South Africa - ZA011 Excess Mortality: Surveillance Episodes Datasets

Prof Steve Tollman - Agincourt
Prof Kathleen Khan - Agincourt
Dr Kobus Herbst - SAPRIN
Dr Chodziwadziwa Kabudula - Agincourt
Dr Beth Tippett-Barr - Nyanja Health Research Institute

Report generated on: December 22, 2023

Visit our data catalog at: https://data.agincourt.co.za/index.php

Overview

Identification

ID NUMBER
ZA011-EXMORTALITY-01

Version

VERSION DESCRIPTION
v1: Dataset for public distribution.

PRODUCTION DATE
2023-06-01

NOTES
v1: Dataset for public distribution.

Overview

ABSTRACT
Following the declaration of COVID-19 as a pandemic by the World Health Organization, there have been high levels of reported deaths, at least in countries with functioning civil registration and vital statistics (CRVS). These largely under-represent the true mortality due to COVID-19. A fundamental question, then, is what is the impact of COVID-19 on mortality and the scale of excess deaths, and the population sub-groups most affected, particularly in low- and middle-income settings? Constructing a true representation of COVID-19 deaths can be useful for social policies and future pandemic preparedness planning. The goal of this initiative is to characterise all-cause mortality rates and trends, by age and sex, across a range of rural and urban sub-Saharan African and South Asian settings under continuous health and demographic surveillance.

This a multinational initiative bringing together 17 sites/centres from Africa and South Asia. This dataset represents a snapshot of the continually evolving data in the underlying longitudinal databases maintained by the nodes.

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

Topic	Vocabulary	URI
Episodes, Mortality, Migration		

KEYWORDS
Episodes, Mortality, Migration

Coverage

GEOGRAPHIC COVERAGE
The MRC/Wits University Agincourt HDSS in Bushbuckridge District, Mpumalanga, which has collected data since 1993. The nodal website is: http://www.agincourt.co.za. 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.

UNIVERSE
Households resident in dwellings within the study area will be eligible for inclusion in the surveillance. 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 Steve Tollman	Agincourt
Prof Kathleen Khan	Agincourt
Dr Kobus Herbst	SAPRIN
Dr Chodziwadziwa Kabudula	Agincourt
Dr Beth Tippett-Barr	Nyanja Health Research Institute

OTHER PRODUCER(S)

Name	Affiliation	Role
Chodziwadziwa Kabudula	Agincourt	Technical Assistance
Kobus Herbst	SAPRIN	Technical Assistance
Daniel Ohene-Kwofie	Agincourt	Technical Assistance
Jean Bashingwa	Agincourt	Technical Assistance
Nkosinathi Masilela	Agincourt	Technical Assistance
Rhulani Silaule	Agincourt	Technical Assistance

FUNDING

Name	Abbreviation	Role

Name	Abbreviation	Role
Bill & Melinda Gates Foundation, Seattle, WA	BMGF	Current Funder
South African Population Research Infrastructure Network, South Africa	SAPRIN	Current Funder
Wellcome Trust, UK	Wellcome	Previous Funder

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
SAMRC/Wits Agincourt Team	Agincourt	Data Review
Kobus Herbst	SAPRIN	Data Review And QA
CHAMPS Team	CHAMPS	Reviews

DDI DOCUMENT ID
ZA011-EXMORTALITY-01

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
2015-01-01	2021-12-31	Agincourt

Time Periods

Start	End	Cycle
2015-01-01		Agincourt

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

ZA011_RawCensoredEpisodes

Content	
Cases	172081
Variable(s)	13
Structure	Type: Keys: ()
Version	
Producer	
Missing Data	

Variables

ID	Name	Label	Type	Format	Question
V132	recnr	RecNr	contin	numeric	
V133	countryid	CountryId	discrete	numeric	
V134	centreid	CentreId	discrete	character	
V135	individualid	IndividualId	discrete	character	
V136	sex	Sex	discrete	numeric	
V137	dob	DoB	discrete	character	
V138	episode	Episode	discrete	numeric	
V139	episodes	Episodes	discrete	numeric	
V140	startevent	StartEvent	discrete	character	
V141	startdate	Start Date	discrete	character	
V142	endevent	EndEvent	discrete	character	
V143	enddate	End Date	discrete	character	
V144	days	Days	contin	numeric	

RecNr (recnr)
File: ZA011_RawCensoredEpisodes

Overview

Type: Continuous
Format: numeric
Width: 12
Decimals: 0
Range: 1-172081

Valid cases: 172081
Invalid: 0
Minimum: 1
Maximum: 172081
Mean: 86041
Standard deviation: 49675.7

CountryId (countryid)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: numeric
Width: 8
Decimals: 0
Range: 710-710

Valid cases: 172081
Invalid: 0

CentreId (centreid)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 5

Valid cases: 172081
Invalid: 0

IndividualId (individualid)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 5

Valid cases: 172081
Invalid: 0

Sex (sex)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: numeric
Width: 8
Decimals: 0
Range: 1-3

Valid cases: 172081
Invalid: 0

DoB (dob)
File: ZA011_RawCensoredEpisodes

Overview

DoB (dob)
File: ZA011_RawCensoredEpisodes

Type: Discrete
Format: character
Width: 11

Valid cases: 172081
Minimum: NaN
Maximum: NaN

Episode (episode)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: numeric
Width: 8
Decimals: 0
Range: 1-2

Valid cases: 172081
Invalid: 0

Episodes (episodes)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: numeric
Width: 8
Decimals: 0
Range: 1-2

Valid cases: 172081
Invalid: 0

StartEvent (startevent)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 3

Valid cases: 172081
Invalid: 0

Start Date (startdate)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 172081
Minimum: NaN
Maximum: NaN

EndEvent (endevent)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 3

Valid cases: 172081
Invalid: 0

15

End Date (enddate)
File: ZA011_RawCensoredEpisodes

Overview

Type: Discrete
Format: character
Width: 11

Valid cases: 172081
Minimum: NaN
Maximum: NaN

Days (days)
File: ZA011_RawCensoredEpisodes

Overview

Type: Continuous
Format: numeric
Width: 8
Decimals: 0
Range: 1-2922

Valid cases: 172081
Invalid: 0
Minimum: 1
Maximum: 2922
Mean: 1961.1
Standard deviation: 1015.4