

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

July 31, 2024

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.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
AN03-03 Dataset 7.4.....	6
Variables List.....	7
AN03-03 Dataset 7.4.....	7
Variables Description.....	8
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
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 2024-06-04
------------------------------	------------------------------------

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 being 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 and must cite the dataset in publication using the citation provided as part of this documentation. 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 relevant 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)

AN03-03 Dataset 7.4	
# Cases	7633032
# Variable(s)	20

Variables List

Dataset contains 20 variable(s)

File AN03-03 Dataset 7.4							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	idno	Person ID number	continuous	numeric-12.0	7633032	0	-
2	study_name	Name of study field site	discrete	character-5	7633032	0	-
3	date_int	Date of interview	discrete	character-11	7633032	-	-
4	survey_n_..	Surveillance round number	continuous	numeric-12.0	7633032	0	-
5	proxy_resp	Proxy respondent or individual interviewed	discrete	numeric-12.0	7633032	0	-
6	pregnant	Currently pregnant	discrete	numeric-12.0	7633032	0	-
7	marital_..	Marital status	discrete	numeric-12.0	7633032	0	-
8	educatio_..	Site specific education level	discrete	numeric-12.0	2189763	5443269	-
9	educatio_..	common education variable	discrete	numeric-12.0	2189763	5443269	-
10	current_..	Currently attending school	discrete	numeric-12.0	2189763	5443269	-
11	mother_a_..	Mother is alive	discrete	numeric-12.0	7633032	0	-
12	mother_l_..	Mother location	discrete	numeric-12.0	7633032	0	-
13	father_a_..	Father is alive	discrete	numeric-12.0	7633032	0	-
14	father_l_..	Father location	discrete	numeric-12.0	7633032	0	-
15	age_firs_..	Age at which first birth occurred	continuous	numeric-12.0	2478959	5154073	-
16	age_firs_..	Age at first marriage	continuous	numeric-12.0	6833374	799658	-
17	ever_sex	Ever had sex	discrete	numeric-12.0	7633032	0	-
18	age_firs_..	Age at sexual debut	continuous	numeric-12.0	633514	6999518	-
19	injectio_..	How many injections in last year	discrete	numeric-12.0	7633032	0	-
20	blood_tr_..	Ever had a blood transfusion	discrete	numeric-12.0	7633032	0	-

Variables Description

Dataset contains 20 variable(s)

File : AN03-03 Dataset 7.4

idno: Person ID number

Information	[Type= continuous] [Format=numeric] [Range= 11-275969] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-] [Mean=96317.241 /-] [StdDev=67576.412 /-]

study_name: Name of study field site

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-]

date_int: Date of interview


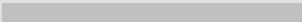
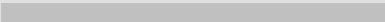
Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-]

survey_name: Surveillance round number

Information	[Type= continuous] [Format=numeric] [Range= 0-2016] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-] [Mean=528.156 /-] [StdDev=859.599 /-]

proxy_resp: Proxy respondent or individual interviewed

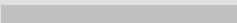


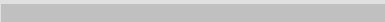
Information	[Type= discrete] [Format=numeric] [Range= 1-9] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Individual interviewed	648553	 8.5%
2	Proxy respondent	3068160	 40.2%
9	Unknown	3916319	 51.3%

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.

pregnant: Currently pregnant

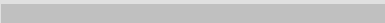







Information	[Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	No	2873637	 37.6%
1	Yes	25300	 0.3%
88	Don't know	120532	 1.6%
99	Question not asked	4613563	 60.4%

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=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	Never married/cohabited	3971132	 52.0%
1	Currently married mono	550581	 7.2%
2	Currently married Poly	59768	 0.8%
3	Currently married	51845	 0.7%
5	Widowed	249518	 3.3%
6	Separated/divorced	15609	 0.2%
9	Currently not married (not further specified)	12284	 0.2%
88	Don't know	21189	 0.3%

File : AN03-03 Dataset 7.4

marital_status: Marital status

Value	Label	Cases	Percentage
93	Refused	299	0.0%
94	Query	0	
95	Default	0	
96	Other	196	0.0%
99	Missing	2700611	35.4%

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: Site specific education level

Information	[Type= discrete] [Format=numeric] [Range= 0-255] [Missing=*]
Statistics [NW/ W]	[Valid=2189763 /-] [Invalid=5443269 /-]

Value	Label	Cases	Percentage
0	Never went to school	138819	6.3%
1	Less than one year	29178	1.3%
2	Grade 1 or Sub A/Class1	61801	2.8%
3	Grade 2 or Sub B/Class2	69885	3.2%
4	Grade 3 / Standard 1	76985	3.5%
5	Grade 4 / Standard 2	88462	4.0%
6	Grade 5 / Standard 3	76744	3.5%
7	Grade 6 / Standard 4	84722	3.9%
8	Grade 7 / Standard 5	115787	5.3%
9	Grade 8 / Standard 6	113884	5.2%
10	Grade 9 / Standard 7	130053	5.9%
11	Grade 10 / Standard 8	191214	8.7%
12	Grade 11 / Standard 9	231561	10.6%
13	Grade 12 / Matric	509421	23.3%
14	Certificate (NTC)	50833	2.3%
15	Diploma	20062	0.9%
16	Bachelor's Degree (University)	33822	1.5%
17	Bachelor's Degree with Diploma	2448	0.1%
18	Honours degree	1069	0.0%
20	Hons/Masters Degree or higher	1349	0.1%
93	_refused	2517	0.1%
94	_query	0	
95	_default	725	0.0%
96	_other	27	0.0%
97	_non applicable	0	
98	_don't know	96295	4.4%
99	_missing	62100	2.8%
255	_deprecated	0	
Sysmiss		5443269	

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_standard: common education variable

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

File : AN03-03 Dataset 7.4

education_level_standard: common education variable

Statistics [NW/ W] [Valid=2189763 /-] [Invalid=5443269 /-]

Value	Label	Cases	Percentage
0	None	167997	7.7%
1	Incomplete Primary	458599	20.9%
2	Complete Primary	359724	16.4%
3	Incomplete Secondary	422775	19.3%
4	Complete Secondary	509421	23.3%
5	Higher education	109583	5.0%
88	Don't know	96295	4.4%
99	Question not asked	65369	3.0%
Sysmiss		5443269	

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=2189763 /-] [Invalid=5443269 /-]

Value	Label	Cases	Percentage
0	no	1186658	54.2%
1	yes	640668	29.3%
88	Don't know	12973	0.6%
99	Not asked	349464	16.0%
Sysmiss		5443269	

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 is alive

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

Statistics [NW/ W] [Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	5015844	65.7%
2	Dead	1621712	21.2%
88	Don't know	101356	1.3%
99	Not asked	894120	11.7%

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= 99-99] [Missing=*]

Statistics [NW/ W] [Valid=7633032 /-] [Invalid=0 /-]

father_alive: Father is alive

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

Statistics [NW/ W] [Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	3551441	46.5%
2	Dead	2584541	33.9%
88	Don't know	279006	3.7%

File : AN03-03 Dataset 7.4

father_alive: Father is alive

Value	Label	Cases	Percentage
99	Not asked	1218044	16.0%

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=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	0	
2	Dead	0	
88	Don't know	0	
99	Not asked	7633032	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.

age_first_birth: Age at which first birth occurred

Information	[Type= continuous] [Format=numeric] [Range= -5-223] [Missing=*]
Statistics [NW/ W]	[Valid=2478959 /-] [Invalid=5154073 /-] [Mean=25.34 /-] [StdDev=6.807 /-]

age_first_marriage: Age at first marriage

Information	[Type= continuous] [Format=numeric] [Range= 10-99999999] [Missing=*]
Statistics [NW/ W]	[Valid=6833374 /-] [Invalid=799658 /-] [Mean=1046.302 /-] [StdDev=76508.234 /-]

ever_sex: Ever had sex

Information	[Type= discrete] [Format=numeric] [Range= 0-99] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
0	no	64424	0.8%
1	yes	212136	2.8%
88	Don't know	58314	0.8%
99	Question not asked	7298158	95.6%

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=633514 /-] [Invalid=6999518 /-] [Mean=778.721 /-] [StdDev=407.65 /-]

injections_last_year: How many injections in last year

Information	[Type= discrete] [Format=numeric] [Range= 1-99] [Missing=*]
Statistics [NW/ W]	[Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	0	
2	Dead	0	
88	Don't know	0	
99	Not asked	7633032	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.

File : AN03-03 Dataset 7.4

blood_trans: Ever had a blood transfusion

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

Statistics [NW/ W] [Valid=7633032 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage
1	Alive	0	
2	Dead	0	
88	Don't know	0	
99	Not asked	7633032	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.