

South Africa

Dr Kobus Herbst, AHRI, SAPRIN

Dr Guy Harling, AHRI

Dr Mark Siedner, AHRI

Prof Willem Hanekom, AHRI

AHRI.Surveillance Episodes Basic

Study Documentation

February 19, 2026

Metadata Production

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

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
SurveillanceEpisodesBasic	7
Variables List	8
SurveillanceEpisodesBasic	8
Variables Description	10
SurveillanceEpisodesBasic	11

AHRI.Surveillance Episodes Basic

Overview

Identification	AHRI.Surveillance.Episodes.Basic
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
 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

Scope & Coverage	
Keywords	Fertility, Mortality, Migration, Social-economic status, HIV, ART
Topics	Surveillance Episodes Basic
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 2000-01-01 end 2024-12-31
------------------------------	------------------------------------

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.Surveillance Episodes Basic [Data set]. Africa Health Research Institute.

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

Files Description

Dataset contains 1 file(s)

SurveillanceEpisodesBasic	
# Cases	592430
# Variable(s)	34

Variables List

Dataset contains 34 variable(s)

File SurveillanceEpisodesBasic							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	Individu..	Unique individual identifier	continuous	numeric-9.0	592430	0	-
2	Sex	Sex of individual	discrete	numeric-9.0	592430	0	-
3	DoB	Date of birth	discrete	character-11	592429	-	-
4	DoD	Date of death	discrete	character-11	52658	-	-
5	MotherId	Mother's IndividualId	continuous	numeric-9.0	402627	189803	-
6	FatherId	Father's IndividualId	continuous	numeric-9.0	207850	384580	-
7	LocationId	Where individual was resident, household residence if non-resident	continuous	numeric-9.0	592430	0	-
8	Househol..	Unique household identifier of the household the individual is a member of	continuous	numeric-9.0	592430	0	-
9	HHRelati..	Relationship to head of household	discrete	numeric-22.0	592430	0	-
10	Resident	Whether individual is resident for duration of episode	discrete	numeric-15.0	592430	0	-
11	Episode	This episode number (first=1, last=Episodes)	discrete	numeric-9.0	592430	0	-
12	Episodes	Total number of episodes for individual	discrete	numeric-9.0	592430	0	-
13	StartDate	Start date of episode (inclusive)	discrete	character-11	592430	-	-
14	EndDate	End date of episode (inclusive)	discrete	character-11	592430	-	-
15	Days	Duration in days of episode	continuous	numeric-9.0	592430	0	-
16	Born	Episode starts with the birth of the individual	discrete	numeric-15.0	592430	0	-
17	Enumerat..	Episode starts with an enumeration	discrete	numeric-15.0	592430	0	-
18	InMigrat..	Episode starts with an in-migration	discrete	numeric-15.0	592430	0	-
19	Location..	Episode starts with an internal migration	discrete	numeric-15.0	592430	0	-
20	ExtResSt..	Flag to indicate start of external residence	discrete	numeric-15.0	592430	0	-
21	Particip..	Resumed participation after refusal	discrete	numeric-15.0	592430	0	-
22	Died	Episode ends with the death of the individual	discrete	numeric-15.0	592430	0	-
23	OutMigra..	Episode ends with the out-migration	discrete	numeric-15.0	592430	0	-
24	Location..	Episode ends with an internal migration	discrete	numeric-15.0	592430	0	-

File SurveillanceEpisodesBasic							
#	Name	Label	Type	Format	Valid	Invalid	Question
25	ExtResEnd	Flag to indicate end of external residence	discrete	numeric-15.0	592430	0	-
26	Refusal	Individual refused follow-up	discrete	numeric-15.0	592430	0	-
27	LostToFo..	Individual was lost to follow-up at the end of the episode	discrete	numeric-15.0	592430	0	-
28	Current	Individual still under surveillance	discrete	numeric-15.0	592430	0	-
29	HHRelati..	Relationship to head of household changed during this episode	discrete	numeric-15.0	592430	0	-
30	Membersh..	Flag to indicate start of household membership	discrete	numeric-15.0	592430	0	-
31	Membersh..	Flag to indicate end of household membership	discrete	numeric-15.0	592430	0	-
32	GapStart	Start of a gap in individual exposure record	discrete	numeric-15.0	592430	0	-
33	GapEnd	End of gap in individual exposure record	discrete	numeric-15.0	592430	0	-
34	Membersh..	Number of concurrent household memberships	continuous	numeric-9.0	592430	0	-

Variables Description

Dataset contains 34 variable(s)

File : SurveillanceEpisodesBasic

IndividualId: Unique individual identifier

Information [Type= continuous] [Format=numeric] [Range= 1-263153] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-] [Mean=120551.372 /-] [StdDev=77495.521 /-]

Sex: Sex of individual

Information [Type= discrete] [Format=numeric] [Range= 0-3] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0		9	0.0%
1	Male	277273	46.8%
2	Female	315148	53.2%
3	Unknown	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.

DoB: Date of birth

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

Statistics [NW/ W] [Valid=592429 /-]

DoD: Date of death

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

Statistics [NW/ W] [Valid=52658 /-]

MotherId: Mother's IndividualId

Information [Type= continuous] [Format=numeric] [Range= 8-263150] [Missing=*]

Statistics [NW/ W] [Valid=402627 /-] [Invalid=189803 /-] [Mean=155358.748 /-] [StdDev=66458.835 /-]

FatherId: Father's IndividualId

Information [Type= continuous] [Format=numeric] [Range= 49-263146] [Missing=*]

Statistics [NW/ W] [Valid=207850 /-] [Invalid=384580 /-] [Mean=163154.15 /-] [StdDev=60487.701 /-]

LocationId: Where individual was resident, household residence if non-resident

Information [Type= continuous] [Format=numeric] [Range= 0-37292] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-] [Mean=18418.826 /-] [StdDev=10750.662 /-]

HouseholdId: Unique household identifier of the household the individual is a member of

Information [Type= continuous] [Format=numeric] [Range= 1-41014] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-] [Mean=20439.449 /-] [StdDev=11795.578 /-]

HHRelationshipTypeId: Relationship to head of household

Information [Type= discrete] [Format=numeric] [Range= 1-12] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Head	59236	10.0%
2	Spouse	23643	4.0%
3	Child	190038	32.1%
4	Son/Daughter-inlaw	2707	0.5%
5	Grandchild	133656	22.6%
6	Parent	3307	0.6%

File : SurveillanceEpisodesBasic

HHRelationshipTypeId: Relationship to head of household

Value	Label	Cases	Percentage
7	Parent-inlaw	192	0.0%
8	Grandparent	871	0.1%
9	Sibling	107282	18.1%
10	Other relative	4112	0.7%
11	Domestic worker/tenant	40978	6.9%
12	Unrelated/Other	26408	4.5%

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.

Resident: Whether individual is resident for duration of episode

Information	[Type= discrete] [Format=numeric] [Range= -1-1] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	215949	36.5%
1	Yes	376481	63.5%

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: This episode number (first=1, last=Episodes)

Information	[Type= discrete] [Format=numeric] [Range= 1-21] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-] [Invalid=0 /-]

Episodes: Total number of episodes for individual

Information	[Type= discrete] [Format=numeric] [Range= 1-21] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-] [Invalid=0 /-]

StartDate: Start date of episode (inclusive)

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-]

EndDate: End date of episode (inclusive)

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-]

Days: Duration in days of episode

Information	[Type= continuous] [Format=numeric] [Range= 1-9132] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-] [Invalid=0 /-] [Mean=1748.351 /-] [StdDev=1857.291 /-]

Born: Episode starts with the birth of the individual

Information	[Type= discrete] [Format=numeric] [Range= -1-1] [Missing=*]
Statistics [NW/ W]	[Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	542974	91.7%
1	Yes	49456	8.3%

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.

File : SurveillanceEpisodesBasic

Enumeration: Episode starts with an enumeration

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	453643	76.6%
1	Yes	138787	23.4%

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.

InMigration: Episode starts with an in-migration

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	434744	73.4%
1	Yes	157686	26.6%

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.

LocationEntry: Episode starts with an internal migration

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	539516	91.1%
1	Yes	52914	8.9%

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.

ExtResStart: Flag to indicate start of external residence

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	416346	70.3%
1	Yes	176084	29.7%

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.

Participation: Resumed participation after refusal

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	592430	100.0%
1	Yes	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.

Died: Episode ends with the death of the individual

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

File : SurveillanceEpisodesBasic

Died: Episode ends with the death of the individual

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	564726	95.3%
1	Yes	27704	4.7%

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.

OutMigration: Episode ends with the out-migration

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	368608	62.2%
1	Yes	223822	37.8%

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.

LocationExit: Episode ends with an internal migration

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	539516	91.1%
1	Yes	52914	8.9%

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.

ExtResEnd: Flag to indicate end of external residence

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	499129	84.3%
1	Yes	93301	15.7%

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.

Refusal: Individual refused follow-up

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	589714	99.5%
1	Yes	2716	0.5%

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.

LostToFollowUp: Individual was lost to follow-up at the end of the episode

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

File : SurveillanceEpisodesBasic

Current: Individual still under surveillance

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	433226	73.1%
1	Yes	159204	26.9%

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.

HHRelationship: Relationship to head of household changed during this episode

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	526437	88.9%
1	Yes	65993	11.1%

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.

MembershipStart: Flag to indicate start of household membership

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	247191	41.7%
1	Yes	345239	58.3%

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.

MembershipEnd: Flag to indicate end of household membership

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	247191	41.7%
1	Yes	345239	58.3%

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.

GapStart: Start of a gap in individual exposure record

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

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	577598	97.5%
1	Yes	14832	2.5%

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.

GapEnd: End of gap in individual exposure record

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

File : SurveillanceEpisodesBasic

GapEnd: End of gap in individual exposure record

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	577598	97.5%
1	Yes	14832	2.5%

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.

Memberships: Number of concurrent household memberships

Information [Type= continuous] [Format=numeric] [Range= 1-28] [Missing=*]

Statistics [NW/ W] [Valid=592430 /-] [Invalid=0 /-] [Mean=1.177 /-] [StdDev=0.422 /-]