

South Africa

Dr. Kobus Herbst, AHRI, SAPRIN

Dr. Guy Harling, AHRI

Dr. Mark Siedner, AHRI

Prof Willem Hanekom, AHRI

AHRI.Coresidency Episodes Dataset

Study Documentation

February 18, 2026

Metadata Production

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

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
CoresidencyEpisodes	7
Variables List	8
CoresidencyEpisodes	8
Variables Description	9
CoresidencyEpisodes	10

AHRI.Coresidency Episodes Dataset

Overview

Identification	AHRI.CoreEpisodes
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	Coresidency Episodes
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.Coresidency Episodes Dataset [Data set]. Africa Health Research Institute.

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

Files Description

Dataset contains 1 file(s)

CoresidencyEpisodes	
# Cases	530673
# Variable(s)	10

Variables List

Dataset contains 10 variable(s)

File CoresidencyEpisodes							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	Individu..	Unique individual identifier	continuous	numeric-9.0	530673	0	-
2	Episode	This episode number	continuous	numeric-9.0	530673	0	-
3	StartDate	Start date of episode (inclusive)	discrete	character-11	530673	-	-
4	EndDate	End date of episode (inclusive)	discrete	character-11	530673	-	-
5	Days	Duration in days of episode	continuous	numeric-9.0	530673	0	-
6	MotherDead	Mother is dead	discrete	numeric-9.0	530673	0	-
7	MotherCo..	Mother is co-resident	discrete	numeric-19.0	530673	0	-
8	FatherDead	Father is dead	discrete	numeric-9.0	530673	0	-
9	FatherCo..	Father is co-resident	discrete	numeric-19.0	530673	0	-
10	Parental..	Change in parental status	discrete	numeric-15.0	530673	0	-

Variables Description

Dataset contains 10 variable(s)

File : CoresidencyEpisodes

IndividualId: Unique individual identifier

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

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-] [Mean=110697.605 /-] [StdDev=77463.3 /-]

Episode: This episode number

Information [Type= continuous] [Format=numeric] [Range= 1-22] [Missing=*]

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-] [Mean=2.197 /-] [StdDev=1.746 /-]

StartDate: Start date of episode (inclusive)

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

Statistics [NW/ W] [Valid=530673 /-]

EndDate: End date of episode (inclusive)

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

Statistics [NW/ W] [Valid=530673 /-]

Days: Duration in days of episode

Information [Type= continuous] [Format=numeric] [Range= 1-9132] [Missing=*]

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-] [Mean=1951.815 /-] [StdDev=2088.028 /-]

MotherDead: Mother is dead

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

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown	91540	17.2%
0	No	360425	67.9%
1	Yes	78708	14.8%

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.

MotherCoresident: Mother is co-resident

Information [Type= discrete] [Format=numeric] [Range= -1-2] [Missing=*]

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown	93676	17.7%
0	No	121126	22.8%
1	Non-resident member	84250	15.9%
2	Resident member	231621	43.6%

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.

FatherDead: Father is dead

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

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown	188790	35.6%
0	No	175972	33.2%
1	Yes	165911	31.3%

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 : CoresidencyEpisodes

FatherCoresident: Father is co-resident

Information [Type= discrete] [Format=numeric] [Range= -1-2] [Missing=*]

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown	194994	36.7%
0	No	196912	37.1%
1	Non-resident member	57250	10.8%
2	Resident member	81517	15.4%

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.

ParentalStatusChange: Change in parental status

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

Statistics [NW/ W] [Valid=530673 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
-1	Unknown/Missing	0	
0	No	0	
1	Yes	530673	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.