

# South Africa - Agincourt INDEPTH Core Dataset 1993-2017 (Release 2019)

**Professor Stephen Tollman - University of the Witwatersrand**  
**Professor Kathleen Khan - University of the Witwatersrand**  
**Dr Mark Collinson - University of the Witwatersrand**  
**Professor Samuel C**

Report generated on: September 24, 2019

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

## Overview

### Identification

---

#### ID NUMBER

INDEPTH.ZA011.CMD2017.v1

### Version

---

#### VERSION DESCRIPTION

ZA011.CMD2017.v1: Edited dataset for public distribution

#### PRODUCTION DATE

2019-06-24

#### NOTES

ZA011.CMD2017.v1

## 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	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>

Topic	Vocabulary	URI
Population [N01.600]	MeSH	<a href="http://www.ncbi.nlm.nih.gov/mesh">http://www.ncbi.nlm.nih.gov/mesh</a>

## 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

The Agincourt HDSS covers an area of 450 square kilometer comprising a sub-district of 31 villages with traditional and elected leadership. Since the democratic transition in 1994, infrastructure development has proceeded but at a rate below expectations: electricity is available in all villages, but the cost is too high for many households; few gravel roads have been tarred within the sub-district; a dam was constructed nearby, but to-date, there is no piped water to dwellings, and sanitation is rudimentary. Every village has a primary school and most a high school; however, the quality of education remains poor.<sup>14</sup> Although almost all children enrol, educational progress is often delayed with few post-secondary opportunities. The area is dry in winter (from May to October), with soil more suited to game farming than agriculture. Households generally purchase maize and other foods, supplementing this with home-grown crops and collection of wild foods.<sup>15</sup> South Africa's non-contributory social grant system is a vital source of household income, notably the old age pension<sup>16</sup> and child support grant.<sup>17</sup><sup>18</sup> There are two health centres and six clinics within the sub-district, with three district hospitals 25-60km away.

## UNIVERSE

Currently resident household members including temporary migrants approximately 117,000 individuals in approximately 22,000 households

## Producers and Sponsors

## PRIMARY INVESTIGATOR(S)

Name	Affiliation
Professor Stephen Tollman	University of the Witwatersrand
Professor Kathleen 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
Chodziwadziwa Kabudula	University of the Witwatersrand	Data Manager
Itayi Adam	University of the Witwatersrand	System Administrator
Rirhandzu Mnisi	University of the Witwatersrand	Data Entry Team leader
Bernard Silaule	University of the Witwatersrand	Field Project Manager
Ryan Wagner	University of the Witwatersrand	Census Project Manager
Daniel Ohene-Kwofie	University of the Witwatersrand	Data Manager

## FUNDING

Name	Abbreviation	Role
------	--------------	------

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
Agincourt Community Members		

## Metadata Production

## METADATA PRODUCED BY

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

## DATE OF METADATA PRODUCTION

2019-06-24

## DDI DOCUMENT VERSION

ZA011.CMD2017.v1: Produced in June 2019

## DDI DOCUMENT ID

DDI.INDEPTH.ZA011.CMD2017.v1

## Sampling

### **Sampling Procedure**

---

Not Applicable

### **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 found on the link <http://www.agincourt.co.za/index.php/data/data-collection-instruments/>

## Data Collection

### Data Collection Dates

---

Start	End	Cycle
1993-01-01	2017-12-31	Release Coverage

### Time Periods

---

Start	End	Cycle
1992-03-01		Round 1
1993-10-25		Round 2
1995-06-10		Round 3
1996-07-15		Round 4
1999-07-12		Round 5
2000-07-19		Round 6
2001-08-01		Round 7
2002-08-01		Round 8
2003-07-01		Round 9
2004-08-01		Round 10
2005-08-01		Round 11
2006-08-01		Round 12
2007-08-01		Round 13
2008-08-01		Round 14
2009-08-03		Round 15
2010-08-02		Round 16
2011-08-01		Round 17
2012-08-01		Round 18
2013-08-01		Round 19
2014-08-01		Round 20
2015-08-01		Round 21
2016-08-01		Round 22
2017-08-01		Round 23

### Data Collection Mode

---

Proxy Respondent [proxy]

### Data Collection Notes

---

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

Interviews in the the repondents home language Xitsonga (Shangaan)

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

### Questionnaires

---

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

### 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 coordinate 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](http://www.agincourt.co.za/wp-content/uploads/2012/10/AGN-zFWGDL_Guideline.pdf)

## Data Processing

### Data Editing

---

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

### Other Processing

---

The dwelling 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 questionnaires

Data Entry was carried using a custom developed front end and a relational database ( in Microsoft 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 271500 2019-06-24 08:16

ZA011 MicroDataCleaned Transitions 0 549682 549682 0. 2019-06-24 08:16

ZA011 MicroDataCleaned Ends 271500 2019-06-24 08:16

ZA011 MicroDataCleaned SexValues 78 549604 549682 0. 2019-06-24 08:16

ZA011 MicroDataCleaned DoBValues 36 549646 549682 0. 2019-06-24 08:16

# File Description

# Variable List

**ZA011.CMD2017.v1**

Content	The file contains raw data on core demographic events collected in the Agincourt HDSS between 1993 and 2017. It contains the INDEPTH core microdataset produced using ETL through Pentaho KETTLE
Cases	738159
Variable(s)	14
Structure	Type: Keys: ()
Version	ZA011.CMD2017.v1
Producer	Agincourt HDSS
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.CMD2017.v1

### Overview

Type: Continuous	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-395414	Maximum: 738159
	Mean: 369080
	Standard deviation: 213088.3

### Description

A sequential number uniquely identifying each record in the data file

## CountryId (CountryId)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
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.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
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.CMD2017.v1

### Overview

Type: Continuous	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-194358	Maximum: 271507
	Mean: 135734
	Standard deviation: 78385.1

### 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.CMD2017.v1

## Sex (Sex)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-2	

### Description

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

## DoB (DoB)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: character	Minimum: NaN
Width: 10	Maximum: NaN

### Description

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

## EventCount (EventCount)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	
Range: 2-10	

### Description

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

## EventNr (EventNr)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-10	

### Description

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

## EventCode (EventCode)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: character	Invalid: 0
Width: 3	

### Description

## EventCode (EventCode)

File: ZA011.CMD2017.v1

A code identifying the type of event that has occurred.

## EventDate (EventDate)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
Format: character	Minimum: NaN
Width: 10	Maximum: NaN

### Description

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

## ObservationDate (ObservationDate)

File: ZA011.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 738159
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.CMD2017.v1

### Overview

Type: Continuous	Valid cases: 738159
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-20325	Maximum: 27434
	Mean: 13663.8
	Standard deviation: 7935.4

### 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.CMD2017.v1

### Overview

Type: Discrete	Valid cases: 98388
Format: numeric	Invalid: 639771
Decimals: 0	

### Description

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

# DeliveryId (DeliveryId)

## File: ZA011.CMD2017.v1

### Overview

Type: Discrete  
Format: numeric  
Decimals: 0

Valid cases: 98388  
Invalid: 639771

### Description

The RecNr of the delivery event associated with this birth

