South Africa - Agincourt INDEPTH Core Dataset 1992-2012 Release 2015

Professor Stephen Tollman - University of the Witwatersrand, Professor Kathy Khan - University of the Witwatersrand, Dr Mark Collinson - University of the Witwatersrand, Professor Samuel Clar

Report generated on: July 27, 2015

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

Overview

Identification

ID NUMBER
INDEPTH.ZA011.CMD2012.V2

Version

VERSION DESCRIPTION
CMD2012.V2: Updated dataset for public distribution.

PRODUCTION DATE
2015-06-09

Overview

ABSTRACT
The Agincourt health and socio-demographic surveillance system (HDSS), located in rural northeast South Africa close to the Mozambique border, was established in 1992 to support district health systems development led by the post-apartheid ministry of health. The HDSS (90 000 people), based on an annual update of resident status and vital events, now supports multiple investigations into the causes and consequences of complex health, population and social transitions. Observational work includes cohorts focusing on different stages along the life course, evaluation of national policy at population, household and individual levels and examination of household responses to shocks and stresses and the resulting pathways influencing health and well-being. Trials target children and adolescents, including promoting psycho-social well-being, preventing HIV transmission and reducing metabolic disease risk. Efforts to enhance the research platform include using automated measurement techniques to estimate cause of death by verbal autopsy, full 'reconciliation' of in- and outmigrations, follow-up of migrants departing the study area, recording of extra-household social connections and linkage of individual HDSS records with those from sub-district clinics. Fostering effective collaborations (including INDEPTH multi-centre work in adult health and ageing and migration and urbanization), ensuring cross site compatibility of common variables and optimizing public access to HDSS data are priorities.

KIND OF DATA
Event history data

UNITS OF ANALYSIS
Individual

Scope

NOTES
Individual : Residence episodes and Residence Initiating and Terminating Events

TOPICS

Topic	Vocabulary	URI
Demography [N01.224]	MesSH	http://www.ncbi.nlm.nih.gov/mesh
Population [N01.600]	MeSH	http://www.ncbi.nlm.nih.gov/mesh

KEYWORDS
Fertility, Mortality, Migration, ADSS, AHDSS, Migrants, SES, Asset, HIV, Causes of Death, dataset, Surveillance, Demographic, Health, death, survey, wits, Agincourt, MRC, Public Health, Health Sciences, Demography, Data, IShare, Indepth

Coverage

GEOGRAPHIC COVERAGE

Demographic Surveillance Survey

UNIVERSE
Currently resident household members including temporary migrants approximately 90,000 indivudals in approximately 16 000 households

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Professor Stephen Tollman	University of the Witwatersrand
Professor Kathy Khan	University of the Witwatersrand
Dr Mark Collinson	University of the Witwatersrand
Professor Samuel Clark	University of the Witwatersrand

OTHER PRODUCER(S)

Name	Affiliation	Role
Xavier Gomez-Olive	University of the Witwatersrand	Research Manager
Sulaimon Afolabi	University of the Witwatersrand	Data Manager
Nkosinathi masilelea	University of the Witwatersrand	Data Manager
Cho Kabudula	University of the Witwatersrand	Data Manager
Itayi Adam	University of the Witwatersrand	System Administrator
Pleasure Ngobeni	University of the Witwatersrand	Data Manager
Bernard Silaule	University of the Witwatersrand	Field Project Manager
Ryan Wagner	University of the Witwatersrand	Census Project Manager
Rirhandzu Mnisi	University of the Witwatersrand	Data Entry Team leader

FUNDING

Name	Abbreviation	Role
Wellcome Trust		
The National Institute of Ageing NIH		
Andrew W Mellon Foundation		
The William and Flora Hewlett Foundation		
South African Medical Research Council		

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Rhian Twine	Agincourt HDSS	LINC
Ngoni	Agincourt HDSS	Senior Administrator

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role

Name	Abbreviation	Affiliation	Role
Agincourt HDSS Data Section	ZA011	Agincourt HDSS Centre, University of the Witwatersrand	Documentation of the Study
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical Support
INDEPTH Network	int.indepth	INDEPTH	Agency

DATE OF METADATA PRODUCTION
2015-07-27

DDI DOCUMENT VERSION
Version 1 (June 2015)

Version 2 (July 2015)

DDI DOCUMENT ID
DDI.INDEPTH.ZA011.CMD2012.v2

Name	Abbreviation	Affiliation	Role
Agincourt HDSS Data Section	ZA011	Agincourt HDSS Centre, University of the Witwatersrand	Documentation of the Study
iSHARE2 Technical Team	isTT	INDEPTH Network	Technical Support
INDEPTH Network	int.indepth	INDEPTH	Agency

Sampling

Sampling Procedure

No sampling

Deviations from Sample Design

Not applicable

Response Rate

On an average around 99 % response over the years for each round.

Weighting

Not Applicable

Questionnaires

Overview

Full details can be foud on the link http://www.agincourt.co.za/index.php/data/data-collection-instruments/

Data Collection

Data Collection Dates

Start	End	Cycle
1992-03-16	1992-10-25	1
1993-10-25	1994-04-21	2
1995-06-10	1995-12-13	3
1996-07-15	1998-08-16	4
1999-07-12	1999-11-30	5
2000-07-19	2000-11-30	6
2001-08-01	2001-12-13	7
2002-08-01	2003-01-31	8
2003-07-01	2004-01-31	9
2004-08-01	2004-12-11	10
2005-08-01	2005-12-31	11
2006-08-01	2006-12-31	12
2007-08-01	2008-06-30	13
2008-08-01	2008-12-31	14
2009-08-03	2009-12-31	15
2010-08-02	2010-12-31	16
2011-08-01	2011-12-31	17
2012-08-01	2012-08-01	18

Time Periods

Start	End	Cycle
1992-03-01	2012-12-31	N/A

Data Collection Mode

Proxy Respondent [proxy]

DATA COLLECTION NOTES

Minimum of 2 weeks training for field team every year adiditonal training for supervisors.

Inteviews in the the repondents home language Xitsonga (Shangaan)

Community Liaison Office (LINC office) works with Community advisory grroup and provides community feedback in each village after each census round

Data Collectors

Name	Abbreviation	Affiliation
Agincourt HDSS	AHDSS	University of the Witwatersrand

SUPERVISION

Interviews were conducted by teams of interviewers. Each interviewing team comprises of 5 -6 individuals a supervisor, and a driver.

The role of the supervisor was to coordinator field data collection activities, including management of the field teams, supplies and equipment, 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 supervisor was responsible for reviewing each questionnaire at the end of the day, checking for missed questions, skip errors, fields incorrectly completed, and checking for inconsistencies in the data. The supervisor also observed interviews and conducted review sessions with interviewers.

The Central Quality Control team reviewed the questionnaires prior to data entry and if necessary sent documents back to

the the field for correction

Full details of the fieldwork process can be found in the Filedworker guidelines document
http://www.agincourt.co.za/wp-content/uploads/2012/10/AGN-zFWGDL_Guideline.pdf

Data Processing

Data Editing

Data checks were acrried out after data entry using SQL queries of the production database. Inconsistencies spotted at this stage were resolved wither by reference to the daabase or by additional filed visits.

Other Processing

The dwellings list was derived from the previous year's census. New dwellings were added and empty dwellings closed each year

A document tracking system is used to maintain a chain of custody for the questionairres

Data Entry was carried using a custom developed front end and a relational database (In SQL-Server since 2003)

Data was extracted using T-SQL queries in SQL Server and processed using Pentaho ETL

Data Appraisal

Estimates of Sampling Error

Not applicable

Other forms of Data Appraisal

CentreId MetricTable QMetric Illegal Legal Total Metric RunDate
ZA011 MicroDataCleaned Starts 221047 2015-06-09 14:24
ZA011 MicroDataCleaned Transitions 0 448794 448794 0. 2015-06-09 14:24
ZA011 MicroDataCleaned Ends 221047 2015-06-09 14:24

File Description

Variable List

ZA011.CMD2012.V2

Content	The file contains raw data on core demogrpahic events collected in the Agincourt HDSS between 1992 and 2010. It contains the INDEPTH core microdataset produced using ETL through Pentaho KETL
Cases	589756
Variable(s)	14
Structure	Type: Keys: ()
Version	ZA011.CMD2012.V2
Producer	
Missing Data	NULL is used for missing values.

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: ZA011.CMD2012.V2

Overview	
Type: Continuous	Valid cases: 589756
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-395414	Maximum: 603962
	Mean: 302022
	Standard deviation: 174351.8

Description

A sequential number uniquely identifying each record in the data file

CountryId (CountryId)
File: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 589756
Format: numeric	Invalid: 0
Decimals: 0	
Range: 710-710	

Description

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

CentreId (CentreId)
File: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 589756
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: ZA011.CMD2012.V2

Overview	
Type: Continuous	Valid cases: 589756
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-194358	Maximum: 221055
	Mean: 110501.8
	Standard deviation: 63832.2

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: ZA011.CMD2012.V2

Sex (Sex)

File: ZA011.CMD2012.V2

Overview

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

Valid cases: 589756
Invalid: 0

Description

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

DoB (DoB)

File: ZA011.CMD2012.V2

Overview

Type: Discrete
Format: character
Width: 10

Valid cases: 589756
Minimum: NaN
Maximum: NaN

Description

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

EventCount (EventCount)

File: ZA011.CMD2012.V2

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 2-10

Valid cases: 589756
Invalid: 0

Description

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

EventNr (EventNr)

File: ZA011.CMD2012.V2

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 1-10

Valid cases: 589756
Invalid: 0

Description

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

EventCode (EventCode)

File: ZA011.CMD2012.V2

Overview

Type: Discrete
Format: character
Width: 3

Valid cases: 589756
Invalid: 0

Description

A code identifying the type of event that has occurred.

EventDate (EventDate)

File: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 589756
Format: character	Minimum: NaN
Width: 10	Maximum: NaN

Description

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

ObservationDate (ObservationDate)

File: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 589756
Format: character	Minimum: NaN
Width: 10	Maximum: NaN

Description

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

LocationId (LocationId)

File: ZA011.CMD2012.V2

Overview	
Type: Continuous	Valid cases: 589756
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-20325	Maximum: 21721
	Mean: 10820.1
	Standard deviation: 6241.8

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: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 79314
Format: numeric	Invalid: 510442
Decimals: 0	

Description

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

DeliveryId (DeliveryId)

File: ZA011.CMD2012.V2

Overview	
Type: Discrete	Valid cases: 79314
Format: numeric	Invalid: 510442
Decimals: 0	

DeliveryId (DeliveryId)
File: ZA011.CMD2012.V2

Description

The RecNr of the delivery event associated with this birth