

# South Africa - AHRI.Education Cleaned Dataset

**Nompumelelo Mkwazi**

Report generated on: February 18, 2026

Visit our data catalog at: <https://data.ahri.org/index.php>

## Identification

### SURVEY ID NUMBER

AHRI.EducationCleaned

### TITLE

AHRI.Education Cleaned Dataset

### COUNTRY

Name	Country code
South Africa	ZA

### ABSTRACT

Africa Health Research Institute (AHRI) is a multidisciplinary, independent research institute. AHRI's goal is to become a source of fundamental discoveries into the susceptibility, transmission and cure of HIV, TB and related diseases. We also seek ways to improve diagnosis, prevention and treatment. To achieve this, we bring together leading researchers from different fields, use cutting edge science to improve people's health, and help to train the next generation of African scientists.

AHRI has conducted individual and household demographic surveillance since 2000. Demographic surveillance data are processed for longitudinal analyses by organizing the demographic surveillance data into individual and household residency episodes at a geographical physical location, and individual membership episodes within a household. Start events of residency and memberships include enumeration, birth, in-migration and relocating into a household from within the study population. Exit events of residency and memberships include death (by cause), out-migration and relocating to another location in the study population. AHRI annually produces a suite of Individual Demographic surveillance datasets which contains detailed records of all individuals under surveillance since January 2000. The datasets detail individual residency and membership patterns, household mortality (including cause of death), fertility, migration, HIV status, health care utilization, and household socio-economic profile.

Each record in the dataset represents a period of observation for an individual during which all the recorded characteristics of the individual stay constant. For example, on the birthday of the individual a new episode will start, because the age of the individual has changed. An out-migration will result in a new episode, because the location or residential status has changed. Any change in one of the status values, such as education, marital status, social-economic status and HIV will likewise result in a new episode on the date of the change.

The following are the versions of the Demographic surveillance datasets:

CoresidencyEpisodes  
 EducationCleaned  
 EducationStatuses  
 HIVStatusEpisodes  
 HouseholdAssets  
 HouseholdMap  
 HouseholdObservations  
 HouseholdSocioEconomicStatus  
 HouseholdStatusEpisodes  
 IndividualHSEEpisodes  
 IndividualMap  
 LabourEpisodes  
 LabourStatuses  
 LocationMap  
 MaritalStatuses  
 PartnerEpisodes  
 SurveillanceEpisodesBasic  
 SurveillanceEpisodesYrAge  
 SurveillanceEpisodes YrAge\_Coresidency  
 SurveillanceEpisodesYrAge\_HIV  
 SurveillanceEpisodes VrAge\_HSE  
 SurveillanceEpisodes YrAge\_Labour

SurveillanceEpisodesYrAge\_Partners  
 SurveillanceEpisodesYrAgeDel  
 SurveillanceEpisodes VrAgeDel\_HSE  
 SurveillanceEpisodesYrAgeDeL\_HSE\_Labour  
 SurveillanceEpisodesYrAgeDeL\_HSE\_Labour\_Partner\_Parents  
 SurveillanceEpisodes VrAgeDel\_HSE\_Labour\_Partner\_Parents\_HIV  
 SurveillanceEpisodesYrAgeDeL\_HSE\_Labour\_Partners

## KIND OF DATA

Event History Data

## UNIT OF ANALYSIS

Episodes of exposure

## Version

---

## VERSION DESCRIPTION

v2.0.0

## Scope

---

## TOPICS

Topic	Vocabulary	URI
Education Cleaned	Africa Health Research Institute	www.ahri.org

## KEYWORDS

Keyword	Vocabulary	URI
Fertility, Mortality, Migration, Social-economic status, HIV, ART	Africa Health Research Institute	www.ahri.org

## Coverage

---

## GEOGRAPHIC COVERAGE

AHRI is situated in the south-east portion of the uMkhanyakude district of KwaZulu-Natal province near the town of Mtubatuba. It is bounded on the west by the Umfolozi-Hluhluwe nature reserve, on the south by the Umfolozi river, on the east by the N2 highway (except form portions where the KwaMsane township strangles the highway) and in the north by the provincial road R22. The surveillance area is approximately 845km<sup>2</sup> in size.

## UNIVERSE

Households resident in dwellings within the study area are eligible for inclusion in the household component of AHRI Demographic Surveillance. All individuals identified by the household proxy informant as members of the households are enumerated. A resident household member is an individual that intends to sleep majority of time at the dwelling occupied by the household over a four-month period. Households include resident and non-resident members. An individual is a non-resident member if they have close ties to the household, but do not physically reside with the household at the dwelling most of the time. Non-resident members can also be called temporary migrants and they are enumerated within the household list. Because household membership is not tied to physical residency, an individual may be a member of more than one household.

## Producers and sponsors

---

## PRIMARY INVESTIGATORS

Name	Affiliation
Dr. Kobus Herbst	AHRI, SAPRIN

Dr. Guy Harling	AHRI
Dr. Mark Siedner	AHRI
Prof Willem Hanekom	AHRI

## PRODUCERS

Name	Abbreviation
Africa Health Research Institute	AHRI

## FUNDING AGENCY/SPONSOR

Name	Abbreviation	Role
Wellcome Trust	WT	Core funding
SAPRIN	SAPRIN	

## OTHER IDENTIFICATIONS/ACKNOWLEDGMENTS

Name	Affiliation	Role
AHRI Research Data Management Team	AHRI	Research Data Management
AHRI Data Collection Team	AHRI	Data Collection
AHRI Community Engagement Team	AHRI	Community Engagement

## Sampling

## SAMPLING PROCEDURE

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

## Data collection

## DATES OF DATA COLLECTION

Start	End
2000-01-01	2024-12-31

## Data Processing

## DATA EDITING

Data collected in this study was centrally stored on AHRI's secure database servers. Data was collected through interviews at the call centre on central server using Survey Solutions, and in the field on paper and encrypted tablet computers using REDCap, and then securely uploaded to a central server. Database rules and constraints were enforced on the master database to ensure that only valid data was stored. Personally identifiable data was restricted to separate database tables accessible only to data management staff .

## Access policy

## ACCESS CONDITIONS

Access to the data requires accurate completion of the online data access application form accessible on the AHRI Data repository(<<https://data.ahri.org/>>). Data users are required to abide by the data use conditions stipulated on the application for access to the data. Failure to do so may result in their data access privileges being revoked by the Data Custodian. In order to recognise the effort and intellectual contributions of AHRI investigators in producing and curating the

data, users of AHRI data must acknowledge the source of the data and abide by the terms and conditions under which the data is accessed and must cite the dataset in publication using the citation provided as part of this documentation. All analytical datasets published on the AHRI Data Repository are assigned digital object identifier (DOIs) and the DOIs can be found on the Data Repository under Study Description tab - Access policy. AHRI data users are required to always cite the dataset using the relevant DOI.

<https://github.com/SAPRIN-MRC/SaprinEpisodes.jl.git>

#### CITATION REQUIREMENTS

Herbst, K., Harling, G., Siedner, M., & Hanekom, W. (2026). AHRI.Education Cleaned Dataset [Data set]. Africa Health Research Institute.

DOI:<https://doi.org/10.23664/AHRI.EDUCATIONCLEANED>

## Metadata production

---

#### DDI DOCUMENT ID

DDI.AHRI.EducationCleaned

#### PRODUCERS

Name	Abbreviation
Africa Health Research Institute	AHRI

**Data Dictionary**

<b>Data file</b>	<b>Cases</b>	<b>Variables</b>
<b>EducationCleaned</b>	3046012	6



**Data file: EducationCleaned**

Cases: 3046012

Variables: 6

**Variables**

ID	Name	Label	Question
V296	IndividualId	Individual ID	
V297	Year	Year in which education level was attained	
V298	DoB	Date of birth of individual	
V299	Age	Age attained during this year	
V300	EducationStatus	School level attained in this year prior to cleaning	
V301	CleanedEducation	School level attained in this year after to cleaning	

Total: 6



### INDIVIDUALID: Individual ID

Data file: EducationCleaned

#### Overview

Valid: 3046012 Invalid: 0 Minimum: 1 Maximum: 263153 Mean: 119070.693 Standard deviation: 77552.346  
 Type: Continuous Decimal: 0 Width: 9 Range: 1 - 263153 Format: Numeric

### YEAR: Year in which education level was attained

Data file: EducationCleaned

#### Overview

Valid: 3046012 Invalid: 0 Minimum: 2000 Maximum: 2024 Mean: 2013.473 Standard deviation: 7.308  
 Type: Continuous Decimal: 0 Width: 9 Range: 2000 - 2024 Format: Numeric

### DOB: Date of birth of individual

Data file: EducationCleaned

#### Overview

Valid: 3046004 Minimum: 1793-01-01 Maximum: 2024-12-31  
 Type: Discrete Width: 11 Range: - Format: character

### AGE: Age attained during this year

Data file: EducationCleaned

#### Overview

Valid: 3046012 Invalid: 0 Minimum: 0 Maximum: 1500 Mean: 25.58 Standard deviation: 18.777  
 Type: Continuous Decimal: 0 Width: 9 Range: 0 - 1500 Format: Numeric

### EDUCATIONSTATUS: School level attained in this year prior to cleaning

Data file: EducationCleaned

#### Overview

Valid: 1737652 Invalid: 1308360  
 Type: Discrete Decimal: 0 Width: 18 Range: -1 - 99 Format: Numeric

#### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
-1	Unknown	86515	5%
0	Grade 0	31894	1.8%

1	Grade 1	59802	3.4%
2	Grade 2	66632	3.8%
3	Grade 3	71886	4.1%
4	Grade 4	80926	4.7%
5	Grade 5	69139	4%
6	Grade 6	73859	4.3%
7	Grade 7	96034	5.5%
8	Grade 8	92848	5.3%
9	Grade 9	102759	5.9%
10	Grade 10	147361	8.5%
11	Grade 11	173666	10%
12	Grade 12	484021	27.9%
13	Special School	0	0%
21	Higher Certificate	0	0%
22	National Diploma	0	0%
23	Degree	0	0%
24	Honours	0	0%
25	Masters	0	0%
26	Doctors	0	0%
31	ABET 1	0	0%
32	ABET 2	0	0%
33	ABET 3	0	0%
34	ABET 4	0	0%
41	FET 1	0	0%
42	FET 2	0	0%
43	FET 3	0	0%
50	LT 1 year	0	0%
51	Creche	0	0%
96	Refused	0	0%
97	NAD	0	0%
98	No education	100310	5.8%
99	Not attending	0	0%
Sysmiss		1308360	

## CLEANEDEDUCATION: School level attained in this year after to cleaning

Data file: EducationCleaned

### Overview

Valid: 2195404 Invalid: 850608

Type: Discrete    Decimal: 0    Width: 18    Range: -1 - 99    Format: Numeric

**Questions and instructions**

## CATEGORIES

<b>Value</b>	<b>Category</b>	<b>Cases</b>	
-1	Unknown	38468	1.8%
0	Grade 0	32154	1.5%
1	Grade 1	67377	3.1%
2	Grade 2	87930	4%
3	Grade 3	89473	4.1%
4	Grade 4	114764	5.2%
5	Grade 5	85390	3.9%
6	Grade 6	92041	4.2%
7	Grade 7	122007	5.6%
8	Grade 8	119479	5.4%
9	Grade 9	113892	5.2%
10	Grade 10	180246	8.2%
11	Grade 11	182734	8.3%
12	Grade 12	696221	31.7%
13	Special School	0	0%
21	Higher Certificate	0	0%
22	National Diploma	0	0%
23	Degree	0	0%
24	Honours	0	0%
25	Masters	0	0%
26	Doctors	0	0%
31	ABET 1	0	0%
32	ABET 2	0	0%
33	ABET 3	0	0%
34	ABET 4	0	0%
41	FET 1	0	0%
42	FET 2	0	0%
43	FET 3	0	0%
50	LT 1 year	0	0%
51	Creche	0	0%
96	Refused	0	0%
97	NAD	0	0%
98	No education	173228	7.9%
99	Not attending	0	0%

---

Sysmiss		850608	
---------	--	--------	--

---

## Download related resources

### Technical documents

#### DDI:Education Cleaned

---

Title	DDI:Education Cleaned
Author(s)	Nompumelelo Mkwanazi
Date	2026-02-18
Country	South Africa
Language	English
Publisher(s)	Nompumelelo Mkwanazi
Filename	Education Cleaned.pdf

---