

ALPHA HIV Incidence and Mortality Data 1994-2016, Kisesa

**Kisesa HIV Open Cohort, 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

TZA_1994-2016_ALPHA-HIV-IM_v01_M

TITLE

ALPHA HIV Incidence and Mortality Data 1994-2016, Kisesa

ABBREVIATION OR ACRONYM

ALPHA-HIV-IM 1994-2016

COUNTRY/ECONOMY

Name	Country code
Tanzania	TAZ

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, Kisesa 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 an Kisesa HIV open cohort Study in Mwanza, Tanzania. 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 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/transaction data [evn]

UNIT OF ANALYSIS

Individuals

Version

VERSION DESCRIPTION

Version 01: Edited, anonymised data available with restrictions.

VERSION DATE

2022

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 INVESTIGATORS

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

FUNDING AGENCY/SPONSOR

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

OTHER IDENTIFICATIONS/ACKNOWLEDGMENTS

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

Data collection

DATES OF DATA COLLECTION

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.

DATA COLLECTORS

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

questionnaires

QUESTIONNAIRES

The original unharmonised data was collected by the Kisesa study

Access policy

CONTACTS

Name	Affiliation	Email	URL
DataFirst helpdesk	University of Cape Town	support@data1st.org	www.support.data1st.org

ACCESS CONDITIONS

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

CITATION REQUIREMENTS

Kisesa HIV Open Cohort study and ALPHA Network. ALPHA HIV Incidence and Mortality Data, 1994-2016, Kisesa, Tanzania [dataset]. Version 1. Kisesa and London: Kisesa HIV Open Cohort study and LSHTM [producers], 2016. Cape Town: DataFirst [distributors], 2022. DOI: <https://doi.org/10.25828/SGBS-Q698>

ACCESS AUTHORITY

Name	Affiliation	Email	URL
DataFirst	University of Cape Town	support@data1st.org	www.datafirst.uct.ac.za

ARCHIVE WHERE STUDY IS ORIGINALLY STORED

Karonga (MW011)

Metadata production

DDI DOCUMENT ID

DDI_TZA_1994-2016_ALPHA-HIV-IM_v01_M

PRODUCERS

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

Development Data Group	DECDG	World Bank Group	Metadata adapted for World Bank Microdata Library
------------------------	-------	------------------	---

DATE OF METADATA PRODUCTION
2022-03-31

DDI DOCUMENT VERSION

Identical to a metadata (tza-alpha-himk-1996-2016-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

Data file	Cases	variables
mortality-kisesa	186358	13
incidence-kisesa	43798	14

Data file: mortality-kisesa

Cases: 186358

variables: 13

variables

ID	Name	Label	Question
V1	study_name	Study name	
V32	idno	Participant ID number for sharing	
V4	sex	Sex	
V5	fouryear	Calendar year, grouped in 4 years post 2005	
V6	agegrp	Five year age group	
V7	hivstatus_broad	HIV status	
V8	timein	Start of episode (age)	
V9	timeout	End of episode (age)	
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)	

total: 13

Data file: incidence-kisumu

Cases: 43798

variables: 14

variables

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

total: 14

STUDY_NAME: Study name**Data file: mortality-kisesa****Overview**

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 2 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Karonga	0	0%
2	Kisesa	186358	100%
3	Manicaland	0	0%
4	Masaka	0	0%
5	Rakai	0	0%
6	uMkhanyakude	0	0%
7	Agincourt	0	0%
8	Kisumu	0	0%
9	Ifakara	0	0%
10	Nairobi	0	0%

FAIL: Seroconversion occurred at the end of this episode**Data file: mortality-kisesa****Overview**

Valid: 185292 Invalid: 1066
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	183785	99.2%
1	Yes	1507	0.8%
Sysmiss		1066	

T0: Stata variable: age at start of episode**Data file: mortality-kisesa**

Overview

Valid: 186358 Invalid: 0 Minimum: 15 Maximum: 49.999 Mean: 27.204 Standard deviation: 9.317
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

IDNO: Participant ID number for sharing

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 5 Format: Numeric Weighted variable: 0

SEX: Sex

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Men	86289	46.3%
2	Women	100061	53.7%
9		8	0%

FOURYEAR: Calendar year, grouped in 4 years post 2005

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	earliest-1999	39736	21.3%
1	2000-04	39862	21.4%
2	2005-08	32364	17.4%
3	2009-12	37740	20.3%

4	2013-16	36656	19.7%
5	2017-20	0	0%
6	2021-24	0	0%

AGEGRP: Five year age group

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 2 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	0-4	0	0%
1	5-9	0	0%
2	10-14	0	0%
3	15-19	47488	25.5%
4	20-24	38033	20.4%
5	25-29	30442	16.3%
6	30-34	25158	13.5%
7	35-39	19402	10.4%
8	40-44	14934	8%
9	45-49	10901	5.8%
10	50-54	0	0%
11	55-59	0	0%
12	60-64	0	0%
13	65-69	0	0%
14	70-74	0	0%
15	75-79	0	0%
16	80-84	0	0%
17	85-89	0	0%
18	90+	0	0%

HIVSTATUS_BROAD: HIV status

Data file: mortality-kisesa

Overview

Valid: 179643 Invalid: 6715
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Negative	66517	37%
2	Positive	4859	2.7%
3	Unknown	108267	60.3%
Sysmiss		6715	

TIMEIN: Start of episode (age)

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0 Minimum: 15 Maximum: 49.999 Mean: 27.204 Standard deviation: 9.317
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

TIMEOUT: End of episode (age)

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0 Minimum: 15.001 Maximum: 50 Mean: 28.523 Standard deviation: 9.448
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

T: Stata variable: age at end of episode

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0 Minimum: 15.001 Maximum: 50 Mean: 28.523 Standard deviation: 9.448
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

ST: Stata variable: in survival analysis

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1		186358	100%

D: Stata variable: failure (seroconversion)

Data file: mortality-kisesa

Overview

Valid: 186358 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0		184851	99.2%
1		1507	0.8%

STUDY_NAME: Study name**Data file: incidence-kisesa****Overview**

Valid: 43798 Invalid: 0
 Type: Discrete Decimal: 0 Width: 2 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Karonga	0	0%
2	Kisesa	43798	100%
3	Manicaland	0	0%
4	Masaka	0	0%
5	Rakai	0	0%
6	uMkhanyakude	0	0%
7	Agincourt	0	0%
8	Kisumu	0	0%
9	Ifakara	0	0%
10	Nairobi	0	0%

SEX: Sex**Data file: incidence-kisesa****Overview**

Valid: 43798 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
1	Men	19147	43.7%
2	Women	24651	56.3%

FOURYEAR: Calendar year, grouped in 4 years post 2005**Data file: incidence-kisesa****Overview**

Valid: 43798 Invalid: 0

Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	1995-99	10409	23.8%
1	2000-04	10958	25%
2	2005-08	8851	20.2%
3	2009-12	8044	18.4%
4	2013-16	5536	12.6%
5	2017-20	0	0%
6	2021-24	0	0%

AGEGRP: Five year age group

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0

Type: Discrete Decimal: 0 Width: 2 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	0-4	0	0%
1	5-9	0	0%
2	10-14	0	0%
3	15-19	9821	22.4%
4	20-24	7576	17.3%
5	25-29	6548	15%
6	30-34	6097	13.9%
7	35-39	5467	12.5%
8	40-44	4691	10.7%
9	45-49	3598	8.2%
10	50-54	0	0%
11	55-59	0	0%
12	60-64	0	0%
13	65-69	0	0%
14	70-74	0	0%

15	75-79	0	0%
16	80-84	0	0%
17	85-89	0	0%
18	90+	0	0%

TIMEIN: Start of episode (age)

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0 Minimum: 15 Maximum: 49.988 Mean: 28.58 Standard deviation: 9.866
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

TIMEOUT: End of episode (age)

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0 Minimum: 15.001 Maximum: 50 Mean: 30.249 Standard deviation: 10.036
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

TIMELASTNEG: Age at last negative test

Data file: incidence-kisesa

Overview

Valid: 43774 Invalid: 24 Minimum: 13.55 Maximum: 68.586 Mean: 34.534 Standard deviation: 11.969
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

TIMEFIRSTPOS: Age at first positive test

Data file: incidence-kisesa

Overview

Valid: 2149 Invalid: 41649 Minimum: 15.023 Maximum: 63.461 Mean: 35.75 Standard deviation: 9.59
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

FAIL: Seroconversion occurred at the end of this episode

Data file: incidence-kisesa

Overview

Valid: 43638 Invalid: 160
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0	No	43138	98.9%
1	Yes	500	1.1%
Sysmiss		160	

IDNO: Participant ID number for sharing

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0
 Type: Discrete Decimal: 0 Width: 5 Format: Numeric Weighted variable: 0

T0: Stata variable: age at start of episode

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0 Minimum: 15 Maximum: 49.988 Mean: 28.58 Standard deviation: 9.866
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

T: Stata variable: age at end of episode

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0 Minimum: 15.001 Maximum: 50 Mean: 30.249 Standard deviation: 10.036
 Type: Continuous Decimal: 0 Width: 16 Format: Numeric Weighted variable: 0

ST: Stata variable: in survival analysis

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0
 Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
-------	----------	-------	--

1		43798	100%
---	--	-------	------

D: Stata variable: failure (seroconversion)

Data file: incidence-kisesa

Overview

Valid: 43798 Invalid: 0

Type: Discrete Decimal: 0 Width: 1 Format: Numeric Weighted variable: 0

Questions and instructions

CATEGORIES

Value	Category	Cases	
0		43298	98.9%
1		500	1.1%

study_resources

technical_documents

Contextual documentation for the shared ALPHA data Kisesa

title	Contextual documentation for the shared ALPHA data Kisesa
date	2022-04-25
country	Tanzania
contributors	This document provided background information on the dataset
filename	alpha-kisesa-1994-2016.pdf
