

South Africa

Kobus Herbst, Africa Health Research Institute
Willem Hanekom, Africa Health Research Institute
Mark Siedner, Africa Health Research Institute
Guy Harling, Africa Health Research Institute

Network for Analysing Longitudinal Population- based HIV/AIDS data on Africa (ALPHA Network)

Study Documentation

August 13, 2026

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	AHRI.AN03-03.Dataset.7.4

Table of Contents

Overview.....	4
Scope & Coverage.....	4
Producers & Sponsors.....	4
Sampling.....	5
Data Collection.....	5
Accessibility.....	5
Files Description.....	6
AHRI.AN03-03.Dataset.7.4.....	6
Variables List.....	7
AHRI.AN03-03.Dataset.7.4.....	7
Variables Description.....	8
AHRI.AN03-03.Dataset.7.4.....	9

Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network)

Overview	
Identification	AHRI.AN03-03.Dataset.7.4
Version	v1.1.0
Abstract The ALPHA Network is a collaboration between 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. AHRI contributes data from ACDIS and PIP to the Network and this is referred to within the Network as uMkhanyakude. This dataset contains the sex and date of birth for the population resident in the study area and dates of migrations and deaths. It is used to construct the population denominator for all analyses.	
Kind of Data	Surveillance data
Unit of Analysis	Each observation describes an individual's period of residence in a household

Scope & Coverage	
Keywords	ALPHA; ALPHA Network; HIV incidence; HIV mortality; ANDLA Network; residency
Topics	Demography Epidemiology HIV Mortality Incidence
Time Period(s)	2002-2024
Countries	South Africa
Geographic Coverage Africa Health Research Institute Population Intervention Study Area (PIPSA)	
Universe Household members of the Africa Health Research Institute Population Intervention Programme Study Area	

Producers & Sponsors	
Primary Investigator(s)	Kobus Herbst, Africa Health Research Institute Willem Hanekom, Africa Health Research Institute Mark Siedner, Africa Health Research Institute Guy Harling, Africa Health Research Institute
Other Producer(s)	Africa Health Research Institute (AHRI)
Funding Agency/ies	Wellcome Trust (WT) , Core funding SAPRIN (SAPRIN)
Other Acknowledgment(s)	Thobeka Mngomezulu , Data collection , AHRI Phumzile Dlamini , Data collection , AHRI Nomathamsanqa Majozi , Public Engagement , AHRI Siyabonga Nxumalo , Data extraction , AHRI Sweetness Dube , Data Documentation , AHRI Dickman Gareta , Data Management oversight , AHRI Eugene Ehlers , Software Development , AHRI Theresa Smit , Laboratory oversight , AHRI Gregory Ording-Jesperse , Laboratory data management , AHRI

Sampling

Sampling Procedure

Household demographic surveillance

Data Collection

Data Collection Dates	start 2002-11-25 end 2024-12-17
------------------------------	------------------------------------

Accessibility

Access Conditions

Access to the data requires accurate completion of the online data access application form accessible on the AHRI Data repository(<<https://data.ahri.org/>>). Data users are required to abide by the data use conditions stipulated on the application for access to the data. Failure to do so may result in their data access privileges revoked by the Data Custodian. In order to recognise the effort and intellectual contributions of AHRI investigators in producing and curating the data, users of AHRI data must acknowledge the source of the data and abide by the terms and conditions under which the data is accessed. All analytical datasets published on the AHRI Data Repository are assigned digital object identifier (DOIs) and the DOIs can be found on the Data Repository under Study Description tab - Access policy. AHRI data users are required to always cite the dataset using the DOI.

Citation Requirements

Herbst, K., Hanekom, W., Siedner, M., & Harling, G. (2024). Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network) [Data set]. Africa Health Research Institute.
DOI:<https://doi.org/10.23664/AHRI.AN03-03.DATASET.7.4>

Files Description

Dataset contains 1 file(s)

AHRI.AN03-03.Dataset.7.4	
# Cases	6317262
# Variable(s)	20

Variables List

Dataset contains 20 variable(s)

File AHRI.AN03-03.Dataset.7.4							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	idno	Person ID number	discrete	character-36	6317262	0	-
2	study_name	Name of your study field site	discrete	character-5	6317262	0	-
3	date_int	Date of interview	discrete	character-11	6317256	-	-
4	survey_n_..	Which survey interview was part of	continuous	numeric-9.0	6317262	0	-
5	pregnant	Currently pregnant	discrete	numeric-18.0	6317262	0	-
6	proxy_resp	Proxy respondent or individual interview	discrete	numeric-22.0	6317262	0	-
7	marital_..	Marital status	discrete	numeric-26.0	6317262	0	-
8	educatio_..	Highest education level completed	discrete	numeric-18.0	6317262	0	-
9	educatio_..	Highest education level completed	discrete	numeric-20.0	6317262	0	-
10	current_..	Currently attending school	discrete	numeric-18.0	6317262	0	-
11	mother_A_..	Mother alive	discrete	numeric-18.0	6317262	0	-
12	mother_l_..	Mother location	discrete	numeric-26.0	6317262	0	-
13	father_a_..	Father alive	discrete	numeric-18.0	6317262	0	-
14	father_l_..	Father location	discrete	numeric-26.0	6317262	0	-
15	age_firs_..	Age at first birth	discrete	numeric-9.0	6317262	0	-
16	age_firs_..	Age at first marriage	continuous	numeric-9.0	386092	5931170	-
17	ever_sex	Ever had sex	discrete	numeric-18.0	386092	5931170	-
18	age_firs_..	Age at sexual debut	continuous	numeric-9.0	386092	5931170	-
19	injectio_..	How many injections in last year	discrete	numeric-18.0	6317262	0	-
20	blood_tr_..	Ever had a blood transfusion	discrete	numeric-18.0	6317262	0	-

Variables Description

Dataset contains 20 variable(s)

File : AHRI.AN03-03.Dataset.7.4

idno: Person ID number

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

study_name: Name of your study field site

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

date_int: Date of interview

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=6317256 /-]

survey_name: Which survey interview was part of

Information	[Type= continuous] [Format=numeric] [Range= 1-182] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-] [Mean=21.106 /-] [StdDev=17.137 /-]

pregnant: Currently pregnant

Information	[Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	no	1812715	28.7%
1	yes	11779	0.2%
88	Don't know	19859	0.3%
99	Question not asked	4472909	70.8%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

proxy_resp: Proxy respondent or individual interview

Information	[Type= discrete] [Format=numeric] [Range= 1-2] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Individual interviewed	0	
2	Proxy respondent	6317262	100.0%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

marital_status: Marital status

Information	[Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Never married/cohabited	3119839	49.4%
1	Currently married mono	0	
2	Currently married Poly	51777	0.8%
3	Currently married	483217	7.6%
4	Currently cohabiting	0	
5	Widowed	234833	3.7%
6	Separated/divorced	14383	0.2%
7	Married but living apart	0	
8	Widowed/divorced/separated	0	

File : AHRI.AN03-03.Dataset.7.4

marital_status: Marital status

Value	Label	Cases	Percentage
88	Donâ€™t know	0	
99	Question not asked	2413213	38.2%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

education_level: Highest education level completed

Information	[Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Grade 0	3200	0.1%
1	Grade 1	61120	1.0%
2	Grade 2	68225	1.1%
3	Grade 3	73658	1.2%
4	Grade 4	82594	1.3%
5	Grade 5	70657	1.1%
6	Grade 6	75494	1.2%
7	Grade 7	97887	1.5%
8	Grade 8	94613	1.5%
9	Grade 9	104754	1.7%
10	Grade 10	149190	2.4%
11	Grade 11	174997	2.8%
12	Grade 12	397329	6.3%
13	Special School	17	0.0%
21	Higher Certificate	40985	0.6%
22	National Diploma	19678	0.3%
23	Degree	32092	0.5%
24	Honours	1113	0.0%
25	Masters	1770	0.0%
26	Doctors	134	0.0%
31	ABET 1	1334	0.0%
32	ABET 2	171	0.0%
33	ABET 3	109	0.0%
34	ABET 4	324	0.0%
41	FET 1	3025	0.0%
42	FET 2	3662	0.1%
43	FET 3	7824	0.1%
45	NQF/N/NTC	3185	0.1%
50	LT 1 year	29019	0.5%
51	Creche	255	0.0%
88	Donâ€™t know	4610839	73.0%
90	No education	102916	1.6%
99	Question not asked	5092	0.1%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

File : AHRI.AN03-03.Dataset.7.4

education_level_standard: Highest education level completed

Information [Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	None	132190	2.1%
1	Incomplete Primary	434948	6.9%
2	Complete Primary	97887	1.5%
3	Incomplete Secondary	523554	8.3%
4	Complete Secondary	397329	6.3%
5	Higher education	115406	1.8%
88	Donâ€™t know	4610856	73.0%
99	Question not asked	5092	0.1%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

current_school: Currently attending school

Information [Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	no	934857	14.8%
1	yes	642809	10.2%
88	Donâ€™t know	12071	0.2%
99	Question not asked	4727525	74.8%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

mother_Alive: Mother alive

Information [Type= discrete] [Format=numeric] [Range= 1-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	42148	0.7%
2	Dead	852421	13.5%
88	Donâ€™t know	5422693	85.8%
99	Question not asked	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

mother_location: Mother location

Information [Type= discrete] [Format=numeric] [Range= 1-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	co-resident	0	
2	elsewhere In DSS area	16787	0.3%
3	elsewhere outside DSS area	25361	0.4%
88	Donâ€™t know	5422693	85.8%
99	Question not asked	852421	13.5%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

father_alive: Father alive

Information [Type= discrete] [Format=numeric] [Range= 1-99] [Missing=*]

File : AHRI.AN03-03.Dataset.7.4

father_alive: Father alive

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	108446	1.7%
2	Dead	1571747	24.9%
88	Donâ€™t know	4637069	73.4%
99	Question not asked	0	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

father_location: Father location

Information [Type= discrete] [Format=numeric] [Range= 1-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	co-resident	0	
2	elsewhere In DSS area	42580	0.7%
3	elsewhere outside DSS area	65866	1.0%
88	Donâ€™t know	4637069	73.4%
99	Question not asked	1571747	24.9%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

age_first_birth: Age at first birth

Information [Type= discrete] [Format=numeric] [Range= 999-999] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

age_first_marriage: Age at first marriage

Information [Type= continuous] [Format=numeric] [Range= 10-999] [Missing=*]

Statistics [NW/ W] [Valid=386092 /-] [Invalid=5931170 /-] [Mean=956.362 /-] [StdDev=198.47 /-]

ever_sex: Ever had sex

Information [Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]

Statistics [NW/ W] [Valid=386092 /-] [Invalid=5931170 /-]

Value	Label	Cases	Percentage
0	no	38153	9.9%
1	yes	102196	26.5%
88	Donâ€™t know	0	
99	Question not asked	245743	63.6%
Sysmiss		5931170	

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

age_first_sex: Age at sexual debut

Information [Type= continuous] [Format=numeric] [Range= 0-999] [Missing=*]

Statistics [NW/ W] [Valid=386092 /-] [Invalid=5931170 /-] [Mean=818.564 /-] [StdDev=378.556 /-]

injections_last_year: How many injections in last year

Information [Type= discrete] [Format=numeric] [Range= 99-99] [Missing=*]

Statistics [NW/ W] [Valid=6317262 /-] [Invalid=0 /-]

File : AHRI.AN03-03.Dataset.7.4

injections_last_year: How many injections in last year

Value	Label	Cases	Percentage
99	Question not asked	6317262	100.0%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

blood_trans: Ever had a blood transfusion

Information	[Type= discrete] [Format=numeric] [Range= 99-99] [Missing=*]
Statistics [NW/ W]	[Valid=6317262 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
99	Question not asked	6317262	100.0%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.