

South Africa - SAPRIN Individual Surveillance Episodes 2020: Detailed Dataset

Prof Mark Collinson - SAPRIN
Dr Kobus Herbst - SAPRIN
Prof Steve Tollman - Agincourt
Dr Eric Maimela - DIMAMO
Prof Willem Hanekom - AHRI

Report generated on: June 1, 2020

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

Overview

Identification

ID NUMBER
SAPRIN.SISEDD2020V1

Version

VERSION DESCRIPTION
v1: Dataset for public distribution.

PRODUCTION DATE
2020-06-01

NOTES
v1: 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 population of 113 113 people; the University of Limpopo DIMAMO HDSS in the Capricorn District of Limpopo, established in 1996, with a current population of 38 479; and the Africa Health Research Institute (AHRI) HDSS in uMkhanyakude District, KwaZulu-Natal, established in 2000, with a current population of 139 250.

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 2017 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 2020 Datasets consists of three types of the Demographic surveillance datasets:

1.SAPRIN Individual Surveillance Episodes 2020: Basic Dataset. This dataset contains only the internal and external residency episodes for an individual.

2.SAPRIN Individual Surveillance Episodes 2020: 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.

3.SAPRIN Individual Surveillance Episodes 2020: Detailed Dataset. This dataset adds to the dataset 2 time-varying attributes such as education, employment, marital status and socio-economic status.

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 December 2017, there were 113 113 people under surveillance of whom 28% were not resident within the surveillance area, with a total of about 2m person years of observation. 33% of the population is under 15 years old. The population is almost exclusively Shangaan-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 December 2017, there were 38 479 people under surveillance, of whom 22% were not resident within the surveillance area, with about 400,000 person years of observation. 30% of the population is under 15 years old. The population is predominantly Sotho 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 December 2017, there were 139 250 people under surveillance of whom 28% were not resident within the surveillance area, with about 1.7m person years of observation. 32% of the population is under 15 years old. The population is almost exclusively Zulu-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
Prof Mark Collinson	SAPRIN
Dr Kobus Herbst	SAPRIN
Prof Steve Tollman	Agincourt
Dr Eric Maimela	DIMAMO
Prof Willem Hanekom	AHRI

OTHER PRODUCER(S)

Name	Affiliation	Role
Taurayi Mudzana	SAPRIN	Technical Assistance
Tinofa Mutevedzi	SAPRIN	Technical Assistance
Chodziwadziwa Kabudula	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
Steve Tollman	Agincourt	
Eric Maimela	DIMAMO	
Willem Hanekom	AHRI	

Name	Affiliation	Role
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
Kobus Herbst	KH	SAPRIN	Documentation of Study and Review of the metadata
Taurayi Mudzana	TAU	SAPRIN	Documentation of Study and Review of the metadata
Tinofa Mutevedzi	TM	SAPRIN	Documentation of Study and Review of the metadata
Mark Collinson	MC	SAPRIN	Documentation of Study and Review of the metadata

DATE OF METADATA PRODUCTION

2020-06-01

DDI DOCUMENT VERSION

Version 1 (June 2020)

DDI DOCUMENT ID

DDI.SAPRIN.SISED2020V1

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	2017-12-31	Agincourt
1996-01-01	2017-12-31	DIMAMO
2000-01-01	2017-12-31	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 .

Other Processing

All datasets have a digital fingerprint. This allows for the verification of the integrity of the dataset files. FCIV (File Checksum Integrity Verifier) is a utility from Microsoft that allows us to compute MD5 hash for each dataset.

```
429a8e105c6f275adfa0f9a8ff0f9589 sisedd2020v1.dta  
8b45fee3180129e298a122d3ef8b1b3e sisedd2020v1.csv
```

Data Appraisal

Estimates of Sampling Error

Not Applicable

Other forms of Data Appraisal

Node	Metric	Table	QMetric	Illegal	Legal	Total	Metric
Agincourt	MicroDataCleaned	Starts	-	223 515	223 515		
DIMAMO	MicroDataCleaned	Starts	-	121 572	121 572		
AHRI	MicroDataCleaned	Starts	-	199 831	199 831		
Agincourt	MicroDataCleaned	Ends	-	223 287	223 287		
DIMAMO	MicroDataCleaned	Ends	-	121 572	121 572		
AHRI	MicroDataCleaned	Ends	-	199 831	199 831		
Agincourt	MicroDataCleaned	Sex Values	78 541	827 541	905		
DIMAMO	MicroDataCleaned	Sex Values	10 251	222 251	232		

File Description

Variable List

SISED2020V1

Content	SAPRIN Individual Surveillance Episodes 2020: Detailed Dataset
Cases	13117161
Variable(s)	56
Structure	Type: Keys: ()
Version	SISED2020V1
Producer	South African Population Research Infrastructure Network
Missing Data	

Variables

ID	Name	Label	Type	Format	Question
V354	Nodeld	Nodeld	discrete	numeric	
V355	IndividualId	IndividualId	contin	numeric	
V356	MotherId	MotherId	contin	numeric	
V357	FatherId	FatherId	contin	numeric	
V358	Sex	Sex	discrete	numeric	
V359	DoB	DoB	discrete	character	
V360	DoD	DoD	discrete	character	
V361	Age	Age	contin	numeric	
V362	Year	Year	contin	numeric	
V363	Episodes	Episodes	contin	numeric	
V364	Episode	Episode	contin	numeric	
V365	StartDate	StartDate	discrete	character	
V366	EndDate	EndDate	discrete	character	
V367	Days	Days	contin	numeric	
V368	Resident	Resident	discrete	numeric	
V369	Gap	Gap	discrete	numeric	
V370	Enumeration	Enumeration	discrete	numeric	
V371	Born	Born	discrete	numeric	
V372	Participation	Participation	discrete	numeric	
V373	InMigration	InMigration	discrete	numeric	
V374	LocationEntry	LocationEntry	discrete	numeric	
V375	ExtResStart	ExtResStart	discrete	numeric	
V376	Died	Died	discrete	numeric	
V377	Refusal	Refusal	discrete	numeric	
V378	LostToFollowUp	LostToFollowUp	discrete	numeric	
V379	Current	Current	discrete	numeric	
V380	OutMigration	OutMigration	discrete	numeric	

ID	Name	Label	Type	Format	Question
V381	LocationExit	LocationExit	discrete	numeric	
V382	ExtResEnd	ExtResEnd	discrete	numeric	
V383	MembershipStart	MembershipStart	discrete	numeric	
V384	MembershipEnd	MembershipEnd	discrete	numeric	
V385	GapStart	GapStart	discrete	numeric	
V386	GapEnd	GapEnd	discrete	numeric	
V387	LocationId	LocationId	contin	numeric	
V388	HouseholdId	HouseholdId	contin	numeric	
V389	HHRelation	HHRelation	discrete	numeric	
V390	Memberships	Memberships	discrete	numeric	
V391	Education	Education	discrete	numeric	
V392	MaritalStatus	MaritalStatus	discrete	numeric	
V393	SEIdx	SEIdx	discrete	character	
V394	CurrentEmployment	CurrentEmployment	contin	numeric	
V395	EmploymentType	EmploymentType	contin	numeric	
V396	MotherCoResident	MotherCoResident	discrete	numeric	
V397	MotherDead	MotherDead	discrete	numeric	
V398	FatherCoResident	FatherCoResident	discrete	numeric	
V399	FatherDead	FatherDead	discrete	numeric	
V400	DwellingIdx	DwellingIdx	discrete	character	
V401	LifeStockIdx	LifeStockIdx	discrete	character	
V402	ModerntAssetIdx	ModerntAssetIdx	discrete	character	
V403	PowerSupplyIdx	PowerSupplyIdx	discrete	character	
V404	WaterSanitationIdx	WaterSanitationIdx	discrete	character	
V405	EducationObsDate	EducationObsDate	discrete	character	
V406	MaritalStatusObsDate	MaritalStatusObsDate	discrete	character	
V407	SEObsDate	SEObsDate	discrete	character	
V408	EmploymentSector	EmploymentSector	contin	numeric	
V409	EmploymentObsDate	EmploymentObsDate	discrete	character	

NodeId (NodeId)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-3	

Description

SAPRIN Node Identifier
 1=Agincourt;
 2=DIMAMO;
 3=AHRI

IndividualId (IndividualId)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1000001
Range: 1000001-3217149	Maximum: 3217149
	Mean: 2084937.5
	Standard deviation: 962082.6

Description

Unique anonymised individual identifier. Only unique within nodal dataset

MotherId (MotherId)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 8530243
Format: numeric	Invalid: 4586918
Decimals: 0	Minimum: 1000003
Range: 1000003-3217143	Maximum: 3217143
	Mean: 2121723.5
	Standard deviation: 962797.4

Description

Mother's IndividualId
 0 = Unknown

FatherId (FatherId)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 4537620
Format: numeric	Invalid: 8579541
Decimals: 0	Minimum: 1000012
Range: 1000012-3217142	Maximum: 3217142
	Mean: 2127752.7
	Standard deviation: 998677.8

Description

Father's IndividualId
 0 = Unknown

Sex (Sex)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-3	

Description

Sex of the individual
 0= Missing
 1 = Male
 2 = Female
 3 = Unknown

DoB (DoB)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: character	Invalid: 0
Width: 9	

Description

Date of birth.
 Format yyyy-MM-dd in csv file
 Date of Birth 1800-01-01 is unknown

DoD (DoD)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 986146
Format: character	Invalid: 0
Width: 9	

Description

Date of Death.
 Format yyyy-MM-dd in csv file

Age (Age)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 0
Range: 0-122	Maximum: 122
	Mean: 25.4
	Standard deviation: 18.7

Description

Age in completed years (valid for the duration of the episode)

Year (Year)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1982
Range: 1982-2017	Maximum: 2017
	Mean: 2008.2
	Standard deviation: 6.5

Description

Calendar year during episode

Episodes (Episodes)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-134	Maximum: 134
	Mean: 46.5
	Standard deviation: 22.5

Description

Total number of episodes for individual

Episode (Episode)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-134	Maximum: 134
	Mean: 23.8
	Standard deviation: 18.7

Description

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

StartDate (StartDate)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: character	Invalid: 0
Width: 9	

Description

Start date of episode (inclusive).
Format yyyy-MM-dd in csv file

EndDate (EndDate)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: character	Invalid: 0
Width: 9	

Description

End date of episode (inclusive).
Format yyyy-MM-dd in csv file

Days (Days)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1
Range: 1-5830	Maximum: 5830
	Mean: 117.1
	Standard deviation: 87

Description

Duration in days of the episode.
EndDate - StartDate + 1

Resident (Resident)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Whether individual is resident for duration of episode
1 = Internal residence; 0 = External residence

Gap (Gap)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Gap in individual exposure

Enumeration (Enumeration)

File: SISEDD2020V1

Overview

Enumeration (Enumeration)

File: SISEDD2020V1

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

Valid cases: 13117161
Invalid: 0

Description

Episode starts with an enumeration

Born (Born)

File: SISEDD2020V1

Overview

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

Valid cases: 13117161
Invalid: 0

Description

Whether this episode starts with the birth of the individual. Can start an internal or external residency episode depending on the residency status of the mother, if the mother is an internal resident, then the child is as well, if the mother is externally resident then child is as well. If the child becomes resident shortly after birth, then the child will be in-migrated and start a residency episode at that point.

Participation (Participation)

File: SISEDD2020V1

Overview

Type: Discrete
Format: numeric
Decimals: 0
Range: 0-0

Valid cases: 13117161
Invalid: 0

Description

Resume participation after refusal.

The individual (or more specifically the individual's household) re-consented to participate after a previous refusal

InMigration (InMigration)

File: SISEDD2020V1

Overview

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

Valid cases: 13117161
Invalid: 0

Description

Whether this episode starts with an in-migration.

Either as a new first episode for the individual; or following the end of an external residency; or after a gap in residency

LocationEntry (LocationEntry)

File: SISEDD2020V1

Overview

LocationEntry (LocationEntry)

File: SISEDD2020V1

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

Valid cases: 13117161
Invalid: 0

Description

Whether this episode starts with an internal migration into the location.
Individual changed residency from one location to another for internal residency episodes; for external residency episodes, the household the individual is a member of, changed residency from one location to another

ExtResStart (ExtResStart)

File: SISEDD2020V1

Overview

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

Valid cases: 13117161
Invalid: 0

Description

External residency starts
Individual out-migrated, but retained membership in the household, in other words the start of a temporary migration episode. If this is the start of the first episode for the individual then the household reported a new household member that is currently externally resident.

Died (Died)

File: SISEDD2020V1

Overview

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

Valid cases: 13117161
Invalid: 0

Description

Whether the episode ends with the death of the individual.

Refusal (Refusal)

File: SISEDD2020V1

Overview

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

Valid cases: 13117161
Invalid: 0

Description

Surveillance end resulting from household refusal.
The household the individual is a member of refused further participation in the surveillance

LostToFollowUp (LostToFollowUp)

File: SISEDD2020V1

LostToFollowUp (LostToFollowUp)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Individual was lost to follow-up at the end of the episode.
Individual was last reported on by the household respondent more than a year before the current right censor date.

Current (Current)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Individual still under surveillance.
Either last reported on by the household respondent after the right censor date, or within a year prior to the right censor date.

OutMigration (OutMigration)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Whether the episode ends with the out-migration of the individual.
If the individual retains membership in a household under surveillance this will be followed by an episode of external residence (temporary migration), otherwise this may be the start of a gap in the surveillance episodes of the individual; or it may be the end of the last surveillance episode for the individual.

LocationExit (LocationExit)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Episode ends with an internal migration.
Individual changed residency from one location to another for internal residency episodes; for external residency episodes, the household the individual is a member of, changed residency from one location to another. LocationExit and LocationEntry always occurs in pairs.

ExtResEnd (ExtResEnd)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

External residency end.
Individual in-migrated after a period of external residency; or the individual is no longer a member of any household under surveillance.

MembershipStart (MembershipStart)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Flag to indicate start of household membership. This flag can be set in conjunction with other start flags, in which case it means that a new membership started at the same time. If none of the other entry flags are set it means that the individual changed household membership without changing residency status or place of residence.

MembershipEnd (MembershipEnd)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Flag to indicate end of household membership. This flag can be set in conjunction with other end flags, in which case it means that the membership ended at the same time. If none of the other exit flags are set it means that the individual changed household membership without changing residency status or place of residence

GapStart (GapStart)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

Start of gap in individual exposure.
E.g. between out-migration without retaining any household membership and a subsequent in-migration taking up household membership and residency in the surveillance area again

GapEnd (GapEnd)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 0-1	

Description

End of gap in individual exposure.
GapStart and GapEnd flags always occur in conjunction with the other entry or exit flags

LocationId (LocationId)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1000001
Range: 1000001-3023818	Maximum: 3023818
	Mean: 1993225.8
	Standard deviation: 960663.4

Description

Where individual is resident for the duration of this episode.
For internal residency episodes it is the actual place of residence, for external residency episodes, it is the place of residence of the individual's household

HouseholdId (HouseholdId)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	Minimum: 1000001
Range: 1000001-3033048	Maximum: 3033048
	Mean: 1996454.5
	Standard deviation: 961739.3

Description

Unique household identifier of the household the individual is a member of during this episode

HHRelation (HHRelation)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-12	

Description

Relationship to the head of the household.

Memberships (Memberships)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: numeric	Invalid: 0
Decimals: 0	
Range: 1-10	

Description

Number of concurrent household memberships.
At the start of the episode

Education (Education)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 9745497
Format: numeric	Invalid: 3371664
Decimals: 0	
Range: -1-17	

Description

Education status of individual during this episode.

- 0. None - could be missing or unknown too
- 1. Grade 1
- 2. Grade 2
- 3. Grade 3
- 4. Grade 4
- 5. Grade 5
- 6. Grade 6
- 7. Grade 7
- 8. Grade 8
- 9. Grade 9
- 10. Grade 10
- 11. Grade 11
- 12. Grade 12
- 13. Some tertiary qualification
- 14. Some tertiary qualification

MaritalStatus (MaritalStatus)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 10627459
Format: numeric	Invalid: 2489702
Decimals: 0	
Range: 0-5	

Description

Marital status of the individual during this episode.

- 0 = Unknown
- 1 = Not Married
- 2 = Married
- 3 = Widowed
- 4 = Separated/divorced
- 5 = Informal Union

SEIdx (SEIdx)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 13117161
Format: character	Invalid: 0
Width: 17	

Description

Socio-economic index of the household the individual belongs to.
 Calculated using the approach described in the paper by Kabudula, et al.

CurrentEmployment (CurrentEmployment)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 12649528
Format: numeric	Invalid: 467633
Decimals: 0	Minimum: 0
Range: 0-202	Maximum: 202
	Mean: 1.4
	Standard deviation: 4.2

Description

Employment status of the individual during this episode.

EmploymentType (EmploymentType)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 12649528
Format: numeric	Invalid: 467633
Decimals: 0	Minimum: 0
Range: 0-203	Maximum: 203
	Mean: 185
	Standard deviation: 52.8

Description

Employment type of the individual during this episode

MotherCoResident (MotherCoResident)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 12018852
Format: numeric	Invalid: 1098309
Decimals: 0	
Range: -1-1	

Description

Whether mother is co-resident during episode, yes, no or unknown (if MotherId = 0)

MotherDead (MotherDead)

File: SISEDD2020V1

Overview

MotherDead (MotherDead)

File: SISEDD2020V1

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

Valid cases: 1417881
Invalid: 11699280

Description

Whether mother is dead for the duration of this episode

FatherCoResident (FatherCoResident)

File: SISEDD2020V1

Overview

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

Valid cases: 12135869
Invalid: 981292

Description

Whether father is co-resident during episode, yes, no or unknown (if FatherId = 0)

FatherDead (FatherDead)

File: SISEDD2020V1

Overview

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

Valid cases: 1573819
Invalid: 11543342

Description

Whether father is dead for the duration of this episode

DwellingIdx (DwellingIdx)

File: SISEDD2020V1

Overview

Type: Discrete
Format: character
Width: 17

Valid cases: 13117161
Invalid: 0

Description

The dwelling SE Index.
Not available in the AHRI data. Based on the construction material and size of the dwelling

LifeStockIdx (LifeStockIdx)

File: SISEDD2020V1

Overview

Type: Discrete
Format: character
Width: 3

Valid cases: 13117161
Invalid: 0

Description

LifeStockIdx (LifeStockIdx)

File: SISEDD2020V1

The livestock SE Index.

Not available in the AHRI data. Based on the construction material and size of the dwelling

ModerntAssetIdx (ModerntAssetIdx)

File: SISEDD2020V1

Overview

Type: Discrete

Format: character

Width: 17

Valid cases: 13117161

Invalid: 0

Description

The Asset SE Index.

Based on the ownership of a list of various assets.

PowerSupplyIdx (PowerSupplyIdx)

File: SISEDD2020V1

Overview

Type: Discrete

Format: character

Width: 2

Valid cases: 13117161

Invalid: 0

Description

The power supply SE Index.

Based on the energy source used for lighting and cooking.

WaterSanitationIdx (WaterSanitationIdx)

File: SISEDD2020V1

Overview

Type: Discrete

Format: character

Width: 4

Valid cases: 13117161

Invalid: 0

Description

The water and sanitation SE Index.

Based on the energy source used for lighting and cooking.

EducationObsDate (EducationObsDate)

File: SISEDD2020V1

Overview

Type: Discrete

Format: character

Width: 9

Valid cases: 9745497

Invalid: 0

Description

The date on which the education status was recorded.

Format yyyy-MM-dd in csv file

MaritalStatusObsDate (MaritalStatusObsDate)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 10627459
Format: character	Invalid: 0
Width: 9	

Description

The date on which the marital status was recorded
 Format yyyy-MM-dd in csv file

SEObsDate (SEObsDate)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 10404288
Format: character	Invalid: 0
Width: 9	

Description

The date on which the socio-economic status was recorded
 Format yyyy-MM-dd in csv file

EmploymentSector (EmploymentSector)

File: SISEDD2020V1

Overview

Type: Continuous	Valid cases: 12649528
Format: numeric	Invalid: 467633
Decimals: 0	
Range: 0-306	

Description

Employment sector of the individual during this episode

EmploymentObsDate (EmploymentObsDate)

File: SISEDD2020V1

Overview

Type: Discrete	Valid cases: 12649528
Format: character	Invalid: 0
Width: 9	

Description

Date on which the employment status was recorded
 Format yyyy-MM-dd in csv file

