

South Africa

Dr Kobus Herbst, AHRI, SAPRIN

Dr Guy Harling, AHRI

Dr Mark Siedner, AHRI

Prof Willem Hanekom, AHRI

AHRI.HIV Status Episodes

Study Documentation

February 19, 2026

Metadata Production

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

Table of Contents

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

AHRI.HIV Status Episodes

Overview

Identification	AHRI.HIVStatus.Episodes
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	HIV Status Episodes
Time Period(s)	2001-2024
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 850km2 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 2001-02-21 end 2024-12-31
------------------------------	------------------------------------

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.HIV Status Episodes [Data set]. Africa Health Research Institute.

DOI:<https://doi.org/10.23664/AHRI.HIVSTATUS.EPISODES>

Files Description

Dataset contains 1 file(s)

HIVStatusEpisodes	
# Cases	191045
# Variable(s)	7

Variables List

Dataset contains 7 variable(s)

File HIVStatusEpisodes							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	Individu..	Individual ID	continuous	numeric-9.0	191045	0	-
2	StartDate	Start Date (inclusive)	discrete	character-11	191045	-	-
3	EndDate	End Date (inclusive)	discrete	character-11	191045	-	-
4	HIVStatus	HIV status during this episode	discrete	numeric-13.0	191045	0	-
5	HIVStatu..	HIV Status changed at the start of this episode	discrete	numeric-9.0	191045	0	-
6	HIVSeroc..	Individual seroconverted at the start of this episode	discrete	numeric-9.0	191045	0	-
7	Episode	Sequential episode for individual	discrete	numeric-9.0	191045	0	-

Variables Description

Dataset contains 7 variable(s)

File : HIVStatusEpisodes

IndividualId: Individual ID

Information [Type= continuous] [Format=numeric] [Range= 2-263149] [Missing=*]

Statistics [NW/ W] [Valid=191045 /-] [Invalid=0 /-] [Mean=125049.928 /-] [StdDev=77011.482 /-]

StartDate: Start Date (inclusive)

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

Statistics [NW/ W] [Valid=191045 /-]

EndDate: End Date (inclusive)

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

Statistics [NW/ W] [Valid=191045 /-]

HIVStatus: HIV status during this episode

Information [Type= discrete] [Format=numeric] [Range= 0-9] [Missing=*]

Statistics [NW/ W] [Valid=191045 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Unknown	0	
1	Pre-negative	59196	31.0%
2	Negative	37080	19.4%
3	Post-negative	53853	28.2%
4	Pre-positive	17632	9.2%
5	Positive	22704	11.9%
9	Error	580	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.

HIVStatusChange: HIV Status changed at the start of this episode

Information [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]

Statistics [NW/ W] [Valid=191045 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	0	
1	Yes	191045	100.0%

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.

HIVSeroconversion: Individual seroconverted at the start of this episode

Information [Type= discrete] [Format=numeric] [Range= 0-1] [Missing=*]

Statistics [NW/ W] [Valid=191045 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	185682	97.2%
1	Yes	5363	2.8%

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.

Episode: Sequential episode for individual

Information [Type= discrete] [Format=numeric] [Range= 1-9] [Missing=*]

Statistics [NW/ W] [Valid=191045 /-] [Invalid=0 /-]