

# South Africa - ZA081 Excess Mortality: Surveillance Episodes Datasets

**Prof Shabir Madhi - Wits VIDA**  
**Prof Ziyaad Dangor - Wits VIDA**

Report generated on: March 15, 2024

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

## Overview

### Identification

ID NUMBER  
ZA081-EXMORTALITY-02

### 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 data are collected from the Soweto and Thembelihle health and socio-demographic surveillance area in a peri-urban township and informal settlement, South Africa. The surveillance area is located approximately 20 km south-west of the City of Johannesburg, covering an area of 36.7 square km.

## 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 Shabir Madhi	Wits VIDA
Prof Ziyaad Dangor	Wits VIDA

## OTHER PRODUCER(S)

Name	Affiliation	Role
Dr Takwanisa Machemedze	Wits VIDA	Technical Assistance
Mr Cleopas Hwinya	Wits VIDA	Technical Assistance
Mr Carlos Herrera	Wits VIDA	Technical Assistance

## FUNDING

Name	Abbreviation	Role
Bill & Melinda Gates Foundation, Seattle, WA	BMGF	Current Funder

## OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Wits VIDA Data Team	Wits VIDA	Providing Data
Solveig Argešanu Cunningham, Jonathan Muir	Programme Office	Technical support
SAMRC/Wits Agincourt Team	Agincourt	Data Review
Kobus Herbst	SAPRIN	Data Review And QA
Dr Beth Tippet-Barr	Nyanja Health Research Institute	Reviews

## DDI DOCUMENT ID

ZA081-EXMORTALITY-02

## 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
2017-11-07	2021-12-31	1

### Time Periods

---

Start	End	Cycle
2017-11-07		1

### 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.

### 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 Data 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 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 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 Data Management hub is satisfied that they have high quality data.

## Data Appraisal

### **Estimates of Sampling Error**

Not Applicable



## File Description

## Variable List

## ZA081\_RawCensoredEpisodes

Content

Cases 75981

Variable(s) 13

Structure Type:  
Keys: ()

Version

Producer

Missing Data

## Variables

ID	Name	Label	Type	Format	Question
V14	recnr	RecNr	contin	numeric	
V15	countryid	CountryId	discrete	numeric	
V16	centreid	CentreId	discrete	character	
V17	individualid	IndividualId	discrete	character	
V18	sex	Sex	discrete	numeric	
V19	dob	DoB	discrete	character	
V20	episode	Episode	discrete	numeric	
V21	episodes	Episodes	discrete	numeric	
V22	startevent	StartEvent	discrete	character	
V23	startdate	Start Date	discrete	character	
V24	endevent	EndEvent	discrete	character	
V25	enddate	End Date	discrete	character	
V26	days	Days	contin	numeric	



## RecNr (recnr)

File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Continuous	Valid cases: 75981
Format: numeric	Invalid: 0
Width: 12	Minimum: 1
Decimals: 0	Maximum: 75981
Range: 1-75981	Mean: 37991
	Standard deviation: 21934

## CountryId (countryid)

File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete	Valid cases: 75981
Format: numeric	Invalid: 0
Width: 8	
Decimals: 0	
Range: 710-710	

## CentreId (centreid)

File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete	Valid cases: 75981
Format: character	Invalid: 0
Width: 5	

## IndividualId (individualid)

File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete	Valid cases: 75981
Format: character	Invalid: 0
Width: 12	

## Sex (sex)

File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete	Valid cases: 75981
Format: numeric	Invalid: 0
Width: 8	
Decimals: 0	
Range: 1-2	

## DoB (dob)

File: ZA081\_RawCensoredEpisodes

**Overview**

## DoB (dob)

## File: ZA081\_RawCensoredEpisodes

Type: Discrete  
 Format: character  
 Width: 11

Valid cases: 75981  
 Minimum: NaN  
 Maximum: NaN

## Episode (episode)

## File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 8  
 Decimals: 0  
 Range: 1-4

Valid cases: 75981  
 Invalid: 0

## Episodes (episodes)

## File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 8  
 Decimals: 0  
 Range: 1-4

Valid cases: 75981  
 Invalid: 0

## StartEvent (startevent)

## File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete  
 Format: character  
 Width: 3

Valid cases: 75981  
 Invalid: 0

## Start Date (startdate)

## File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete  
 Format: character  
 Width: 11

Valid cases: 75981  
 Minimum: NaN  
 Maximum: NaN

## EndEvent (endevent)

## File: ZA081\_RawCensoredEpisodes

**Overview**

Type: Discrete  
 Format: character  
 Width: 3

Valid cases: 75981  
 Invalid: 0

End Date (enddate)

File: ZA081\_RawCensoredEpisodes

#### Overview

Type: Discrete  
Format: character  
Width: 11

Valid cases: 75981  
Minimum: NaN  
Maximum: NaN

Days (days)

File: ZA081\_RawCensoredEpisodes

#### Overview

Type: Continuous  
Format: numeric  
Width: 8  
Decimals: 0  
Range: 1-1864

Valid cases: 75981  
Invalid: 0  
Minimum: 1  
Maximum: 1864  
Mean: 1178.5  
Standard deviation: 542.2

