South Africa - SAPRIN Individual Surveillance Episodes 2021 Datasets

Dr Kobus Herbst, Prof Mark Collinson, Tinofa Mutevedzi, Prof Steve Tollman, Dr Eric Maimela, Prof Willem Hanekom

Report generated on: January 24, 2024

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

Overview

Identification

ID NUMBER
SAPRIN.SISED2021V2

Version

VERSION DESCRIPTION
v2: Dataset for public distribution.

PRODUCTION DATE
2024-01-23

NOTES
v2: Dataset for public distribution.

Overview

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 112 831 people; the University of Limpopo DIMAMO HDSS in the Capricorn District of Limpopo, established in 1996, with a current population of 32 026; and the Africa Health Research Institute (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, established in 2000, with a current population of 146 751.

For an individual to 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 extend 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 2018 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 2021 Datasets consists of two types of the Demographic surveillance datasets:

1.SAPRIN Individual Surveillance Episodes 2021: Basic Dataset. This dataset contains only the internal and external residency episodes for an individual.

2.SAPRIN Individual Surveillance Episodes 2021: Age-Year-Delivery Dataset. This dataset splits the basic surveillance episodes at calendar year end and at the date when the age in years (birth-day) of an individual changes. In the case of women who have given births, episodes are split at the time of delivery as well.

KIND OF DATA
Event history data

UNITS OF ANALYSIS
Exposure Episodes

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 marital status, will likewise result in a new episode on the date of the change.

TOPICS

Topic	Vocabulary	URI
Fertility, Mortality, Migration		

KEYWORDS
Fertility, Mortality, Migration

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 2018, there were 112 831 people under surveillance of whom 28% were not resident within the surveillance area, with a total of about 2.2m person years of observation. 32% 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 2018, there were 32 026 people under surveillance, of whom 18% were not resident within the surveillance area, with about 440,000 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 form portions where the Kwamsane township stradles 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 2018, there were 146 751 people under surveillance of whom 28% were not resident within the surveillance area, with about 1.9m person years of observation. 32% 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 INVESTIGATOR(S)

Name	Affiliation
Dr Kobus Herbst	SAPRIN
Prof Mark Collinson	SAPRIN
Tinofa Mutevedzi	SAPRIN
Prof Steve Tollman	Agincourt
Dr Eric Maimela	DIMAMO
Prof Willem Hanekom	AHRI

OTHER PRODUCER(S)

Name	Affiliation	Role
Molulaqhooa Linda Maoyi	SAPRIN	Technical Assistance
Dr Kobus Herbst	SAPRIN	Technical Assistance
Daniel Ohene-Kwofie	Agincourt	Technical Assistance
Joseph Tlouyamma	DIMAMO	Technical Assistance
Dickman Gareta	AHRI	Technical Assistance

FUNDING

Name	Abbreviation	Role
Department of Science and Innovation	DSI	Current Funder

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Agincourt Data Team	Agincourt	Providing Data
DIMAMO Data Team	DIMAMO	Providing Data
AHRI Data Team	AHRI	Providing Data
Centre for High Performance Computing	Centre for High Performance Computing	Providing IT Infrastructure for Data Processing

Metadata Production

METADATA PRODUCED BY

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

DATE OF METADATA PRODUCTION
2024-01-23

DDI DOCUMENT VERSION

Version 2 (January 2024)

Sampling

Sampling Procedure

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

Questionnaires

Overview

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 Collection

Data Collection Dates

Start	End	Cycle
1993-01-01	2018-12-31	Agincourt
1996-01-01	2018-12-31	DIMAMO
2000-01-01	2018-12-31	AHRI

Time Periods

Start	End	Cycle
1993-01-01		Agincourt
1996-01-01		DIMAMO
2000-01-01		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

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 iSHARE data quality assessment and indicator process. The data quality checking process is conducted using Pentaho Data Integration (PDI). PDI 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 SAPRIN Management hub is satisfied that they have high quality data.To produce this final standard dataset, the data is processed using PDI on the Centre for High Performance Computing cluster .

Data Appraisal

Estimates of Sampling Error

Not Applicable

File Description

Variable List

SISEAYDD2021V1

Content	
Cases	9821411
Variable(s)	41
Structure	Type: Keys: ()
Version	
Producer	
Missing Data	

Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V74	NodeId	SAPRIN Node Identifier	discrete	numeric	
V75	IndividualId	Unique individual identifier	contin	numeric	
V76	CalendarYear	The calendar year in which the episode falls	contin	numeric	
V77	Age	The age in completed years of the individual	contin	numeric	
V78	ChildrenEverBorn	Number of children ever born	discrete	numeric	
V79	Resident	Whether individual is resident for duration of episode	discrete	numeric	
V80	LocationId	Where individual was resident, household residence if non-resident	contin	numeric	
V81	HouseholdId	Unique household identifier of the household the individual is a member of	contin	numeric	
V82	Sex	Sex of individual	discrete	numeric	
V83	DoB	Date of birth	discrete	character	
V84	DoD	Date of death	discrete	character	
V85	MotherId	Mother's IndividualId	contin	numeric	
V86	FatherId	Father's IndividualId	contin	numeric	
V87	StartDate	Start date of episode (inclusive)	discrete	character	
V88	EndDate	End date of episode (inclusive)	discrete	character	
V89	Days	Duration in days of episode	contin	numeric	
V90	ChildrenBorn	Number of children born to this woman during episode	discrete	numeric	
V91	Born	Episode starts with the birth of the individual	discrete	numeric	
V92	Enumeration	Episode starts with an enumeration	discrete	numeric	
V93	InMigration	Episode starts with an in-migration	discrete	numeric	
V94	LocationEntry	Episode starts with an internal migration	discrete	numeric	
V95	ExtResStart	Flag to indicate start of external residence	discrete	numeric	
V96	Participation	Resume participation after refusal	discrete	numeric	
V97	Died	Episode ends with the death of the individual	discrete	numeric	
V98	OutMigration	Episode ends with the out-migration	discrete	numeric	
V99	LocationExit	Episode ends with an internal migration	discrete	numeric	
V100	ExtResEnd	Flag to indicate end of external residence	discrete	numeric	

V101	Refusal	Individual refused follow-up	discrete	numeric
V102	LostToFollowUp	Individual was lost to follow-up at the end of the episode	discrete	numeric
V103	Current	Individual still under surveillance	discrete	numeric
V104	MembershipStart	Flag to indicate start of household membership	discrete	numeric
V105	MembershipEnd	Flag to indicate end of household membership	discrete	numeric
V106	Memberships	Number of concurrent household memberships	discrete	numeric
V107	Gap	Individual exposure Gap flag	discrete	numeric
V108	YrStart	Flag episode start due to calendar year change	discrete	numeric
V109	YrEnd	Flag episode end due to calendar year change	discrete	numeric
V110	AgeStart	Flag episode start due to age change	discrete	numeric
V111	AgeEnd	Flag episode end due to age change	discrete	numeric
V112	Delivery	Flag a delivery to this woman at start of episode	discrete	numeric
V113	Episode	This episode number (first=1, last=Episodes)	contin	numeric
V114	Episodes	Total number of episodes for individual	contin	numeric

SISEBD2021V1

Content	
Cases	1067511
Variable(s)	32
Structure	Type: Keys: ()
Version	
Producer	
Missing Data	

Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V115	NodeId	SAPRIN Node Identifier	discrete	numeric	
V116	IndividualId	Unique individual identifier	contin	numeric	
V117	Resident	Whether individual is resident for duration of episode	discrete	numeric	
V118	LocationId	Where individual was resident, household residence if non-resident	contin	numeric	
V119	HouseholdId	Unique household identifier of the household the individual is a member of	contin	numeric	
V120	Sex	Sex of individual	discrete	numeric	
V121	DoB	Date of birth	discrete	character	
V122	DoD	Date of death	discrete	character	
V123	MotherId	Mother's IndividualId	contin	numeric	
V124	FatherId	Father's IndividualId	contin	numeric	
V125	StartDate	Start date of episode (inclusive)	discrete	character	
V126	EndDate	End date of episode (inclusive)	discrete	character	
V127	Days	Duration in days of episode	contin	numeric	
V128	Born	Episode starts with the birth of the individual	discrete	numeric	
V129	Enumeration	Episode starts with an enumeration	discrete	numeric	
V130	InMigration	Episode starts with an in-migration	discrete	numeric	
V131	LocationEntry	Episode starts with an internal migration	discrete	numeric	
V132	ExtResStart	Flag to indicate start of external residence	discrete	numeric	
V133	Participation	Resume participation after refusal	discrete	numeric	
V134	Died	Episode ends with the death of the individual	discrete	numeric	
V135	OutMigration	Episode ends with the out-migration	discrete	numeric	
V136	LocationExit	Episode ends with an internal migration	discrete	numeric	
V137	ExtResEnd	Flag to indicate end of external residence	discrete	numeric	
V138	Refusal	Individual refused follow-up	discrete	numeric	
V139	LostToFollowUp	Individual was lost to follow-up at the end of the episode	discrete	numeric	
V140	Current	Individual still under surveillance	discrete	numeric	
V141	MembershipStart	Flag to indicate start of household membership	discrete	numeric	

V142	MembershipEnd	Flag to indicate end of household membership	discrete	numeric
V143	Memberships	Number of concurrent household memberships	discrete	numeric
V144	Gap	Individual exposure Gap flag	discrete	numeric
V145	Episode	This episode number (first=1, last=Episodes)	contin	numeric
V146	Episodes	Total number of episodes for individual	contin	numeric

SAPRIN Node Identifier (NodeId)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 9
Decimals: 0
Range: 1-3

Valid cases: 9821411
Invalid: 0

Unique individual identifier (IndividualId)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-239251

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 239251
Mean: 104046.7
Standard deviation: 65521.3

The calendar year in which the episode falls (CalendarYear)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1992-2018

Valid cases: 9821411
Invalid: 0
Minimum: 1992
Maximum: 2018
Mean: 2008.4
Standard deviation: 7

The age in completed years of the individual (Age)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 0-129

Valid cases: 9821411
Invalid: 0
Minimum: 0
Maximum: 129
Mean: 24.6
Standard deviation: 18.6

Number of children ever born (ChildrenEverBorn)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 0-16

Valid cases: 9821411
Invalid: 0

Whether individual is resident for duration of episode (Resident)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Where individual was resident, household residence if non-resident (LocationId)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-32172

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 32172
Mean: 14783.3
Standard deviation: 8741.3

Unique household identifier of the household the individual is a member of (HouseholdId)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-36453

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 36453
Mean: 16609.8
Standard deviation: 9886

Sex of individual (Sex)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 0-3

Valid cases: 9821411
Invalid: 0

Date of birth (DoB)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 9821411
Minimum: NaN
Maximum: NaN

Date of death (DoD)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 806080
Minimum: NaN
Maximum: NaN

Mother's IndividualId (MotherId)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 0-239249

Valid cases: 9458867
Invalid: 362544
Minimum: 0
Maximum: 239249
Mean: 91754
Standard deviation: 77000.9

Father's IndividualId (FatherId)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 48-239245

Valid cases: 3547231
Invalid: 6274180
Minimum: 48
Maximum: 239245
Mean: 142177.6
Standard deviation: 51097

Start date of episode (inclusive) (StartDate)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 9821411
Minimum: NaN
Maximum: NaN

End date of episode (inclusive) (EndDate)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 9821411
Minimum: NaN
Maximum: NaN

Duration in days of episode (Days)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-366

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 366
Mean: 167.5
Standard deviation: 101.5

Number of children born to this woman during episode (ChildrenBorn)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 0-4

Valid cases: 9821411
Invalid: 0

Episode starts with the birth of the individual (Born)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode starts with an enumeration (Enumeration)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode starts with an in-migration (InMigration)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode starts with an internal migration (LocationEntry)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag to indicate start of external residence (ExtResStart)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Resume participation after refusal (Participation)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode ends with the death of the individual (Died)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode ends with the out-migration (OutMigration)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Episode ends with an internal migration (LocationExit)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag to indicate end of external residence (ExtResEnd)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Individual refused follow-up (Refusal)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Individual was lost to follow-up at the end of the episode (LostToFollowUp)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Individual still under surveillance (Current)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag to indicate start of household membership (MembershipStart)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag to indicate end of household membership (MembershipEnd)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Number of concurrent household memberships (Memberships)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 1-10

Valid cases: 9821411
Invalid: 0

Individual exposure Gap flag (Gap)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag episode start due to calendar year change (YrStart)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag episode end due to calendar year change (YrEnd)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag episode start due to age change (AgeStart)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag episode end due to age change (AgeEnd)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

Flag a delivery to this woman at start of episode (Delivery)
File: SISEAYDD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 9821411
Invalid: 0

This episode number (first=1, last=Episodes) (Episode)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-69

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 69
Mean: 16.7
Standard deviation: 13.2

Total number of episodes for individual (Episodes)
File: SISEAYDD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-69

Valid cases: 9821411
Invalid: 0
Minimum: 1
Maximum: 69
Mean: 32.4
Standard deviation: 16.2

SAPRIN Node Identifier (NodeId)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 9
Decimals: 0
Range: 1-3

Valid cases: 1067511
Invalid: 0

Unique individual identifier (IndividualId)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-239251

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 239251
Mean: 103671.1
Standard deviation: 65642.6

Whether individual is resident for duration of episode (Resident)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Where individual was resident, household residence if non-resident (LocationId)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-32172

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 32172
Mean: 14721.3
Standard deviation: 8743

Unique household identifier of the household the individual is a member of (HouseholdId)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-36453

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 36453
Mean: 16486.1
Standard deviation: 9901.2

Sex of individual (Sex)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 0-3

Valid cases: 1067511
Invalid: 0

Date of birth (DoB)
File: SISEBD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 1067511
Minimum: NaN
Maximum: NaN

Date of death (DoD)
File: SISEBD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 79737
Minimum: NaN
Maximum: NaN

Mother's IndividualId (MotherId)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 0-239249

Valid cases: 1013271
Invalid: 54240
Minimum: 0
Maximum: 239249
Mean: 90696.4
Standard deviation: 76794.9

Father's IndividualId (FatherId)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 48-239245

Valid cases: 349525
Invalid: 717986
Minimum: 48
Maximum: 239245
Mean: 141693.6
Standard deviation: 52452.2

Start date of episode (inclusive) (StartDate)
File: SISEBD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 1067511
Minimum: NaN
Maximum: NaN

End date of episode (inclusive) (EndDate)
File: SISEBD2021V1

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 1067511
Minimum: NaN
Maximum: NaN

Duration in days of episode (Days)
File: SISEBD2021V1

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-9802

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 9802
Mean: 1541.1
Standard deviation: 1764.9

Episode starts with the birth of the individual (Born)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode starts with an enumeration (Enumeration)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode starts with an in-migration (InMigration)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode starts with an internal migration (LocationEntry)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Flag to indicate start of external residence (ExtResStart)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Resume participation after refusal (Participation)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode ends with the death of the individual (Died)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode ends with the out-migration (OutMigration)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Episode ends with an internal migration (LocationExit)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Flag to indicate end of external residence (ExtResEnd)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Individual refused follow-up (Refusal)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Individual was lost to follow-up at the end of the episode (LostToFollowUp)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Individual still under surveillance (Current)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Flag to indicate start of household membership (MembershipStart)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Flag to indicate end of household membership (MembershipEnd)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

Number of concurrent household memberships (Memberships)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 12
Decimals: 0
Range: 1-10

Valid cases: 1067511
Invalid: 0

Individual exposure Gap flag (Gap)
File: SISEBD2021V1

Overview

Type: Discrete
Format: numeric
Width: 15
Decimals: 0
Range: -1-1

Valid cases: 1067511
Invalid: 0

This episode number (first=1, last=Episodes) (Episode)
File: SISEBD2021V1

Overview	

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-22

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 22
Mean: 2
Standard deviation: 1.5

Total number of episodes for individual (Episodes)
File: SISEBD2021V1

Overview	

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-22

Valid cases: 1067511
Invalid: 0
Minimum: 1
Maximum: 22
Mean: 3.1
Standard deviation: 2.2