

# ALPHA HIV Incidence and Mortality Data 2000-2016, uMkhanyakude

**Africa Health Research Institute, Network for Analysing Longitudinal Population  
HIV/ AIDS data on Africa**

report\_generated\_on: May 29, 2026

visit\_data\_catalog\_at: <https://microdata.worldbank.org/>

## Identification

---

### SURVEY ID NUMBER

ZAF\_2000-2016\_ALPHA-HIV-IM\_v01\_M

### TITLE

ALPHA HIV Incidence and Mortality Data 2000-2016, uMkhanyakude

### ABBREVIATION OR ACRONYM

ALPHA-HIV-IM 2000-2016

### COUNTRY/ECONOMY

Name	Country code
South Africa	ZAF

### STUDY TYPE

Other Household Survey [hh/oth]

### SERIES INFORMATION

The ALPHA (Analysing Longitudinal Population-based HIV/AIDS Data in Africa) Network is a collaboration among 10 longitudinal studies in sub-Saharan Africa. These studies collect data on HIV infection alongside demographic, behavioural, socio-economic and clinical data from residents of the study areas. The Network harmonises these data and conducts comparable and pooled analyses on HIV-related research questions. The ALPHA HIV Incidence and Mortality data, uMkhanyakude is part of the international data harmonisation and analysis programme of the ALPHA network.

### ABSTRACT

This dataset is created from harmonizing longitudinal population-based demographic surveillance and repeated serological survey data collected by the uMkhanyakude study in South Africa. The data harmonization is coordinated by the ALPHA Network which curates individual-level data on demographic surveillance, verbal autopsy interviews, serological and sexual behavior surveys, and individually-linked data from HDSS and medical facilities.

The data from the Umkhanyakude site in South Africa includes two data files:

- HIV incidence data

This file includes data on resident study participants aged 15-49 who have had one negative HIV test and at least one subsequent HIV test done within the study for research purposes. Data may be split into multiple observations for each person to describe time spent in different residencies, different age groups, and different calendar year periods. All timings are given as age at event, instead of dates. Dates of birth are rounded to the 15th of the month, and all other dates are offset by the same amount and the time elapsed between the rounded dates is used as the age at event. To enable multiple imputation of seroconversion dates the times (ages) of last negative test and first positive test have been provided.

- Mortality data

This data file was created as described at <https://gatesopenresearch.org/articles/1-4>

### KIND OF DATA

Event/transaction data [evn]

### UNIT OF ANALYSIS

Individuals

## Version

---

### VERSION DESCRIPTION

Version 01: Edited, anonymised data available with restrictions.

## Scope

---

### NOTES

The dataset includes individual-level HIV data from demographic surveillance, verbal autopsy interviews, serological and sexual behaviour surveys, and individually-linked data from HDSS and medical facilities.

## Coverage

### GEOGRAPHIC COVERAGE

The data covers the uMkhanyakude study area in KwaZulu-Natal, South Africa

### GEOGRAPHIC UNIT

The data is at the level of village.

## Producers and sponsors

### PRIMARY INVESTIGATORS

Name	Affiliation
Africa Health Research Institute	
Network for Analysing Longitudinal Population HIV/ AIDS data on Africa	London School of Hygiene and Tropical Medicine

### FUNDING AGENCY/SPONSOR

Name	Role
Wellcome Trust	Funding agency
Bill and Melinda Gates Foundation	Funding agency

### OTHER IDENTIFICATIONS/ACKNOWLEDGMENTS

Name	Affiliation	Role
Dr Kobus Herbst	AHRI	Director of Population Science
Dickman Gareta	AHRI	Data Scientist

## Data collection

### DATES OF DATA COLLECTION

Start	End	Cycle
2000	2016	Incidence data
2000	2016	Mortality data

### DATA COLLECTION MODE

Face-to-face

### DATA COLLECTION NOTES

ALPHA data are harmonised datasets from ALPHA member sites and this dataset is from the Africa Health Research Institute.

### DATA COLLECTORS

Name
Africa Health Research Institute

## Access policy

### ACCESS CONDITIONS

Access under a Creative Commons CC-BY-NC (Attribution, Non-Commercial use only) License

### CITATION REQUIREMENTS

Herbst, K. et al. ALPHA HIV Incidence and Mortality Data, uMkhanyakude 2000-2016 [dataset].Version 1. Africa Health Research Institute [producer and distributor], 2022. <https://data.ahri.org/index.php/catalog/1029>

### ACCESS AUTHORITY

Name	URL
Africa Health Research Institute	<a href="#">Link</a>

### ARCHIVE WHERE STUDY IS ORIGINALLY STORED

Karonga (MW011)

## Metadata production

### DDI DOCUMENT ID

DDI\_ZAF\_2000-2016\_ALPHA-HIV-IM\_v01\_M

### PRODUCERS

Name	Abbreviation	Affiliation	Role
Africa Health Research Institute	AHRI		Metadata creator
Network for Analysing Longitudinal Population HIV/ AIDS data on Africa	ALPHA	London School of Hygiene and Tropical Medicine	Metadata creator
DataFirst		University of Cape Town	Metadata creator
Development Data Group	DECDG	World Bank Group	Metadata adapted for World Bank Microdata Library

### DATE OF METADATA PRODUCTION

2024-08-23

### DDI DOCUMENT VERSION

Identical to a metadata (zaf-alpha-himu-2000-2016-v1) published on DataFirst microdata repository (<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog>). Some of the metadata fields have been edited.

**data\_dictionary**

<b>Data file</b>	<b>Cases</b>	<b>variables</b>
<b>incidence_uMkhanyakude</b>	0	15
<b>mortality_uMkhanyakude</b>	0	14



**Data file: incidence\_uMkhanyakude**

Cases:	0
variables:	15

**variables**

ID	Name	Label	Question
V1	study_name	Study name	
V2	idno	Participant ID number	
V3	sex	Sex	
V4	fouryear	Calendar year, grouped in 4 years post 2005	
V5	agegrp	Five year age group	
V6	timein	Start of episode (age)	
V7	timeout	End of episode (age)	
V8	timelastneg	Age at last negative test	
V9	timefirstpos	Age at first positive test	
V10	fail	Seroconversion occurred at the end of this episode	
V11	t0	Stata variable: age at start of episode	
V12	t	Stata variable: age at end of episode	
V13	st	Stata variable: in survival analysis	
V14	d	Stata variable: failure (seroconversion)	
V15	collapsegrp		

total: 15

**Data file: mortality\_uMkhanyakude**

Cases:	0
variables:	14

**variables**

ID	Name	Label	Question
V16	study_name	Study name	
V17	idno	Participant ID number	
V18	sex	Sex	
V19	fouryear	Calendar year, grouped in 4 years post 2005	
V20	agegrp	Five year age group	
V21	hivstatus_broad	HIV status	
V22	timein	Start of episode (age)	
V23	timeout	End of episode (age)	
V24	fail	Seroconversion occurred at the end of this episode	
V25	t0	Stata variable: age at start of episode	
V26	t	Stata variable: age at end of episode	
V27	st	Stata variable: in survival analysis	
V28	d	Stata variable: failure (seroconversion)	
V29	collapsegrp		

total: 14



**STUDY\_NAME: Study name****Data file: incidence\_uMkhanyakude****Overview**

Valid: Invalid:

Type: Discrete Width: 12 Range: - Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category
1	Karonga
2	Kisesa
3	Manicaland
4	Masaka
5	Rakai
6	uMkhanyakude
7	Agincourt
8	Kisumu
9	Ifakara
10	Nairobi

**FAIL: Seroconversion occurred at the end of this episode****Data file: incidence\_uMkhanyakude****Overview**

Valid: Invalid:

Type: Discrete Width: 9 Range: - Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category
	No
1	Yes

**IDNO: Participant ID number****Data file: incidence\_uMkhanyakude****Overview**

Valid: Invalid:

Type: Continuous    Width: 9    Range: -    Format: Numeric

## SEX: Sex

Data file: incidence\_uMkhanyakude

### Overview

Valid:    Invalid:

Type: Discrete    Width: 12    Range: -    Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category
1	Men
2	Women

## FOURYEAR: Calendar year, grouped in 4 years post 2005

Data file: incidence\_uMkhanyakude

### Overview

Valid:    Invalid:

Type: Discrete    Width: 9    Range: -    Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category
	1995-99
1	2000-04
2	2005-08
3	2009-12
4	2013-16
5	2017-20
6	2021-24

## AGEGRP: Five year age group

Data file: incidence\_uMkhanyakude

### Overview

Valid:    Invalid:

Type: Discrete    Width: 9    Range: -    Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category
	0-4
1	5-9
2	10-14
3	15-19
4	20-24
5	25-29
6	30-34
7	35-39
8	40-44
9	45-49
10	50-54
11	55-59
12	60-64
13	65-69
14	70-74
15	75-79
16	80-84
17	85-89
18	90+

### TIMEIN: Start of episode (age)

Data file: incidence\_uMkhanyakude

#### Overview

Valid: Invalid:  
Type: Continuous Width: 10 Range: - Format: Numeric

### TIMEOUT: End of episode (age)

Data file: incidence\_uMkhanyakude

#### Overview

Valid: Invalid:  
Type: Continuous Width: 10 Range: - Format: Numeric

**TIMELASTNEG: Age at last negative test****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Continuous   Width: 10   Range: -   Format: Numeric

---

**TIMEFIRSTPOS: Age at first positive test****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Continuous   Width: 10   Range: -   Format: Numeric

---

**T0: Stata variable: age at start of episode****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Continuous   Width: 10   Range: -   Format: Numeric

---

**T: Stata variable: age at end of episode****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Continuous   Width: 10   Range: -   Format: Numeric

---

**ST: Stata variable: in survival analysis****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Discrete   Width: 8   Range: -   Format: Numeric

---

**D: Stata variable: failure (seroconversion)****Data file:** incidence\_uMkhanyakude**Overview**

Valid:   Invalid:  
 Type: Discrete   Width: 8   Range: -   Format: Numeric

**COLLAPSEGRP:**

**Data file: incidence\_uMkhanyakude**

**Overview**

Valid:    Invalid:  
Type: Continuous    Width: 9    Range: -    Format: Numeric

---

**STUDY\_NAME: Study name****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:

Type: Discrete Width: 12 Range: - Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category
1	Karonga
2	Kisesa
3	Manicaland
4	Masaka
5	Rakai
6	uMkhanyakude
7	Agincourt
8	Kisumu
9	Ifakara
10	Nairobi

**IDNO: Participant ID number****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:

Type: Continuous Width: 9 Range: - Format: Numeric

**SEX: Sex****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:

Type: Discrete Width: 8 Range: - Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category
1	Men

2	Women
---	-------

## FOURYEAR: Calendar year, grouped in 4 years post 2005

Data file: mortality\_uMkhanyakude

### Overview

Valid: Invalid:

Type: Discrete Width: 13 Range: - Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category
	earliest-1999
1	2000-04
2	2005-08
3	2009-12
4	2013-16
5	2017-20
6	2021-24

## AGEGRP: Five year age group

Data file: mortality\_uMkhanyakude

### Overview

Valid: Invalid:

Type: Discrete Width: 9 Range: - Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category
	0-4
1	5-9
2	10-14
3	15-19
4	20-24
5	25-29
6	30-34
7	35-39

8	40-44
9	45-49
10	50-54
11	55-59
12	60-64
13	65-69
14	70-74
15	75-79
16	80-84
17	85-89
18	90+

### HIVSTATUS\_BROAD: HIV status

Data file: mortality\_uMkhanyakude

#### Overview

Valid:    Invalid:  
 Type: Discrete    Width: 9    Range: -    Format: Numeric

#### Questions and instructions

#### CATEGORIES

Value	Category
1	Negative
2	Positive
3	Unknown

### TIMEIN: Start of episode (age)

Data file: mortality\_uMkhanyakude

#### Overview

Valid:    Invalid:  
 Type: Continuous    Width: 10    Range: -    Format: Numeric

### TIMEOUT: End of episode (age)

Data file: mortality\_uMkhanyakude

#### Overview

Valid:    Invalid:  
 Type: Continuous    Width: 10    Range: -    Format: Numeric

**FAIL: Seroconversion occurred at the end of this episode****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:  
 Type: Discrete Width: 9 Range: - Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category
	No
1	Yes

**T0: Stata variable: age at start of episode****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:  
 Type: Continuous Width: 10 Range: - Format: Numeric

**T: Stata variable: age at end of episode****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:  
 Type: Continuous Width: 10 Range: - Format: Numeric

**ST: Stata variable: in survival analysis****Data file:** mortality\_uMkhanyakude**Overview**

Valid: Invalid:  
 Type: Discrete Width: 8 Range: - Format: Numeric

**D: Stata variable: failure (seroconversion)****Data file:** mortality\_uMkhanyakude

## Overview

Valid:   Invalid:  
Type: Discrete   Width: 8   Range: -   Format: Numeric

---

## COLLAPSEGRP:

**Data file: mortality\_uMkhanyakude**

## Overview

Valid:   Invalid:  
Type: Continuous   Width: 9   Range: -   Format: Numeric

---

## **study\_resources**