

# South Africa - SAPRIN Individual Surveillance Episodes Year Age Household Socioeconomic Status 2024 Datasets

**Dr Kobus Herbst, Prof Mark Collinson, Tinofa Mutevedzi, Dr Molulaqhooa Linda  
Maoyi**

Report generated on: April 15, 2025

Visit our data catalog at: <https://saprindata.samrc.ac.za/index.php>

## Identification

### SURVEY ID NUMBER

SAPRIN.SISEYAHSSD2024V1

### TITLE

SAPRIN Individual Surveillance Episodes Year Age Household Socioeconomic Status 2024 Datasets

### COUNTRY

Name	Country code
South Africa	RSA

### STUDY TYPE

Demographic Surveillance

### SERIES INFORMATION

This dataset contains demographic surveillance data covering the period from 1 Jan 1993 to 31 December 2023.

### ABSTRACT

The 'South African Population Research Infrastructure Network' (SAPRIN) is a national research infrastructure funded through the Department of Science and Innovation and hosted by the South African Medical Research Council. One of SAPRIN's initial goals has been to harmonise and share the longitudinal data from the three current Health and Demographic Surveillance System \*(HDSS)\* Nodes. These long-standing nodes are the MRC/Wits University Agincourt HDSS in Bushbuckridge District, Mpumalanga, established in 1993, with a current population of 104 679 people; the University of Limpopo DIMAMO HDSS in the Capricorn District of Limpopo, established in 1996, with a current population of 101 376; and the Africa Health Research Institute (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, established in 2000, with a current population of 156 751.

For an individual to be eligible for inclusion in the surveillance, the individual must be a member of a household resident within the geographic boundaries of a SAPRIN node. For a household to be resident, it must have at least one household member who is resident within the surveillance area. Households and household membership are self-defined by the household informant interviewed by the fieldworker at their place of residence (or during a telephonic interview with the household informant). Household members so identified could be resident, that is sleep the majority of nights at this household's place of residence, or could be resident elsewhere (usually outside the surveillance area, but potentially within the surveillance area, in which case they will be a resident household member of the household resident at that place - In this dataset, individuals are members of a single household at a time, and in this example, the non-resident member of the household who is resident elsewhere in the surveillance area, will be reflected in the dataset as a resident member of the household this individual is co-resident with and not also as a non-resident member of this household). The resident status of household members can change: they can move out of the surveillance area to be resident elsewhere, but still be considered household members (so-called 'temporary migration'), such cases are reflected in the data as episodes of external residence; or temporary migrants can return to take up residence again with the household, initiating a new episode of residence internal to the surveillance area.

In addition to these periods of internal and external residence punctuated by in- and outmigration, surveillance episodes can be started by the birth of an individual, if the child is born to a resident mother, their birth starts a period of internal residency for the child; if the child is born to a mother who is a temporary migrant (externally resident) and the child is considered to be a member of the household, a period of external residency ensues for the child. Residency episodes (whether internal or external) are of course terminated by the death of the individual, if that happens whilst the individual is under surveillance.

All SAPRIN nodes conducted a baseline household census at their beginning and all individuals enrolled at this point start their surveillance episode with enumeration. However, nodes may extend their area of surveillance at certain points after the initial household census, by doing another baseline census in these new areas, and all individuals enrolled then, also start their surveillance episodes with enumeration. For integrity in the longitudinal surveillance of individuals, the identity of newly encountered individuals is checked against the database and merged with prior records if the individual is already in the database. In the case of newly incorporated areas into the surveillance area, it is entirely possible to find individuals that have previously out-migrated from the surveillance area to reside in this new area, in such cases an individual will have more than one surveillance episode that starts with enumeration, their enumeration in the original baseline census as well as their enumeration in the newly extended surveillance area.

This dataset represents a snapshot of the continually evolving data in the underlying longitudinal databases maintained by the SAPRIN nodes. In these databases the rightmost extent of the individual's surveillance episode is indicated by the data

collection date of the last time the individual's membership of a household under surveillance has been confirmed. Each dataset has a right censor date (31 December 2023 for the current version of the dataset) and individual surveillance episodes are terminated at that point if the individual is still under surveillance beyond the cut-off date. Each individual surveillance episode is associated with a physical location, for internal residency episodes it is the actual place of residence of the individual, for external residence episodes (periods of temporary migration) it is the place of residence of the individual's household. If an individual change their place of residency from one location within the surveillance area to another location still within the surveillance area, the episode at the original location is terminated with a location exit event, and a new episode starts with a location entry event at the destination location. It is also possible for the household the individual is a member of, to change their place of residency in the surveillance area, whilst the individual is externally resident (is a temporary migrant), in which case the individual's external resident episode will also be split with a location exit-entry pair of events.

At every household visit written consent is obtained from the household respondent for continued participation in the surveillance and such consent can be withdrawn. When this happens all household members' surveillance episodes are terminated with a refusal event. It is possible for households to again provide consent to participate in the surveillance after some time, in such cases surveillance events are restarted with a permission event.

As mentioned previously, surveillance episodes are continually extended by the last data collection event if the individual remains under surveillance. In certain cases, individuals may be lost to follow-up and surveillance episodes where the date of last data collection is more than one year prior to the right censor data are terminated as lost to follow up at that last data collection date. Individuals with data collection dates within a year of the right censor date is considered still to be under surveillance up to this last data collection date.

Each surveillance episode contains the identifier of the household the individual is a member of during that episode. Under relatively rare circumstances it is possible for an individual to change household membership whilst still resident at the same location, or to change membership whilst externally resident, in these cases the surveillance episode will be split with a pair of membership end and membership start events. More commonly membership start and end events coincide with location exit and entry events or in- and out-migration events. Memberships also obviously start at birth or enumeration and end at death, refusal to participate or lost to follow-up.

In about half of the cases, individuals have a single episode from first enumeration, birth or in-migration, to their eventual death, out-migration or currently still under surveillance. In the remaining cases, individuals transition from internal residency to external residency via out-migration, or from one location to another via internal migration with a location exit and entry event, or some other rarer form of transition involving membership change, refusal or lost to follow-up. Usually these series of surveillance episodes are continuous in time, with no gaps between episodes, but gaps can form, e.g. when an individual out-migrates and end membership with the household and so is no longer under surveillance, only to return via in-migration at some future date and take up membership with same or different household.

The SAPRIN Individual Surveillance Episodes Year Age Household Socioeconomic Status 2024 Datasets consists of the Demographic surveillance dataset. This dataset splits the basic surveillance episodes at calendar year end, age in years (birth-day) and Household status of an individual changes.

#### KIND OF DATA

Event history data

#### UNIT OF ANALYSIS

Exposure Episodes

## Version

---

#### VERSION DESCRIPTION

v1: Dataset for public distribution.

#### VERSION DATE

2025-04-14

#### VERSION NOTES

v1: Dataset for public distribution.

## Scope

---

#### NOTES

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 or household socioeconomic status, will likewise result in a new episode on the date of the change.

## TOPICS

Topic
Fertility, Mortality, Migration, Household Socioeconomic Status

## KEYWORDS

Keyword
Fertility, Mortality, Migration, Household Socioeconomic Status

# Coverage

## GEOGRAPHIC COVERAGE

The South African Population Research Infrastructure Network (SAPRIN) currently represents a network of three Health and Demographic Surveillance System \*(HDSS)\* nodes located in rural South Africa, namely: 1) MRC/Wits University Agincourt HDSS in Bushbuckridge District, Mpumalanga, which has collected data since 1993. The nodal website is: <http://www.agincourt.co.za>; 2) the University of Limpopo DIMAMO HDSS in the Capricorn District of Limpopo, which has collected data since 1996. The nodal website is: N/A; 3) and the Africa Health Research Institute (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, which has collected data since 2000. The nodal website is: <http://www.ahri.org>.

The Agincourt HDSS covers a surveillance area of approximately 420 square kilometres and is located in the Bushbuckridge District, Mpumalanga in the rural northeast of South Africa close to the Mozambique border. At baseline in 1992, 57 600 people were recorded in 8900 households in 20 villages; by 2006, the population had increased to about 70 000 people in 11 700 households. As of 1st July 2023, there were 110 896 people under surveillance of whom 24% were not resident within the surveillance area, with a total of about 2.8 m person years of observation. 29% of the population is under 15 years old. The population is almost exclusively Xitsonga speaking. The Agincourt HDSS has population density of over 200 persons per square kilometre. The Agincourt HDSS extends between latitudes 24° 50' and 24° 56' S and longitudes 31°08' and 31° 25' E. The altitude is about 400-600m above sea level.

DIMAMO is located in the Capricorn district, Limpopo Province approximately 40 kilometres from Polokwane, the capital city of Limpopo Province and 15-50 kilometres from the University of Limpopo. The site covers an area of approximately 400 square kilometres. The initial total population observed was about 8 000 but the field site was expanded in 2010. As of 1st July 2023, there were 99 087 people under surveillance, of whom 24% were not resident within the surveillance area, with about 961,000 person years of observation. 29% of the population is under 15 years old. The population is predominantly Sepedi speaking. Most households have electricity. Some households have piped water either inside the house or in their yards, but most fetch water from taps situated at strategic points in the villages. Most households have a pit latrine in their yards. The area lies between latitudes and 23°65' and 23°90' S and longitudes 29°65' and 29°85' E. The HDSS is located on a high plateau area (approximately 1250 m above sea level) where communities typically consist of households clustered in villages, with access to local land for small-scale food production.

Africa Health Research Institute (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 for portions where the Kwamsane township straddles the highway) and in the north by the Inyalazi river for portions of the boundary. The surveillance area is approximately 850 square kilometres. As of 1st July 2023, there were 154 815 people under surveillance of whom 25% were not resident within the surveillance area, with about 2.6 m person years of observation. 29 % of the population is under 15 years old. The population is almost exclusively isiZulu speaking. The surveillance area is typical of many rural areas of South Africa in that while predominantly rural, it contains an urban township and informal peri-urban settlements. The area is characterized by large variations in population densities (20-3000 people per square kilometre). The area lies between latitudes -28°24' and 28°20' N and longitudes 32°10' and 31°58' E.

## UNIVERSE

Households resident in dwellings within the study area will be eligible for inclusion in the household component of SAPRIN.

All individuals identified by the household proxy informant as a member of the household will be enumerated. A resident household member is an individual that intends to sleep the majority of time at the dwelling occupied by the household over a four-month period. Households will 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 most of the time. They 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	SAPRIN
Prof Mark Collinson	SAPRIN
Tinofa Mutevedzi	SAPRIN
Dr Molulaqhooa Linda Maoyi	SAPRIN

### PRODUCERS

Name	Affiliation	Role
Dr Molulaqhooa Linda Maoyi	SAPRIN	Technical Assistance
Dr Kobus Hersbt	SAPRIN	Technical Assistance
Augustine Khumalo	SAPRIN	Technical Assistance

### FUNDING AGENCY/SPONSOR

Name	Abbreviation	Role
Department of Science,Technology and Innovation	DSTI	Current Funder

### OTHER IDENTIFICATIONS/ACKNOWLEDGMENTS

Name	Affiliation	Role
Agincourt Data Team	Agincourt	Providing Data
DIMAMO Data Team	DIMAMO	Providing Data
AHRI Data Team	AHRI	Providing Data
Prof Steve Tollman	Agincourt	
Dr Joseph Tlouyamma	DIMAMO	
Prof Collins Iwuji	AHRI	
Centre for High Performance Computing	Centre for High Performance Computing	Providing IT Infrastructure for Data Processing

## 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	Cycle
1993-01-01	2023-12-31	Agincourt
1996-01-01	2023-12-31	DIMAMO
2000-01-01	2023-12-31	AHRI

## TIME PERIODS

Start date	End date	Cycle
1993-01-01	2023-12-31	Agincourt
1996-01-01	2023-12-31	DIMAMO
2000-01-01	2023-12-31	AHRI

## DATA COLLECTION NOTES

In all the HDSS nodes, data are collected from a household proxy respondent, preferably the head of household or any next available senior adult resident household member, after informed consent was obtained by trained fieldworkers. Respondents are informed of the purpose and confidentiality of the interview, their right to refuse participation or withdraw from the study, and that scientists would be given access to anonymised data to analyse and publish information. Informed consent was verbal in all HDSS nodes until 2016. Written informed consent started in 2017 in AHRI, and 2018 in DIMAMO and 2019 in Agincourt. Until 2016 for Agincourt and AHRI, and 2017 for DIMAMO, data collection was field-based 'paper and pen' personal interviews (PAPI), before changing to field-based computer-assisted personal interviews (CAPI). Since 2019, all SAPRIN HDSS nodes collect data in 3 annual rounds over a 45-week data collection schedule; one field-based CAPI round, sandwiched on either side by a Call-Centre-based computer assisted telephonic interview (CATI), to create 3 data points at an interval of approximately 4 months in each calendar year. In the past HDSS nodes had different data collection frequencies. AHRI data collection was 2 PAPI rounds per year from inception to 2011, changing to 3 PAPI rounds per year between 2012 and 2016, before becoming 1 PAPI round and 2 CATI rounds from 2017. Agincourt and DIMAMO have been collecting data once annually in a census-type format, over 4-5-month period until 2018.

## Questionnaires

## QUESTIONNAIRES

The data on this Repository is not the result of a single questionnaire but is a result of harmonised data from three different sites longitudinally collected over more than twenty years using different questionnaires that varied over time and site.

## Data Processing

## DATA EDITING

The first step in the data preparation process is quality assurance. The SAPRIN Management hub team assess the data submitted to ensure it is in the correct format and falls within expected value ranges. Other potential issues checked include: missing data, incorrect data types, unexpected duplicate or orphan records. The SAPRIN Management hub assess this conversion by running both original operational database and the SAPRIN database created from the operational database through the SAPRINQA data quality assessment and indicator process. The data quality checking process is conducted using the SAPRIN QA Julia Code. The Julia Code provides the Extract, Transform, and Load (ETL) capabilities that facilitates the process of capturing, cleansing, and storing data using a uniform and consistent format that is accessible and relevant to end users. The principle of the data quality checks is that if the data conversion conducted by the nodes was complete and accurate, there should be little or no difference in the data quality and demographic indicators between the base and SAPRIN versions of the nodal data. If the data submitted by the nodes meets the criteria for inclusion into the consolidated dataset the data moves to the second step of the data production process. However, if the data fail the inclusion checks, this could then lead to another iteration of data submission and quality control checks until the SAPRIN Management hub is satisfied that they have high quality data. To produce this final standard dataset, the data is processed using Julia Code on the Centre for High Performance Computing cluster.

## Data Appraisal

## ESTIMATES OF SAMPLING ERROR

Not Applicable

## Access policy

## CONTACTS

Name	Affiliation	Email	URL
Dr Molulaqhooa Linda Maoyi	SAPRIN	<a href="mailto:linda.maoyi@mrc.ac.za">linda.maoyi@mrc.ac.za</a>	<a href="#">Link</a>
Dr Kobus Herbst	SAPRIN	<a href="mailto:kobus.herbst@mrc.ac.za">kobus.herbst@mrc.ac.za</a>	<a href="#">Link</a>

## ACCESS CONDITIONS

This data is made available for access under the following conditions:

- 1)The data and other materials provided by SAPRIN will not be redistributed or sold to other individuals, institutions, or organizations without the written agreement of SAPRIN.
- 2)The data will be used for statistical and scientific research purposes only. They will be used solely for reporting of aggregated information, and not for investigation of specific individuals or organisations. The Data User will neither use nor permit others to use the data in any way other than listed in the original application (Analysis Plan) for access to the dataset.
- 3)No attempt will be made to re-identify respondents, and no use will be made of the identity of any person or establishment discovered inadvertently. Any such discovery should immediately be reported to SAPRIN.
- 4)No attempt will be made to produce links among datasets provided by SAPRIN, or among data from SAPRIN and other datasets that could identify individuals or organizations.
- 5)The Data User will ensure that the data are kept in a secured environment and that only authorized users have access to the data.
- 6)Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from SAPRIN will cite the source of data in accordance with the Citation Requirement provided with each dataset.
- 7)An electronic copy of all reports and publications based on the requested data will be sent to SAPRIN.
- 8)The original collector of the data, SAPRIN, and relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
- 9) Once the data set has served its indicated purpose it must be destroyed. If the dataset needs to be lodged for publication purposes, a reference (a digital object identifier will be maintained by SAPRIN for this purpose) to the original dataset on the SAPRIN data repository should be used. Derived or aggregated datasets produced from the original dataset do not fall within this provision and may be lodged as publication datasets. If the same dataset is needed for a different purpose, the dataset should be re-requested and the new purposes indicated.

## CITATION REQUIREMENTS

Any use of this data must cite the digital object identifier (doi) associated with the appropriate dataset. Using the following form:

Maoyi, ML; Mutevedzi, T; Collinson, M; Herbst, K (2024) : SAPRIN Individual Surveillance Episodes 2024: Age-Year-Household Socioeconomic Status Dataset. South African Population Research Infrastructure Network.  
<https://doi.org/10.23667/SAPRIN.SISEYAHSSD2024V1>

Maoyi, ML; Mutevedzi, T; Collinson, M; Herbst, K (2024) : SAPRIN Individual Surveillance Episodes 2024: Age-Year-Household Observation Status Dataset. South African Population Research Infrastructure Network.  
<https://doi.org/10.23667/SAPRIN.SISEYAHOD2024V1>

Maoyi, ML; Mutevedzi, T; Collinson, M; Herbst, K (2024) : SAPRIN Individual Surveillance Episodes 2024: Age-Year-Household Socioeconomic Status Observation Dataset. South African Population Research Infrastructure Network.  
<https://doi.org/10.23667/SAPRIN.SISEYAHSSOD2024V1>

## ACCESS AUTHORITY

Name	Affiliation	Email	URL
Dr Molulaqhooa Linda Maoyi	SAPRIN	<a href="mailto:linda.maoyi@mrc.ac.za">linda.maoyi@mrc.ac.za</a>	<a href="#">Link</a>
Dr Kobus Herbst	SAPRIN	<a href="mailto:kobus.herbst@mrc.ac.za">kobus.herbst@mrc.ac.za</a>	<a href="#">Link</a>

## Disclaimer and copyrights

DISCLAIMER

The user of the data acknowledges that the original collector of the data and the relevant funding agencies bear no responsibility for the data's use or interpretation or inferences based upon it.

COPYRIGHT

This dataset documentation is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License. The dataset is shared in terms of the data-use agreement accepted at the time of data download.

## Metadata production

DDI DOCUMENT ID  
DDI.SAPRIN.SISEYAHSSD2024V1

PRODUCERS

Name	Abbreviation	Affiliation	Role
Molulaqhooa Linda Maoyi	MLM	SAPRIN	Documentation of Study and Review of the metadata
Kobus Herbst	KH	SAPRIN	Documentation of Study and Review of the metadata
Mark Collinson	MC	SAPRIN	Documentation of Study and Review of the metadata
Tinofa Mutevedzi	TM	SAPRIN	Documentation of Study and Review of the metadata

DATE OF METADATA PRODUCTION  
2025-04-15

DDI DOCUMENT VERSION  
Version 1 (April 2025)



Data Dictionary

Data file	Cases	Variables
SISEYAHOD2024V1	576907	6
SISEYAHSSD2024V1	609826	8
SISEYAHSSOD2024V1	18037487	43



**Data file: SISEYAHOD2024V1**

Cases:	576907
Variables:	6

**Variables**

ID	Name	Label	Question
V93	NodeId	SAPRIN Node identifier	
V94	HouseholdId	Unique household identifier	
V95	ObservationId	Unique observation identifier	
V96	StartDate	Start Date (inclusive)	
V97	EndDate	End Date (inclusive)	
V98	ObservationDate	Observation date	

Total: 6

Data file: SISEYAHSSD2024V1

Cases:	609826
Variables:	8

Variables

ID	Name	Label	Question
V99	NodeId	SAPRIN Node identifier	
V100	ObservationId	Unique observation identifier	
V101	ModernIdx	Modern Asset index	
V102	LivestockIdx	Livestock index	
V103	DwellingIdx	Dwelling index	
V104	WaterSanitationIdx	Water and sanitation index	
V105	PowerSupplyIdx	Power supply index	
V106	SEIdx	Socio-economic index	

Total: 8

**Data file: SISEYAHSSOD2024V1**

Cases: 18037487

Variables: 43

**Variables**

ID	Name	Label	Question
V107	NodeId	SAPRIN Node identifier	
V108	IndividualId	Unique individual identifier	
V109	Sex	Sex of individual	
V110	DoB	Date of birth	
V111	DoD	Date of death	
V112	MotherId	Mother's IndividualId	
V113	FatherId	Father's IndividualId	
V114	LocationId	Where individual was resident, household residence if non-resident	
V115	HouseholdId	Unique household identifier of the household the individual is a member of	
V116	HHRelationshipTypeId	Relationship to head of household	
V117	Resident	Whether individual is resident for duration of episode	
V118	Episode	This episode number (first=1, last=Episodes)	
V119	Episodes	Total number of episodes for individual	
V120	StartDate	Start date of episode (inclusive)	
V121	EndDate	End date of episode (inclusive)	
V122	Days	Duration in days of episode	
V123	CalendarYear	Calendar year during episode	
V124	Age	Age in completed years during episode	
V125	AgeGroup	Age group during episode	
V126	Born	Episode starts with the birth of the individual	
V127	Enumeration	Episode starts with an enumeration	
V128	InMigration	Episode starts with an in-migration	
V129	LocationEntry	Episode starts with an internal migration	
V130	ExtResStart	Flag to indicate start of external residence	
V131	Participation	Resumed participation after refusal	
V132	Died	Episode ends with the death of the individual	
V133	OutMigration	Episode ends with the out-migration	
V134	LocationExit	Episode ends with an internal migration	
V135	ExtResEnd	Flag to indicate end of external residence	
V136	Refusal	Individual refused follow-up	
V137	LostToFollowUp	Individual was lost to follow-up at the end of the episode	
V138	Current	Individual still under surveillance	
V139	HHRelationship	Relationship to head of household changed during this episode	
V140	MembershipStart	Flag to indicate start of household membership	
V141	MembershipEnd	Flag to indicate end of household membership	
V142	GapStart	Start of a gap in individual exposure record	
V143	GapEnd	End of gap in individual exposure record	
V144	AgeChange	Flag age changed	

ID	Name	Label	Question
V145	YrChange	Flag calendar year changed	
V146	Memberships	Number of concurrent household memberships	
V147	EducationStatus	Education level achieved during this calendar year	
V148	ObservationId	Link to HouseholdObservations & HouseholdSocioEconomicStatus	
V149	HouseholdStatusChange	Change in household socio-economic status	

Total: 43



**NODEID: SAPRIN Node identifier****Data file: SISEYAHOD2024V1****Overview**

Valid: 576907    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 9    Range: 1 - 3    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
1	Agincourt	284710	49.4%
2	DIMAMO	51115	8.9%
3	AHRI	241082	41.8%

**HOUSEHOLDID: Unique household identifier****Data file: SISEYAHOD2024V1****Overview**

Valid: 576907    Invalid: 0    Minimum: 1    Maximum: 39752    Mean: 17672.898    Standard deviation: 10465.42  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 39752    Format: Numeric

**OBSERVATIONID: Unique observation identifier****Data file: SISEYAHOD2024V1****Overview**

Valid: 576907    Invalid: 0    Minimum: 1    Maximum: 301330    Mean: 127030.236    Standard deviation: 83739.117  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 301330    Format: Numeric

**STARTDATE: Start Date (inclusive)****Data file: SISEYAHOD2024V1****Overview**

Valid: 576907    Minimum: 1970-01-01    Maximum: 2023-12-22  
 Type: Discrete    Width: 11    Range: -    Format: character

**ENDDATE: End Date (inclusive)****Data file: SISEYAHOD2024V1**



Overview

Valid: 576907    Minimum: 1992-09-30    Maximum: 2023-12-30  
Type: Discrete    Width: 11    Range: -    Format: character

---

**OBSERVATIONDATE: Observation date**

Data file: SISEYAHOD2024V1

Overview

Valid: 576907    Minimum: 1970-01-01    Maximum: 2023-12-22  
Type: Discrete    Width: 11    Range: -    Format: character

---

**OBSERVATIONID: Unique observation identifier****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 1   Maximum: 301330   Mean: 128365.878   Standard deviation: 84018.62  
Type: Continuous   Decimal: 0   Width: 9   Range: 1 - 301330   Format: Numeric

---

**MODERNIDX: Modern Asset index****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 0   Maximum: 1   Mean: 0.359   Standard deviation: 0.189  
Type: Continuous   Decimal: 0   Width: 9   Range: 0 - 1   Format: Numeric

---

**LIVESTOCKIDX: Livestock index****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 0   Maximum: 1   Mean: 0.547   Standard deviation: 0.4  
Type: Continuous   Decimal: 0   Width: 9   Range: 0 - 1   Format: Numeric

---

**DWELLINGIDX: Dwelling index****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 0   Maximum: 1   Mean: 0.614   Standard deviation: 0.369  
Type: Continuous   Decimal: 0   Width: 9   Range: 0 - 1   Format: Numeric

---

**WATERSANITATIONIDX: Water and sanitation index****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 0   Maximum: 1.875   Mean: 0.567   Standard deviation: 0.239  
Type: Continuous   Decimal: 0   Width: 9   Range: 0 - 1.875   Format: Numeric

---

**POWERSUPPLYIDX: Power supply index****Data file: SISEYAHSSD2024V1****Overview**

Valid: 609826   Invalid: 0   Minimum: 0   Maximum: 1   Mean: 0.731   Standard deviation: 0.338  
Type: Continuous   Decimal: 0   Width: 9   Range: 0 - 1   Format: Numeric

**SEIDX: Socio-economic index**

Data file: SISEYAHSSD2024V1

**Overview**

Valid: 609826    Invalid: 0    Minimum: 0.0133    Maximum: 0.987    Mean: 0.564    Standard deviation: 0.2  
Type: Continuous    Decimal: 0    Width: 9    Range: 0.01333333336561918 - 0.986666679382324    Format: Numeric

**NODEID: SAPRIN Node identifier**

Data file: SISEYAHSSD2024V1

**Overview**

Valid: 609826    Invalid: 0  
Type: Discrete    Decimal: 0    Width: 9    Range: 1 - 3    Format: Numeric

**Questions and instructions**

CATEGORIES

Value	Category	Cases	
1	Agincourt	301330	49.4%
2	DIMAMO	59320	9.7%
3	AHRI	249176	40.9%

**NODEID: SAPRIN Node identifier****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0

Type: Discrete Decimal: 0 Width: 9 Range: 1 - 3 Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
1	Agincourt	7792577	43.2%
2	DIMAMO	2499816	13.9%
3	AHRI	7745094	42.9%

**INDIVIDUALID: Unique individual identifier****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0 Minimum: 1 Maximum: 308940 Mean: 127490.441 Standard deviation: 83202.946

Type: Continuous Decimal: 0 Width: 9 Range: 1 - 308940 Format: Numeric

**SEX: Sex of individual****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0

Type: Discrete Decimal: 0 Width: 9 Range: 0 - 3 Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
0		338	0%
1	Male	8595493	47.7%
2	Female	9441651	52.3%
3	Unknown	5	0%

**DOB: Date of birth****Data file: SISEYAHSSOD2024V1**

## Overview

Valid: 18037487    Minimum: 1800-01-01    Maximum: 2023-12-30  
 Type: Discrete    Width: 11    Range: -    Format: character

---

## DOD: Date of death

Data file: SISEYAHSSOD2024V1

## Overview

Valid: 1306172    Minimum: 1992-03-15    Maximum: 2024-05-03  
 Type: Discrete    Width: 11    Range: -    Format: character

---

## MOTHERID: Mother's IndividualId

Data file: SISEYAHSSOD2024V1

## Overview

Valid: 11711100    Invalid: 6326387    Minimum: 13    Maximum: 308937    Mean: 155220.539    Standard deviation: 73467.547  
 Type: Continuous    Decimal: 0    Width: 9    Range: 13 - 308937    Format: Numeric

---

## FATHERID: Father's IndividualId

Data file: SISEYAHSSOD2024V1

## Overview

Valid: 5974857    Invalid: 12062630    Minimum: 49    Maximum: 308936    Mean: 165379.286    Standard deviation: 68645.883  
 Type: Continuous    Decimal: 0    Width: 9    Range: 49 - 308936    Format: Numeric

---

## LOCATIONID: Where individual was resident, household residence if non-resident

Data file: SISEYAHSSOD2024V1

## Overview

Valid: 18037487    Invalid: 0    Minimum: 0    Maximum: 35891    Mean: 15921.444    Standard deviation: 9469.692  
 Type: Continuous    Decimal: 0    Width: 9    Range: 0 - 35891    Format: Numeric

---

## HOUSEHOLDID: Unique household identifier of the household the individual is a member of

Data file: SISEYAHSSOD2024V1

## Overview

Valid: 18037487    Invalid: 0    Minimum: 1    Maximum: 40054    Mean: 17439.203    Standard deviation: 10424.522  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 40054    Format: Numeric

**HHRELATIONSHIPTYPEID: Relationship to head of household****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 22    Range: 1 - 12    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
1	Head	2898026	16.1%
2	Spouse	1126839	6.2%
3	Child	6859594	38%
4	Son/Daughter-inlaw	208665	1.2%
5	Grandchild	3806616	21.1%
6	Parent	127190	0.7%
7	Parent-inlaw	5471	0%
8	Grandparent	12695	0.1%
9	Sibling	1680014	9.3%
10	Other relative	507519	2.8%
11	Domestic worker/tenant	417946	2.3%
12	Unrelated/Other	386912	2.1%

**RESIDENT: Whether individual is resident for duration of episode****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	4604080	25.5%
1	Yes	13433407	74.5%

**EPISODE: This episode number (first=1, last=Episodes)****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0    Minimum: 1    Maximum: 98    Mean: 22.061    Standard deviation: 18.415  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 98    Format: Numeric

---

**EPISODES: Total number of episodes for individual****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0    Minimum: 1    Maximum: 98    Mean: 43.122    Standard deviation: 23.509  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 98    Format: Numeric

---

**STARTDATE: Start date of episode (inclusive)****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Minimum: 1992-03-01    Maximum: 2023-12-31  
 Type: Discrete    Width: 11    Range: -    Format: character

---

**ENDDATE: End date of episode (inclusive)****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Minimum: 1992-03-01    Maximum: 2023-12-31  
 Type: Discrete    Width: 11    Range: -    Format: character

---

**DAYS: Duration in days of episode****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0    Minimum: 1    Maximum: 366    Mean: 130.848    Standard deviation: 93.141  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1 - 366    Format: Numeric

---

**CALENDARYEAR: Calendar year during episode****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0    Minimum: 1992    Maximum: 2023    Mean: 2012.566    Standard deviation: 8.005  
 Type: Continuous    Decimal: 0    Width: 9    Range: 1992 - 2023    Format: Numeric

**AGE: Age in completed years during episode**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 18037487    Invalid: 0    Minimum: 0    Maximum: 199    Mean: 25.477    Standard deviation: 18.954  
Type: Continuous    Decimal: 0    Width: 9    Range: 0 - 199    Format: Numeric

**AGEGROUP: Age group during episode**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 18037487    Invalid: 0  
Type: Discrete    Decimal: 0    Width: 9    Range: 0 - 18    Format: Numeric

**Questions and instructions**

CATEGORIES

Value	Category	Cases	
0	<1	373460	2.1%
1	01-04	1619875	9%
2	05-09	2070015	11.5%
3	10-14	2022623	11.2%
4	15-19	1976167	11%
5	20-24	1943867	10.8%
6	25-29	1752293	9.7%
7	30-34	1437159	8%
8	35-39	1115643	6.2%
9	40-44	854084	4.7%
10	45-49	677040	3.8%
11	50-54	548224	3%
12	55-59	447782	2.5%
13	60-64	361107	2%
14	65-69	278838	1.5%
15	70-74	211775	1.2%
16	75-79	149472	0.8%
17	80-84	100946	0.6%
18	85+	97117	0.5%



**BORN: Episode starts with the birth of the individual****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0

Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17906690	99.3%
1	Yes	130797	0.7%

**ENUMERATION: Episode starts with an enumeration****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0

Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17724321	98.3%
1	Yes	313166	1.7%

**INMIGRATION: Episode starts with an in-migration****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487 Invalid: 0

Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%

0	No	17600652	97.6%
1	Yes	436835	2.4%

## LOCATIONENTRY: Episode starts with an internal migration

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17983552	99.7%
1	Yes	53935	0.3%

## EXTRESSTART: Flag to indicate start of external residence

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17629757	97.7%
1	Yes	407730	2.3%

## PARTICIPATION: Resumed participation after refusal

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	18037487	100%
1	Yes	0	0%

## DIED: Episode ends with the death of the individual

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
 Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17981438	99.7%
1	Yes	56049	0.3%

## OUTMIGRATION: Episode ends with the out-migration

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
 Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17439109	96.7%
1	Yes	598378	3.3%

**LOCATIONEXIT: Episode ends with an internal migration****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17983552	99.7%
1	Yes	53935	0.3%

**EXTRESEND: Flag to indicate end of external residence****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17817660	98.8%
1	Yes	219827	1.2%

**REFUSAL: Individual refused follow-up****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%

0	No	18034249	100%
1	Yes	3238	0%

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

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: 0 - 1 Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
0		17992519	99.8%
1		44968	0.2%

## CURRENT: Individual still under surveillance

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17674689	98%
1	Yes	362798	2%

## HHRELATIONSHIP: Relationship to head of household changed during this episode

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17934022	99.4%
1	Yes	103465	0.6%

## MEMBERSHIPSTART: Flag to indicate start of household membership

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
 Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17236864	95.6%
1	Yes	800623	4.4%

## MEMBERSHIPEND: Flag to indicate end of household membership

Data file: SISEYAHSSOD2024V1

### Overview

Valid: 18037487 Invalid: 0  
 Type: Discrete Decimal: 0 Width: 15 Range: -1 - 1 Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	17236864	95.6%
1	Yes	800623	4.4%

**GAPSTART: Start of a gap in individual exposure record****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	18020741	99.9%
1	Yes	16746	0.1%

**GAPEND: End of gap in individual exposure record****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	18020741	99.9%
1	Yes	16746	0.1%

**AGECHANGE: Flag age changed****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 18037487    Invalid: 0  
 Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%

0	No	11623495	64.4%
1	Yes	6413992	35.6%

**YRCHANGE: Flag calendar year changed**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 18037487    Invalid: 0  
Type: Discrete    Decimal: 0    Width: 15    Range: -1 - 1    Format: Numeric

**Questions and instructions**

CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	11799570	65.4%
1	Yes	6237917	34.6%

**MEMBERSHIPS: Number of concurrent household memberships**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 18037487    Invalid: 0  
Type: Discrete    Decimal: 0    Width: 9    Range: 1 - 11    Format: Numeric

**Questions and instructions**

CATEGORIES

Value	Category	Cases	
1		17528563	97.2%
2		481576	2.7%
3		24740	0.1%
4		1943	0%
5		540	0%
6		68	0%
8		32	0%
9		10	0%
10		1	0%
11		14	0%



**EDUCATIONSTATUS: Education level achieved during this calendar year****Data file: SISEYAHSSOD2024V1****Overview**

Valid: 13190309 Invalid: 4847178

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

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
-1	Unknown	265936	2%
0	Grade 0	307444	2.3%
1	Grade 1	400472	3%
2	Grade 2	512579	3.9%
3	Grade 3	550930	4.2%
4	Grade 4	655403	5%
5	Grade 5	568705	4.3%
6	Grade 6	646135	4.9%
7	Grade 7	737534	5.6%
8	Grade 8	771339	5.8%
9	Grade 9	755819	5.7%
10	Grade 10	1088400	8.3%
11	Grade 11	1109197	8.4%
12	Grade 12	4138910	31.4%
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	681506	5.2%
99	Not attending	0	0%
Sysmiss		4847178	

**OBSERVATIONID: Link to HouseholdObservations & HouseholdSocioEconomicStatus**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 13635976   Invalid: 4401511   Minimum: 1   Maximum: 301330   Mean: 121651.107   Standard deviation: 83321.164  
Type: Continuous   Decimal: 0   Width: 9   Range: 1 - 301330   Format: Numeric

**HOUSEHOLDSTATUSCHANGE: Change in household socio-economic status**

Data file: SISEYAHSSOD2024V1

**Overview**

Valid: 13635976   Invalid: 4401511  
Type: Discrete   Decimal: 0   Width: 15   Range: -1 - 1   Format: Numeric

**Questions and instructions**

CATEGORIES

Value	Category	Cases	
-1	Unknown/Missing	0	0%
0	No	9165395	67.2%
1	Yes	4470581	32.8%
Sysmiss		4401511	