

South Africa

Dr Kobus Herbst, AHRI, SAPRIN

Dr Guy Harling, AHRI

Dr Mark Siedner, AHRI

Prof Willem Hanekom, AHRI

AHRI.Labour Statuses

Study Documentation

February 19, 2026

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.AHRI.LabourStatuses

Table of Contents

Overview	4
Scope & Coverage	5
Producers & Sponsors	5
Sampling	5
Data Collection	6
Data Processing & Appraisal	6
Accessibility	6
Files Description	7
LabourStatuses	7
Variables List	8
LabourStatuses	8
Variables Description	9
LabourStatuses	10

AHRI.Labour Statuses

Overview

Identification	AHRI.LabourStatuses
Version	v2.0.0

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
 SurveillanceEpisodesYTAge
 SurveillanceEpisodes YrAge_Coresidency
 SurveillanceEpisodes YrAge_HIV
 SurveillanceEpisodes VrAge_HSE
 SurveillanceEpisodes YrAge_Labour
 SurveillanceEpisodes YrAge_Partners
 SurveillanceEpisodes YrAgeDel

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

Scope & Coverage	
Keywords	Fertility, Mortality, Migration, Social-economic status, HIV, ART
Topics	Labour Statuses
Time Period(s)	2000-2024
Countries	South Africa
Geographic Coverage	
<p>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 845km² in size.</p>	
Universe	
<p>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.</p>	

Producers & Sponsors	
Primary Investigator(s)	Dr Kobus Herbst, AHRI, SAPRIN Dr Guy Harling, AHRI Dr Mark Siedner, AHRI Prof Willem Hanekom, AHRI
Other Producer(s)	Africa Health Research Institute (AHRI)
Funding Agency/ies	Wellcome Trust (WT) , Core funding SAPRIN (SAPRIN)
Other Acknowledgment(s)	AHRI Research Data Management Team , Research Data Management , AHRI AHRI Data Collection Team , Data Collection , AHRI AHRI Community Engagement Team , Community Engagement , AHRI

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

Data Collection

Data Collection Dates	start 2000-02-22 end 2024-12-17
------------------------------	------------------------------------

Data Processing & Appraisal

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 .

Accessibility

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.Labour Statuses [Data set]. Africa Health Research Institute.

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

Files Description

Dataset contains 1 file(s)

LabourStatuses	
# Cases	1176085
# Variable(s)	7

Variables List

Dataset contains 7 variable(s)

File LabourStatuses							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	Individu..	Individual ID	continuous	numeric-9.0	1176085	0	-
2	Observat..	Observation Date	discrete	character-11	1176085	-	-
3	CurrentE..	Current Employment	discrete	numeric-63.0	1176085	0	-
4	Employme..	Employment Sector	discrete	numeric-35.0	1176085	0	-
5	Employme..	Employment Type	discrete	numeric-23.0	1176085	0	-
6	Employer	Employer	discrete	numeric-42.0	1176085	0	-
7	Unemploy..	Reason/s for unemployment	discrete	numeric-23.0	772210	403875	-

Variables Description

Dataset contains 7 variable(s)

File : LabourStatuses

IndividualId: Individual ID

Information [Type= continuous] [Format=numeric] [Range= 2-263153] [Missing=*]

Statistics [NW/ W] [Valid=1176085 /-] [Invalid=0 /-] [Mean=134200.15 /-] [StdDev=75271.889 /-]

ObservationDate: Observation Date

Information [Type= discrete] [Format=character] [Missing=*]

Statistics [NW/ W] [Valid=1176085 /-]

CurrentEmployment: Current Employment

Information [Type= discrete] [Format=numeric] [Range= 0-202] [Missing=*]

Statistics [NW/ W] [Valid=1176085 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Unknown employment status	89969	7.6%
1	Currently employed	331305	28.2%
2	Currently part-time employed	49067	4.2%
3	Pensioner	0	
4	Not looking for work	0	
5	Other	0	
10	Unemployed due to caring for others/household duties	0	
11	Unemployed due to caring for others/household duties	0	
12	Unemployed and looking for a job	0	
13	Unemployed but in school or undergoing other education/training	106937	9.1%
14	Unemployed and unable to work due to illness/disability	0	
15	Not currently in employment	598807	50.9%
202	Work for themselves	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

EmploymentSector: Employment Sector

Information [Type= discrete] [Format=numeric] [Range= 0-306] [Missing=*]

Statistics [NW/ W] [Valid=1176085 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Unknown employment sector	0	
100	Unknown employment sector	1022906	87.0%
101	Agriculture, Fishing, or Forestry	7301	0.6%
102	Mining	3780	0.3%
103	Manufacturing	14214	1.2%
104	Electricity and water	3233	0.3%
105	Construction	17919	1.5%
106	Wholesale/retail	14190	1.2%
107	Restaurant/Hotels/Tourism/Sport	6249	0.5%
108	Transport and communication	9821	0.8%
109	Financial services	2032	0.2%
110	Educational services	11564	1.0%
111	Health services	5672	0.5%
112	Legal services	2965	0.3%
113	Research	1285	0.1%

File : LabourStatuses

EmploymentSector: Employment Sector

Value	Label	Cases	Percentage
114	Domestic services	15641	1.3%
115	Armed forces and police	9213	0.8%
116	Informal sector, e.g. street vendor	28100	2.4%
301	National/Central government	0	
305	Private sector employer	0	
306	Non-profit institution	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

EmploymentType: Employment Type

Information	[Type= discrete] [Format=numeric] [Range= 0-203] [Missing=*]
Statistics [NW/ W]	[Valid=1176085 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Unknown employment type	778	0.1%
200	Unknown employment type	1031288	87.7%
201	Works as employee	135376	11.5%
202	Work for themselves	5046	0.4%
203	Do odd jobs/piece jobs	3597	0.3%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

Employer: Employer

Information	[Type= discrete] [Format=numeric] [Range= 0-308] [Missing=*]
Statistics [NW/ W]	[Valid=1176085 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Unknown Employer	816	0.1%
201	Works as employee	0	
300	Unknown Employer	1033195	87.9%
301	National/Central government	6435	0.5%
302	Provincial government	13632	1.2%
303	Local / district authority or municipality	3820	0.3%
304	Public Corporation	11350	1.0%
305	Private sector employer	95523	8.1%
306	Non-profit institution	4616	0.4%
307	Self-employment	3139	0.3%
308	Another household member	3559	0.3%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

Unemployment: Reason/s for unemployment

Information	[Type= discrete] [Format=numeric] [Range= 0-8480] [Missing=*]
Statistics [NW/ W]	[Valid=772210 /-] [Invalid=403875 /-]

Value	Label	Cases	Percentage
0	Unknown	17856	2.3%
1	Studying	116010	15.0%
2	Looking for work	236879	30.7%
3		2	0.0%

File : LabourStatuses

Unemployment: Reason/s for unemployment

Value	Label	Cases	Percentage
4	Retired	39479	5.1%
8	Sick/injured	16958	2.2%
16	Pregnant/Own Child Care	14778	1.9%
32	Other's Child Care	9362	1.2%
64	Care for sick/injured	1266	0.2%
128	Retrenched	5339	0.7%
256	Nothing (Not looking)	295571	38.3%
512	Other	8175	1.1%
1024	Don't know	9632	1.2%
2048	Refused	694	0.1%
4096	Seasonal	8	0.0%
8192	Disabled	200	0.0%
8480		1	0.0%
Sysmiss		403875	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.