

• Records details of pregnancies and their outcomes

• Only if woman is a new member

• Only if woman has never completed WHL or WGH

Death notification form (DTN)

• Records all deaths that have recently occurred

• includes information about time, place, circumstances and possible cause of death

Data Collectors

Name	Abbreviation	Affiliation
Dikgale Health and Demographic Surveillance System	ZA021	UL

Supervision

There are two supervisor, one for each team. They are responsible for making sure that field operations are running smoothly. Taking forms to/from the field.

Data Processing

Data Editing

On data entry data consistency and plausibility were checked by 455 data validation rules at database level. If data validation failure was due to a data collection error, the questionnaire was referred back to the field for revisit and correction.

No imputations were done on the resulting micro data set, except for:

a. If an out-migration (OMG) event is followed by a homestead entry event (ENT) and the gap between OMG event and ENT event is greater than 180 days, the ENT event was changed to an in-migration event (IMG).

b. If an out-migration (OMG) event is followed by a homestead entry event (ENT) and the gap between OMG event and ENT event is less than 180 days, the OMG event was changed to an homestead exit event (EXT) and the ENT event date changed to the day following the original OMG event.

c. If a homestead exit event (EXT) is followed by an in-migration event (IMG) and the gap between the EXT event and the IMG event is greater than 180 days, the EXT event was changed to an out-migration event (OMG).

d. If a homestead exit event (EXT) is followed by an in-migration event (IMG) and the gap between the EXT event and the IMG event is less than 180 days, the IMG event was changed to an homestead entry event (ENT) with a date equal to the day following the EXT event.

e. If the last recorded event for an individual is homestead exit (EXT) and this event is more than 180 days prior to the end of the surveillance period, then the EXT event is changed to an out-migration event (OMG)

Other Processing

All homesteads in the Dikgale HDSS were geocoded and entered into a geographic information system (GIS) prior to the start of surveillance.

Before the round, a SQL script generated a list of questionnaires to be printed for each household resident in the surveillance area. Each questionnaire is given a unique integer key which is printed as a barcode on the questionnaire. A series of web-based reports called 'Unified Reports' are then used to track and control the status of each questionnaire from document production, data collection, data entry and document archiving. A strict chain of custody is enforced for all questionnaire movements.

A data entry is performed by one data capturer with one supervisor using in-house developed software (Delphi and .NET C#).
Data is stored in a MS SQL database, with transaction logging, daily backups and twice weekly off-site backups. Constraints and validation rules placed on the database help in checking data quality during data entry.

Because of error in data entry 100% recheck was done on all entries.

Field QC Procedures

The team leader checks the conduct of the interviewer.

After data collection and before data entry, the questionnaires are checked for completeness, consistency and accuracy. If a questionnaire failed to meet the quality standard requirements, the questionnaire is sent back the questionnaires to the team leader.

Specify how the data was extracted (including which software program was used) to produce the core micro data set. How was inconsistent records dealt with during this process?

Following data collection and data entry completion a snapshot of the operational database is created as an analytical database. This snapshot is uniquely identified and analytical datasets must reference the analytical database they originated from..

An SQL script produces a normalised episode table each time an analytical database is created. This episode table contains an exposure record for each exposure episode for an individual, from initial enumeration, birth or in-migration, up to eventual death or out-migration. The episode table contains the start event and date of the exposure as well as the end event and date of the end of exposure. Individuals that out-migrate and later in-migrate are reconciled as far as possible using individual identifiers (national identity number, names, sex and date of birth) under a single individual identity. All internal movements (migrations) are reconciled and residencies at different homesteads within the surveillance area are reflected as separate episodes in the episode table.

To produce this micro-data set, the episode table is processed using Pentaho Kettle ETL program to produce this standard event-history format dataset.

Data Appraisal

Estimates of Sampling Error

Not Applicable.

Other forms of Data Appraisal

CentreId MetricTable QMetric Illegal Legal Total Metric RunDate

ZA021 MicroDataCleaned Starts 55087 2018-05-23 13:14
ZA021 MicroDataCleaned Transitions 0 118228 118228 0. 2018-05-23 13:14
ZA021 MicroDataCleaned Ends 55087 2018-05-23 13:14
ZA021 MicroDataCleaned SexValues 6 118222 118228 0. 2018-05-23 13:14
ZA021 MicroDataCleaned DoBValues 8 118220 118228 0. 2018-05-23 13:14

File Description

Variable List

ZA021.CMD2016.v1

Content	This file contains Dikgale HDSS (ZA021) INDEPTH Core Mirco Datasetmin Event Hostory Format
Cases	135115
Variable(s)	14
Structure	Type: Keys: ()
Version	CMD2016.v1: For public distribution
Producer	Dikgale HDSS (ZA021)
Missing Data	

Variables

ID	Name	Label	Type	Format	Question
V1	RecNr	RecNr	contin	numeric	
V2	CountryId	CountryId	discrete	numeric	
V3	CentreId	CentreId	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	discrete	numeric	
V14	DeliveryId	DeliveryId	discrete	numeric	

RecNr (RecNr)
File: ZA021.CMD2016.v1

Overview

Type: Continuous
Format: numeric
Decimals: 0
Range: 1-74249

Valid cases: 135115
Invalid: 0
Minimum: 1
Maximum: 135115
Mean: 67558
Standard deviation: 39004.5

Description

A sequential number uniquely identifying each record in the data file

CountryId (CountryId)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 710-710

Valid cases: 135115
Invalid: 0

Description

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

CentreId (CentreId)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: character
Width: 5

Valid cases: 135115
Invalid: 0

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: ZA021.CMD2016.v1

Overview

Type: Continuous
Format: numeric
Decimals: 0
Range: 1-36405

Valid cases: 135115
Invalid: 0
Minimum: 1
Maximum: 55088
Mean: 27550.2
Standard deviation: 15894.6

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: ZA021.CMD2016.v1

Sex (Sex)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 0-2

Valid cases: 135115
Invalid: 0

Description

Sex of the individual. 1 for Male and 2 for Female

DoB (DoB)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: character

Valid cases: 135115
Minimum: NaN
Maximum: NaN

Description

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

EventCount (EventCount)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 2-4

Valid cases: 135115
Invalid: 0

Description

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

EventNr (EventNr)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 1-4

Valid cases: 135115
Invalid: 0

Description

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

EventCode (EventCode)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: character
Width: 3

Valid cases: 135115
Invalid: 0

Description

EventCode (EventCode)
File: ZA021.CMD2016.v1

A code identifying the type of event that has occurred.

EventDate (EventDate)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: character

Valid cases: 135115
Minimum: NaN
Maximum: NaN

Description

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

ObservationDate (ObservationDate)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: character

Valid cases: 135115
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: ZA021.CMD2016.v1

Overview

Type: Continuous
Format: numeric
Decimals: 0
Range: 1-6607

Valid cases: 135115
Invalid: 0
Minimum: 1
Maximum: 7964
Mean: 4005
Standard deviation: 2316.8

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: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0

Valid cases: 12942
Invalid: 122173

Description

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

DeliveryId (DeliveryId)
File: ZA021.CMD2016.v1

Overview

Type: Discrete
Format: numeric
Decimals: 0

Valid cases: 12942
Invalid: 122173

Description

The RecNr of the delivery event associated with this birth