

# Tanzania - ALPHA HIV Incidence and Mortality Data, Kisesa 1994-2016

**Kisesa HIV Open Cohort, Network for Analysing Longitudinal Population HIV/ AIDS data on Africa**

Report generated on: June 12, 2022

Visit our data portal at: <https://www.datafirst.uct.ac.za/dataportal/index.php>

## Overview

### Identification

---

#### ID NUMBER

tza-alpha-himk-1996-2016-v1

### Version

---

#### VERSION DESCRIPTION

Version 1: Edited, anonymised data available with restrictions

#### PRODUCTION DATE

2022

## Overview

---

#### ABSTRACT

This dataset is created from harmonising longitudinal population-based demographic surveillance and repeated serological survey data collected by an Kisesa HIV open cohort Study in Mwanza, Tanzania. The data harmonisation is coordinated by the ALPHA Network which curates individual-level data on demographic surveillance, verbal autopsy interviews, serological and sexual behaviour surveys, and individually-linked data from HDSS and medical facilities.

The data from the Kisesa site in Tanzania includes two data files:

HIV incidence data data file:

This data 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 the dates. Dates of birth were rounded to the 15th of the month, and all other dates were 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 we have provided the times (ages) of last negative test and first positive test.

Mortality data

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

#### KIND OF DATA

Event History data

#### UNITS OF ANALYSIS

Individuals

## Scope

---

#### NOTES

The data covers demographic surveillance and repeated serological survey data from the Kisesa site in Mwanza, Tanzania, collected through demographic surveillance, verbal autopsy interviews, serological and sexual behaviour surveys. Also included is data from HDSS and medical facilities.

## Coverage

### GEOGRAPHIC COVERAGE

The data covers the Kisesa HIV open cohort study area in Mwanza, Tanzania

### GEOGRAPHIC UNIT

The lowest level of geographic aggregation of the data is village.

## Producers and Sponsors

### PRIMARY INVESTIGATOR(S)

Name	Affiliation
Kisesa HIV Open Cohort	National Institute for Medical Research, Mwanza
Network for Analysing Longitudinal Population HIV/ AIDS data on Africa	London School of Hygiene and Tropical Medicine

### OTHER PRODUCER(S)

Name	Affiliation	Role

### FUNDING

Name	Abbreviation	Role
The Global Fund to Fight AIDS, Tuberculosis and Malaria		
Wellcome Trust		
Bill and Melinda Gates Foundation		

### OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Mark Urassa	National Institute for Medical Research, Mwanza	Demographer and Kisesa Study Director
Dr David Beckles		Database programmer, consultant
Prof Basia Zaba		Demographer, former ALPHA network lead
Prof Jim Todd	LSHTM	Professor of statistics and technical support
Baltazar Mtenga	National Institute for Medical Research, Mwanza	Data manager and programmer
Dr Michael Denna	National Institute for Medical Research, Mwanza	Epidemiologist
Dr Milly Marston	LSHTM	Demographer, Technical support
Dr Emma Slaymaker	LSHTM	Epidemiologist, Technical support

## Metadata Production

### METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Kisesa Open HIV Cohort study	Kisesa	National Institute for Medical Research	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

DATE OF METADATA PRODUCTION  
2022-03-31

DDI DOCUMENT VERSION  
v1

## Sampling

No content available

# Questionnaires

## Overview

---

The original unharmonised data was collected by the Kisesa study

## Data Collection

### Data Collection Dates

Start	End	Cycle
1994	2016	Mortality data
1996	2015	Incidence 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 Kisesa study site.

### Questionnaires

The original unharmonised data was collected by the Kisesa study

### Data Collectors

Name	Abbreviation	Affiliation
Kisesa HIV Open Cohort		National Institute for Medical Research, Mwanza

## Data Processing

No content available



## Data Appraisal

No content available

## File Description

## Variable List

## mortality-kisesa

### Content

Cases	186358
Variable(s)	13
Structure	Type: relational Keys: idno(Participant ID number for sharing)
Version	
Producer	Malawi Epidemiology and Interventions Research Unit
Missing Data	

## Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V1	study_name	Study name	discrete	numeric	
V32	idno	Participant ID number for sharing	discrete	numeric	
V4	sex	Sex	discrete	numeric	
V5	fouryear	Calendar year, grouped in 4 years post 2005	discrete	numeric	
V6	agegrp	Five year age group	discrete	numeric	
V7	hivstatus_broad	HIV status	discrete	numeric	
V8	timein	Start of episode (age)	contin	numeric	
V9	timeout	End of episode (age)	contin	numeric	
V10	fail	Seroconversion occurred at the end of this episode	discrete	numeric	
V11	t0	Stata variable: age at start of episode	contin	numeric	
V12	t	Stata variable: age at end of episode	contin	numeric	
V13	st	Stata variable: in survival analysis	discrete	numeric	
V14	d	Stata variable: failure (seroconversion)	discrete	numeric	

## incidence-kisesa

### Content

Cases	43798
Variable(s)	14
Structure	Type: relational Keys: idno(Participant ID number for sharing)
Version	
Producer	Malawi Epidemiology and Interventions Research Unit
Missing Data	

## Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V16	study_name	Study name	discrete	numeric	
V33	idno	Participant ID number for sharing	discrete	numeric	
V19	sex	Sex	discrete	numeric	
V20	fouryear	Calendar year, grouped in 4 years post 2005	discrete	numeric	
V21	agegrp	Five year age group	discrete	numeric	
V22	timein	Start of episode (age)	contin	numeric	
V23	timeout	End of episode (age)	contin	numeric	
V24	timelastneg	Age at last negative test	contin	numeric	
V25	timefirstpos	Age at first positive test	contin	numeric	
V26	fail	Seroconversion occurred at the end of this episode	discrete	numeric	
V27	t0	Stata variable: age at start of episode	contin	numeric	
V28	t	Stata variable: age at end of episode	contin	numeric	
V29	st	Stata variable: in survival analysis	discrete	numeric	
V30	d	Stata variable: failure (seroconversion)	discrete	numeric	



## Study name (study\_name)

File: mortality-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 2  
 Decimals: 0  
 Range: 1-10

Valid cases: 186358  
 Invalid: 0

## Participant ID number for sharing (idno)

File: mortality-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 5  
 Decimals: 0

Valid cases: 186358  
 Invalid: 0

## Sex (sex)

File: mortality-kisesa

**Overview**

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

Valid cases: 186358  
 Invalid: 0

## Calendar year, grouped in 4 years post 2005 (fouryear)

File: mortality-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 1  
 Decimals: 0  
 Range: 0-6

Valid cases: 186358  
 Invalid: 0

## Five year age group (agegrp)

File: mortality-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 2  
 Decimals: 0  
 Range: 0-18

Valid cases: 186358  
 Invalid: 0

## HIV status (hivstatus\_broad)

File: mortality-kisesa

**Overview**

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

Valid cases: 179643  
 Invalid: 6715

Start of episode (age) (timein)  
 File: mortality-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15-49.9986310746064

Valid cases: 186358  
 Invalid: 0  
 Minimum: 15  
 Maximum: 50  
 Mean: 27.2  
 Standard deviation: 9.3

End of episode (age) (timeout)  
 File: mortality-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15.0006844626968-50

Valid cases: 186358  
 Invalid: 0  
 Minimum: 15  
 Maximum: 50  
 Mean: 28.5  
 Standard deviation: 9.4

Seroconversion occurred at the end of this episode (fail)  
 File: mortality-kisesa

**Overview**

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

Valid cases: 185292  
 Invalid: 1066

Stata variable: age at start of episode (t0)  
 File: mortality-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15-49.9986310746064

Valid cases: 186358  
 Invalid: 0  
 Minimum: 15  
 Maximum: 50  
 Mean: 27.2  
 Standard deviation: 9.3



Stata variable: age at end of episode (t)

File: mortality-kisesa

#### Overview

Type: Continuous	Valid cases: 186358
Format: numeric	Invalid: 0
Width: 16	Minimum: 15
Decimals: 0	Maximum: 50
Range: 15.0006844626968-50	Mean: 28.5
	Standard deviation: 9.4

Stata variable: in survival analysis (st)

File: mortality-kisesa

#### Overview

Type: Discrete	Valid cases: 186358
Format: numeric	Invalid: 0
Width: 1	
Decimals: 0	
Range: 1-1	

Stata variable: failure (seroconversion) (d)

File: mortality-kisesa

#### Overview

Type: Discrete	Valid cases: 186358
Format: numeric	Invalid: 0
Width: 1	
Decimals: 0	
Range: 0-1	

## Study name (study\_name)

File: incidence-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 2  
 Decimals: 0  
 Range: 1-10

Valid cases: 43798  
 Invalid: 0

## Participant ID number for sharing (idno)

File: incidence-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 5  
 Decimals: 0

Valid cases: 43798  
 Invalid: 0

## Sex (sex)

File: incidence-kisesa

**Overview**

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

Valid cases: 43798  
 Invalid: 0

## Calendar year, grouped in 4 years post 2005 (fouryear)

File: incidence-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 1  
 Decimals: 0  
 Range: 0-6

Valid cases: 43798  
 Invalid: 0

## Five year age group (agegrp)

File: incidence-kisesa

**Overview**

Type: Discrete  
 Format: numeric  
 Width: 2  
 Decimals: 0  
 Range: 0-18

Valid cases: 43798  
 Invalid: 0

## Start of episode (age) (timein)

File: incidence-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15-49.9876796714579

Valid cases: 43798  
 Invalid: 0  
 Minimum: 15  
 Maximum: 50  
 Mean: 28.6  
 Standard deviation: 9.9

End of episode (age) (timeout)

File: incidence-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15.0006844626968-50

Valid cases: 43798  
 Invalid: 0  
 Minimum: 15  
 Maximum: 50  
 Mean: 30.2  
 Standard deviation: 10

Age at last negative test (timelastneg)

File: incidence-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 13.5496235455168-68.5859000684463

Valid cases: 43774  
 Invalid: 24  
 Minimum: 13.6  
 Maximum: 68.6  
 Mean: 34.5  
 Standard deviation: 12

Age at first positive test (timefirstpos)

File: incidence-kisesa

**Overview**

Type: Continuous  
 Format: numeric  
 Width: 16  
 Decimals: 0  
 Range: 15.0225872689938-63.460643394935

Valid cases: 2149  
 Invalid: 41649  
 Minimum: 15  
 Maximum: 63.5  
 Mean: 35.8  
 Standard deviation: 9.6

Seroconversion occurred at the end of this episode (fail)

File: incidence-kisesa

**Overview**

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

Valid cases: 43638  
 Invalid: 160

## Stata variable: age at start of episode (t0)

File: incidence-kisesa

**Overview**

Type: Continuous	Valid cases: 43798
Format: numeric	Invalid: 0
Width: 16	Minimum: 15
Decimals: 0	Maximum: 50
Range: 15-49.9876796714579	Mean: 28.6
	Standard deviation: 9.9

## Stata variable: age at end of episode (t)

File: incidence-kisesa

**Overview**

Type: Continuous	Valid cases: 43798
Format: numeric	Invalid: 0
Width: 16	Minimum: 15
Decimals: 0	Maximum: 50
Range: 15.0006844626968-50	Mean: 30.2
	Standard deviation: 10

## Stata variable: in survival analysis (st)

File: incidence-kisesa

**Overview**

Type: Discrete	Valid cases: 43798
Format: numeric	Invalid: 0
Width: 1	
Decimals: 0	
Range: 1-1	

## Stata variable: failure (seroconversion) (d)

File: incidence-kisesa

**Overview**

Type: Discrete	Valid cases: 43798
Format: numeric	Invalid: 0
Width: 1	
Decimals: 0	
Range: 0-1	

# Documentation

## Other material

### Contextual documentation for the shared ALPHA data Kisesa

---

Title	Contextual documentation for the shared ALPHA data Kisesa
Date	2022-04-25
Country	Tanzania
Language	English
Contributor(s)	This document provided background information on the dataset
Filename	alpha-kisesa-1994-2016.pdf

---