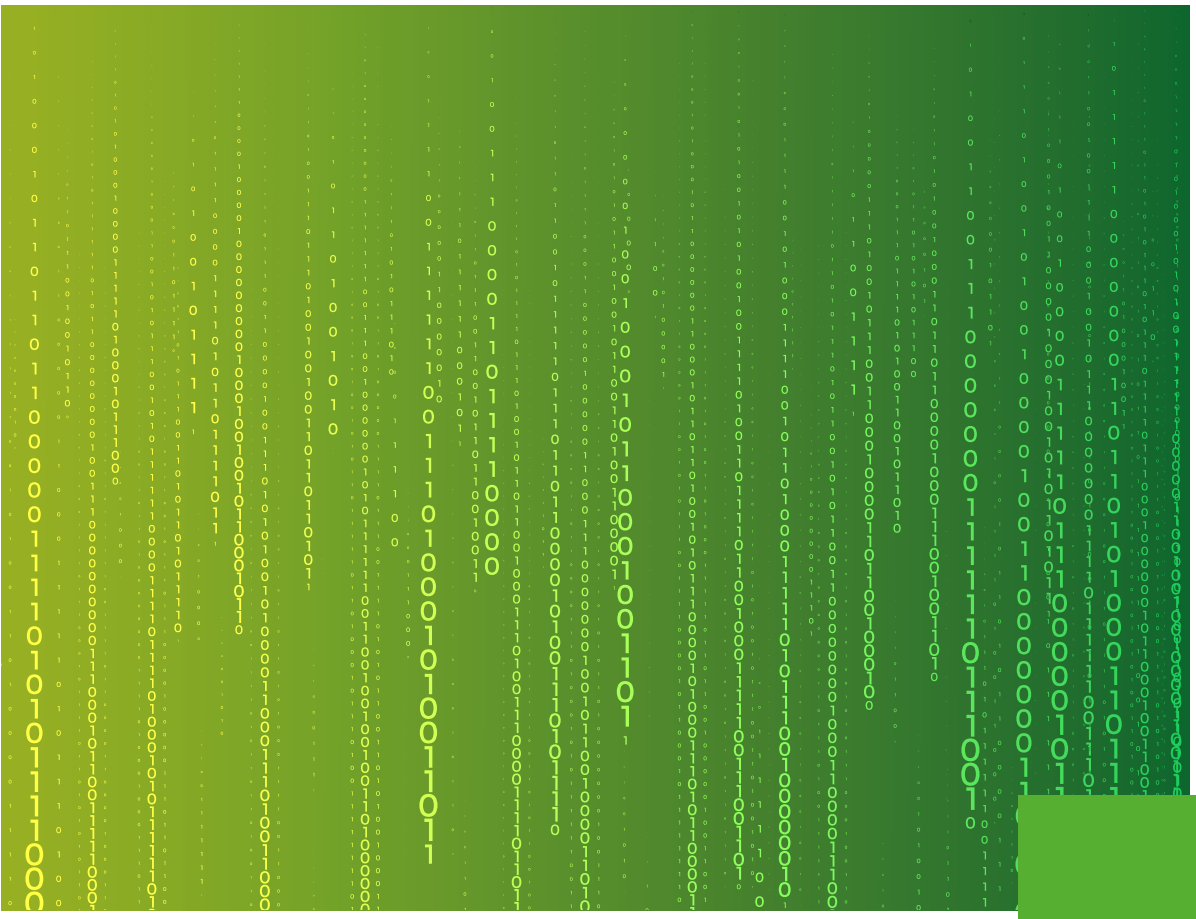
INSIDA

Mozambique Population-based HIV Impact Assessment



DATA USE MANUAL SUPPLEMENT

INSTITUTO NACIONAL DE SAÚDE MOÇAMBIQUE

MINISTÉRIO DA SAÚDE

Conselho Nacional de Combate ao HIV/SIDA CNCS

INSTITUTO NACIONAL DE ESTATÍSTICA

PEPFAR U.S. President's Emergency Plan for AIDS Relief

CDC Centers for Disease Control and Prevention

Westat

icap Global Health

Mozambique Population-based HIV Impact Assessment 2021

INSIDA 2021

This project is supported by the US President's Emergency Plan for AIDS Relief (PEPFAR) through CDC under the terms of cooperative agreement U2GGH002173. The findings and conclusions are those of the authors and do not necessarily represent the official position of the funding agencies.

INSIDA 2021 Collaborating Institutions

Instituto Nacional de Saúde (INS), Mozambique
Ministry of Health (MoH), Mozambique
Centers for Disease Control and Prevention (CDC), Mozambique
Centers for Disease Control and Prevention (CDC), Atlanta
ICAP at Columbia University, New York
ICAP at Columbia University, Mozambique
Westat, Rockville
Instituto Nacional de Estatística (INE), Mozambique
Conselho Nacional de Combate ao SIDA (CNCS), Mozambique

Donor Support

This project is supported by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of cooperative agreement U2GGH002173.The findings and conclusions are those of the authors and do not necessarily represent the official position of the funding agencies.

Suggested Citation

Manual

Mozambique Population-based HIV Impact Assessment (Inquérito Nacional sobre o impacto do HIV e SIDA em Moçambique) (INSIDA) 2021 Data Use Manual Supplement. New York, NY. January 2024.

Single dataset

Ministry of Health (MOH), Mozambique Population-based HIV Impact Assessment (Inquérito Nacional sobre o impacto do HIV e SIDA em Moçambique) (INSIDA) 2021 [Dataset]. <Insert dataset name used, e.g. insida2021adultbio>. Maputo: MOH, Instituto Nacional de Saúde, Instituto Nacional de Estatística, Conselho Nacional de Combate ao SIDA, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], January 2024.

Multiple datasets (more than 3)

Ministry of Health (MOH), Mozambique Population-based HIV Impact Assessment (Inquérito Nacional sobre o impacto do HIV e SIDA em Moçambique) (INSIDA) 2021 [Dataset]. <Insert dataset name used, e.g. insida2021adultbio>. Maputo: MOH, Instituto Nacional de Saúde, Instituto Nacional de Estatística, Conselho Nacional de Combate ao SIDA, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], January 2024.

Access this Manual Online

https://phia-data.icap.columbia.edu/datasets?country_id=15

Contact Information

ICAP at Columbia University
722 West 168th Street
New York, NY 10032
Website: icap.columbia.edu
Email: icap-communications@columbia.edu

Table of Contents

Abbreviations

ART	Antiretroviral Therapy
ARV	Antiretroviral
CA	Control Area
CAP/CTM	COBAS AmpliPrep/COBAS Taqman HIV-1 Qualitative Test
CD4	CD4+ T-Cell
CI	Confidence Interval
CONSORT	Consolidated Standard of Reporting Trials
DHS	Demographic and Health Surveys
DNA	Deoxyribonucleic Acid
EA	Enumeration Area
HIV	Human Immunodeficiency Virus
ID	Identification
LAg-EIA	Limiting-Antigen Avidity Enzyme Immunoassay
INSDIA	Inquérito Nacional sobre o impacto do HIV e SIDA em Moçambique (Mozambique Population-based HIV Impact Assessment)
OVC	Orphans and Vulnerable Children
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PHIA	Population-based HIV Impact Assessment
PSU	Primary Sampling Unit

1 Background

1.1 What is INSIDA 2021?

The Mozambique Population-based HIV Impact Assessment 2021 (INSIDA 2021) was a cross-sectional household-based survey conducted in Mozambique. INSIDA 2021 is part of the PHIA Project, a series of population-based surveys, which are designed to assess the burden of HIV disease and impact of the health sector response to national HIV epidemics.

1.2 Purpose of the INSIDA 2021 Data Use Manual Supplement

The ***INSIDA 2021 Data Manual Supplement*** (hereafter, ***Supplement***) contains INSIDA 2021 survey specifications, including survey-specific eligibility criteria, sampling approaches and measures, and survey-specific technical documentation such as codebooks and questionnaires. It is intended to accompany the ***PHIA Data Manual*** (hereafter, "***Manual***"), which contains information on PHIA data generally applicable to all PHIA surveys, including general information on the data and documentation packages and their contents, a guide to getting started with the PHIA data, and details on the files and variables available for all PHIAs. Users should refer to both the ***Manual*** and this ***Supplement*** for a complete discussion of the INSIDA 2021 data.

1.3 Other documentation and resources

In addition to the ***Manual*** and this ***Supplement***, users should refer to INSIDA 2021 publications such as the ***INSIDA 2021 Summary Sheet*** and ***INSIDA 2021 Final Report***. The ***INSIDA 2021 Summary Sheet*** contains highlights and summary results from the survey. The ***INSIDA 2021 Final Report*** contains detailed results from INSIDA 2021 along with information on survey data collection procedures, establishing participation by the household head, procedures for individual consent, maintaining confidentiality during data collection and testing procedures, procedures for returning/obtaining test results, and referral for or direct linkage to services are included.

Important, INSIDA 2021-specific documentation attached to ***Supplement***, includes:

- ***INSIDA 2021 Survey Questionnaires***: The INSIDA 2021 household, roster, and adult questionnaires. These questionnaires illustrate the questionnaire's structure, including the order that the questions were asked, each question's wording, variable names and labels, value coding and labels, and skip patterns. The question number on the questionnaire is referenced in the variable label on the datasets and in the "variable label" of the codebook, where applicable.
- ***INSIDA 2021 Codebook with Frequencies***: Codebooks are provided for each dataset and document each variable's name, category (i.e., the questionnaire module or source data of the variable), full question text or variable description, variable label (i.e., a condensed label used on the datasets), type and width (e.g., numeric, text), coding values and labels, and the frequency and percent of records containing each value. Summary statistics are provided in the coding values and labels for selected numeric variables, such as counts.
- ***INSIDA 2021 Analytic Variable Flow Diagrams***: Flow diagrams illustrate the logic used to create key analytic variables.

- ***INSIDA 2021 Testing Methodology Diagram***: Flow diagram illustrating household-based HIV testing algorithm.
- ***INSIDA 2021 Sampling and Weighting Technical Report***: Details of INSIDA 2021 sampling and weighting procedures.
- ***INSIDA 2021 Survey-Specific Table Specifications****:* Table shells and technical specifications for INSIDA 2021-specific final report tabulations. (Shells and specifications for all standard final report tabulations are included in the ***Manual***.)

Utility statistical programs are provided for Stata, SAS, or R depending on the file format requested.

  - ***INSIDA 2021 Stata Intro Code.do***: Stata do-file
  - ***INSIDA 2021 SAS Intro Code.sas***: SAS program
  - ***INSIDA 2021 R Intro Code.R***: R script

Datasets and labels are incorporated into the Stata datasets. For SAS, there is a second statistical program containing code to label all values for variables on each of the data sets.

  - ***INSIDA 2021 SAS Formats.sas***: SAS program

2 Survey design and data collection

INSIDA 2021 was a nationally representative, cross-sectional, multi-stage, population-based survey of households across Mozambique. Its target population corresponded to adults, defined in this survey as those aged 15 years and older.

Table 1. INSIDA 2021 survey design characteristics

Survey design characteristics	Description
Survey design	
Data source for survey weighting[1]	2021 Mozambique National Population Projections (INE)
Sampling stratum	Province
Primary sampling unit	INE Control Areas (CA)
Secondary sampling unit	Census enumeration area (EA)
Tertiary sampling unit	Dwelling units
Quaternary sampling unit	Household units
Urban/rural categorization	Urban/rural
Survey administration	
Data collection dates	April 2021 – July 2021 (wave 1) September 2021 – February 2022 (wave 2)
Languages	Portuguese, Emakhuwa, Xichangana, Cisena, Elomwe, Echuwabo, Cinyanja, Cindau, Xitswa, Cinyungwe, Ciyao, Shona
Sample size[1]	
Number of selected EAs	311[2]
Number of selected dwelling units	11,375
Number of selected households	9,015
Number of rostered individuals	38,825
Survey participation	
Number of completed household interviews[3]	8,690
Number of completed individual interviews	17,105
Number of completed biomarker tests	14,488

[1] See **Attachment 9.5: *Sampling and Weighting Technical Report*** for more details on the sampling and weighting approach for the survey and for response rates.

[2] The original sampling process selected 324 CA/EAs. Twelve of these were later excluded due to security concerns and/or community-level refusal to participate. One EA had no inhabitants.

[3] One household completed the household interview, but had no individual interviews carried out because the occupants did not have time to continue the survey.

Exceptions to the general PHIA design

Sample Design

The EA selection process in INSIDA 2021 was slightly more complex than the general PHIA design, consisting of 4 stages: control areas, enumeration areas, dwelling units and households. INE selected the sample of control areas (CAs) from a sampling frame developed from the 2017 population and housing census. CAs typically comprised 3-4 EAs along with a measure of size

based on the household count from the 2017 census. Within each CA, an equal probability random sample of one EA was taken to obtain the final list of selected EAs.

INSIDA 2021 also had an additional dwelling unit sampling step compared with the typical PHIA design. In the EAs selected for the survey, all potentially eligible dwelling units and households were listed and used to create a dwelling unit sampling frame (a dwelling unit is defined as a collection of eligible households sharing the same housing structure). The dwelling units were sampled with probability proportional to the number of eligible households they contained. The households within each selected and responding dwelling unit were re-listed, and one household was randomly selected from each dwelling unit to participate in the survey.

For a full explanation of the sample design and sample selection procedures, see the Sampling and Weighting Technical Report.

Dataset Changes

PSU-level weights are included in an additional intermediary psu weights file, with the PSUs being uniquely identified by the centroidid variable which can be used to link to the other INSDIA 2021 datasets.

Dwelling unit weights are included in the household file with the prefix ‘duwt’, and supplementary dwelling unit weights are included in the household supplementary weights dataset.

In the INSIDA 2021 household dataset, there is one row for each dwelling unit, instead of one row for each household unit as in standard PHIA surveys. The dwellingid variable is included in the household dataset to uniquely identify the dwellings, and a du_status variable records the survey outcome at the dwelling level. One household was selected per dwelling, so each row of this dataset also corresponds to one household, except for non-responding dwellings where no household was selected. Subsetting using the household status variable is the simplest way to ensure that only the households of interest are selected for analysis.

See the Intermediary Weights Codebook and the Sampling and Weighting Technical Report for full details of the weight variables included in the INSIDA 2021 package.

Questionnaire Changes

There were several country-specific changes to the questionnaire in INSIDA 2021 when compared to standard PHIA surveys. Questions with differences that could cause misinterpretation or incomparability with the corresponding questions in other PHIA countries have had their dataset variables renamed to use a “_mz” suffix. For full details regarding the questions, variables, and response options, refer to the questionnaire (Data Manual Supplement Attachment 1 – Questionnaires). The full list of questions having country-specific changes is as follows:

Table 2. INSIDA 2021 questions and variables with country-specific changes

Household Dataset		
Questions	**Variable Names**	**Differences from PHIA standard**

What is the main source of drinking water for members of your household?	watersource_mz	Additional response options
What kind of toilet facility do members of your household usually use?	toilettype_mz	Additional response options
What type of fuel does your household mainly use for lighting?	lightingfuel_mz	New question not in standard template
Whom does this dwelling belong to?	dwellbelong_mz	New question not in standard template
Altogether, how many pigs do members of your household own?	ownpignum_mz	New question not in standard template
Has your household received any financial assistance, material goods, or food support related to COVID-19?	econsupcovid_mz	New question not in standard template
Roster Dataset		
Questions	**Variable Names**	**Differences from PHIA standard**
Province or country individual is currently in	liveregionlivecountry_mz	Survey-specific response options
Adult Interview Dataset		
Questions	**Variable Names**	**Differences from PHIA standard**
Interview language and native language	lngvqx_lng_mz, lngvint_lng_mz, lngnat_lng_mz	Survey-specific response options
What is the highest level of school you attended?	schcom_mz	Survey-specific response options
What is the highest class/year you completed at that level?	schcomc1_mz	New question not in standard template
Just before you moved here, did you live in an urban area or in a rural area?	outregiontype_mz	Survey-specific response options
Before you moved here, which province did you live in? If you lived outside of Mozambique, which country did you live in?	outregionwhr_mz	Survey-specific response options
The last time you were away from home for more than a month, where were you?	whereout_mz	INSIDA-specific response options
What was the main reason you went there?	reasonaway_mz	New COVID option added
What is your occupation? That is, what kind of work do you mainly do?	workind_mz	Additional response options
What was the main reason you did not take ARVs while you were pregnant with your last child?	arvnrpg_mz	New options added, including COVID

When you were pregnant with your last child, were you offered testing for syphilis?	prgsyphoffer[1-5]_mz	New question not in standard template
When you were pregnant with your last child, were you tested for syphilis?	prgsyphtest[1-5]_mz	New question not in standard template
Did you test positive for syphilis during your last pregnancy?	prgsyphpos[1-3]_mz	New question not in standard template
Did you get treatment for syphilis during your last pregnancy?	prgsyphtreat[1-2]_mz	New question not in standard template
Who did your circumcision?	mcwho_mz	New question not in standard template
Could you point to the image that best describes the type of circumcision you went through?	mcilltype_mz	New question not in standard template
Where were you circumcised?	mcloc_mz	New question not in standard template
Why haven't you used a condom during the last sexual intercourse?	partlastcndmrsn_mz	New question not in standard template
Why have you never been tested for HIV?	hivtstnvrrsn_[a-l, x]_mz	COVID-related option added
Where was your last HIV test done?	hivtstlocation_mz	Additional response option
At which facility are you currently receiving HIV care? (province of facility)	hivclinic_gv1_mz	New question not in standard template
What is the main reason why you have never received care or treatment for HIV from a doctor, clinical officer, or nurse?	hivcnotrsn_mz	COVID-related option added
What is the main reason you have never taken ARVs?	arvsnottake_mz	Additional response options added, including COVID
Can you tell me the main reason why you stopped taking ARVs?	arvsnotcurrsn_mz	COVID-related option added
How do you normally receive your ARVs?	arvloc_mz	Additional response options
Was there any period since March 2020, when the state of emergency caused by COVID was announced, when you obtained (or were told to obtain) your ART in a different way or place than where you usually receive them?	arvloccovid_mz	New question not in standard template
Why were your ARVs changed?	arvswitchwhy_mz	Added in a response option
Have you ever been told by a doctor or health worker that you have any of the following chronic health conditions?	chroniccond_[a-i, x]_mz	Additional response option – H - Epilepsy
Are you currently taking medication for any of the following chronic health conditions?	chronicmed_[a-i, x]_mz	Additional response option – H - Epilepsy
Have you ever heard of Hepatitis B or C?	hepatitis_mz	New question not in standard template

Have you ever received a blood transfusion?	bloodtrans_mz	New question not in standard template
Where do you normally have alcohol?	alcloc_mz	New question not in standard template
In the past 12 months, have you used non-prescription drugs for recreation or pleasure?	drug_mz	New question not in standard template
Which non-prescription drugs have you ever consumed, even one time?	drugtype_[a-e, x]_mz	New question not in standard template
Why is it not easy for you to get a condom?	condomnoteasyrsn_[a-f,x]_mz	COVID-related option added
Have you ever discussed HIV with your parents or guardian?	addishiv_mz	New question not in standard template
2 year groups for <25yrs, age groups	agegroup2years_mz	Survey-specific analytic variable for final report tabulations
How do you normally receive your ARVs? (final report table version with “Not on ARVs” option)	arvlocalt_mz	Additional response options.
Has given birth in 2 years preceding survey start date	delivered2years_mz	Survey-specific analytic variable for final report tabulations
First sex before age 18	firstsexbefore18_mz	Survey-specific analytic variable for final report tabulations

3 Overview of survey questionnaires

In participating households, a household questionnaire was administered to the designated household head. Household head is defined as an individual age 18 or older and emancipated minors (defined in Mozambique as an individual ages 15-17 who is or has ever been married or pregnant, lived alone, or headed a household). The household head provided consent for the household to participate in the survey, after which individual members were rostered during the household interview.

Then, adult individual questionnaires were administered to eligible and consenting individuals aged 15 and older in the household. Specific consent criteria are determined in each country. It should be noted that non-emancipated minors provided consent via a different process than adults although they are grouped as adults for sampling and reporting. The consent criteria included:

- Women and men aged 18 years and older, living in the selected households, and visitors who slept in the household the night before the survey, who were willing and able to provide written consent
- Adolescents aged 15-17 years, living in the selected households and visitors of the same age who slept in the household the night before the survey, who were willing and able to provide written assent, and whose parents or guardians were willing and able to provide written permission for their participation
    - Parental permission was not required for emancipated minors

Modules included in each questionnaire and their associated eligibility criteria are listed in the table below. The content and order of each module may differ between INSIDA 2021 and other PHIA surveys. Users should refer to the ***Supplement*** for any other surveys they intend to analyze to familiarize themselves with surveys differences.

Table 3. INSIDA 2021 questionnaire

Questionnaire module	Eligibility criteria
Household questionnaire	Sample of households within selected EAs
Household roster	
Household roster for minors	
Deaths in the household	
Household characteristics	
Economic support	
Individual questionnaire – adults (15 years and older)	All eligible[1] and consenting individuals
Respondent background	
Marriage	
Reproductive history	All women
Male circumcision	All men
Sexual activity	
HIV testing history	
HIV status, care and treatment	All self-reporting HIV-positive adults
Tuberculosis and other health issues	
Alcohol use	
Exposure to prevention intervention	All individuals age 15-24

[1] Household members are eligible if they were confirmed to have slept in the household the night before the interview.

4 Biomarker testing

In INSIDA 2021, biomarker testing was offered to all rostered and consenting adults (15+ years). Eligibility criteria for receiving tests for specific biomarkers are provided in the table below.

Table 4. INSIDA 2021 biomarker testing

Biomarker	Eligibility criteria
HIV serostatus[1]	All participants
Limiting Antigen Enzyme (LAg-Avidity)[2]	All HIV+ individuals
CD4+ cell count	All HIV+ individuals
HIV RNA viral load	All HIV+ individuals
Antiretroviral (ARV) drug presence	All HIV+ individuals
ARV drug resistance	All HIV+ individuals with viral load > 200

[1] See HIV testing algorithm below.

[2] Recency of HIV infection is determined via a combination of Limiting Antigen Enzyme (LAg-Avidity) Immunoassay, viral load and ARV results. See “New HIV infections and annual HIV incidence” in the ***PHIA Data Use Manual***.

INSIDA 2021 HIV testing algorithm

For participants 15 years of age or over, initial household-based HIV testing was performed with the national HIV testing algorithm using two HIV rapid tests, as shown in the HIV testing methodology diagram attachment. Individuals with a nonreactive result on an initial HIV rapid screening test (Alere Determine™ HIV-1/2 Ag/Ab) were classified as HIV-negative. Then, individuals with a reactive screening test underwent confirmatory testing using the Trinity Biotech Uni-Gold HIV1/2 rapid diagnostic test. Individuals with a reactive screening test result followed by a non-reactive confirmatory test results were retested by repeating the same series testing process again. Individuals with a second discordant outcome were considered to have inconclusive results and referred to a health facility for retesting in four weeks as per the national guidelines.

All HIV positives identified in the field received confirmatory testing in satellite laboratory using the BioRad Geenius™ HIV 1/2 Supplemental Assay. Individuals who self-reported being HIV-positive but tested HIV-negative received additional laboratory-based HIV testing via DNA qualitative polymerase chain reaction (PCR) test (Roche COBAS AmpliPrep/COBAS Taqman (CAP/CTM) HIV-1 Qualitative Test).

5 Data confidentiality

As noted in the ***Manual***, various risk mitigation actions were used to protect the privacy and confidentiality of respondents in the public use data. Some of these actions apply to all PHIA surveys, while other actions are data-driven decisions motivated by various risk disclosure concerns. These concerns include small counts as a result of certain combinations of variables and values which may introduce individual disclosure risk concerns. This section outlines the variables that have been identified for disclosure risk remediation and the specific data action taken to address the risk concern.

The following date variables were redacted for all PHIA surveys prior to public release:

Table 5. Date variables redacted for all PHIA surveys

Dataset(s)	Filename	Variable
Household	insida2021hh	dieddated01- dieddated07
Adult individual	insida2021adultind	surveystday birthday birthmon

Top-coding is the process of re-coding values above an upper bound to the value of the upper bound. Age for all respondents was top coded at 80. There was also top-coding to collapse small counts with nearby values, in which the data were re-coded so that the highest category contains at least 25 cases or 1 percent of households or individuals reporting the category. Variables that underwent top-coding are listed below:

Table 6. Variables that underwent top-coding

Dataset(s)	Filename	Variable	Top-coding upper bound
Household	insida2021household	ownchiknnum	30
Household	insida2021household	owncownum	10
Household	insida2021household	owndognum	4
Household	insida2021household	owngoatnum	12
Household	insida2021household	roomsleep	6
Household	insida2021household	toiletsharenum	5
Household	insida2021household	diedagey_01-07	80
Household	insida2021household	ownpignum_mz	6
Adult individual	insida2021adultind	agemar	35
Adult individual	insida2021adultind	arvsmissdays	4
Adult individual	insida2021adultind	firstsxage	25
Adult individual	insida2021adultind	husnwif	4
Adult individual	insida2021adultind	lifetimesex	23
Adult individual	insida2021adultind	liveb	11
Adult individual	insida2021adultind	mcage	40
Adult individual	insida2021adultind	medinhmonths	6
Adult individual	insida2021adultind	monthtimes	6
Adult individual	insida2021adultind	numwif	3
Adult individual	insida2021adultind	part12monum	6
Adult individual	insida2021adultind	partage1	80
Adult individual	insida2021adultind	wifliveew	2

Adult individual	insida2021adultind	arvamtm	6
Adult individual	insida2021adultind	chtsthivagelastm1	24
Adult individual	insida2021adultind	chtsthivagem1	36
Adult individual	insida2021adultind	livetimey	66
Adult individual	insida2021adultind	schcomc1_mz	13

Bottom-coding is the process of re-coding values below a lower bound to the value of the lower bound. Bottom-coding was used collapse small counts with nearby values, in which the data were re-coded so that the lowest category contains at least 25 cases or 1 percent of households or individuals reporting the category. Variables that underwent bottom-coding are listed below:

Table 7. Variables that underwent bottom-coding

Dataset(s)	Filename	Variable	Bottom-coding lower bound
Roster	insida2021roster	liveawayy	2019
Household	insida2021household	diedagey_01-07	5
Adult individual	insida2021adultind	agemar	14
Adult individual	insida2021adultind	arvfty	2004
Adult individual	insida2021adultind	cervcntsy	2010
Adult individual	insida2021adultind	firstsxage	14
Adult individual	insida2021adultind	hivcly	2019
Adult individual	insida2021adultind	hivlastnegy	2004
Adult individual	insida2021adultind	hivtesty	2007
Adult individual	insida2021adultind	hivtfposy	2007
Adult individual	insida2021adultind	medinhmonths	1
Adult individual	insida2021adultind	monthwheny	2005
Adult individual	insida2021adultind	partage1-3	14
Adult individual	insida2021adultind	vltestlsty	2019
Adult individual	insida2021adultind	chtsthivagem1	2

The following variables and values were combined with the code for “other” due to small counts or percentages:

Table 8. Variables and values collapsed in to the “other” classification

Dataset(s)	Variable	Value(s)
Roster	liveregionlivecountry_mz	1, 2, 3, 4, 5, 6, 7, 8, 9, 10,11, 14, 15, 16, 22
Household	cookingfuel	8, 9, 95
Household	matexwalls	23, 25, 35
Household	matfloor	21, 31, 35
Household	matroof	11, 34, 35, 36
Household	econsup12_c	1
Household	lightingfuel_mz	8
Household	toilettype_mz	62
Household	watersource_mz	61, 71, 91
Adult Individual	arvloc_mz	3, 5, 6, 7, 8
Adult Individual	hivtstlocation_mz	8, 9, 11, 12
Adult Individual	partrelation1	6, 7
Adult Individual	partrelation2	6, 7
Adult Individual	partrelation3	2, 6, 7
Adult Individual	partlastcndmrsn3_mz	3, 6
Adult Individual	workind_mz	9
Adult Individual	adhivprev_b	1
Adult Individual	arvswitchwhy_mz	3, 6
Adult Individual	chronicmed_e_mz	1
Adult Individual	chronicmed_g_mz	1
Adult Individual	cmethod_a	1
Adult Individual	cmethod_b	1
Adult Individual	cmethod_i	1
Adult Individual	cmethod_j	1
Adult Individual	cmethod_k	1
Adult Individual	drugtype_c_mz	1
Adult Individual	drugtype_d_mz	1
Adult Individual	drugtype_e_mz	1
Adult Individual	hivtstnvrrsn_j_mz	1
Adult Individual	hivtstnvrrsn_k_mz	1
Adult Individual	outregionwhr_mz	13, 15, 17, 19, 21, 22
Adult Individual	reasonaway_mz	6
Adult Individual	whereout_mz	15, 17, 19, 21, 22, 23, 24

The following variables were redacted entirely due to small counts or percentages:

Table 9. Variables that were redacted

Dataset(s)	Filename	Variable
Roster	insida2021roster	supportemot3
Roster	insida2021roster	supportsocial12
Roster	insida2021roster	supportsocial3
Household	insida2021hh	dieddatem_01-05
Household	insida2021hh	ownhorsenum
Household	insida2021hh	hhqown_d
Adult Individual	insida2021adultind	childalive2-5
Adult Individual	insida2021adultind	childbrstfd2-5
Adult Individual	insida2021adultind	childbrstfdnow2-5
Adult Individual	insida2021adultind	chtsthivbirth2-5
Adult Individual	insida2021adultind	chtsthivbrstfd2-3
Adult Individual	insida2021adultind	chtsthivresult2-3
Adult Individual	insida2021adultind	chtsthivresultlast2
Adult Individual	insida2021adultind	deathagemo1
Adult Individual	insida2021adultind	deathagemo2
Adult Individual	insida2021adultind	deathageyr1
Adult Individual	insida2021adultind	deathageyr2
Adult Individual	insida2021adultind	deathageyr4
Adult Individual	insida2021adultind	firstsxagec
Adult Individual	insida2021adultind	partrecent1
Adult Individual	insida2021adultind	prgsyphoffer2_mz
Adult Individual	insida2021adultind	prgsyphoffer3_mz
Adult Individual	insida2021adultind	prgsyphoffer4_mz
Adult Individual	insida2021adultind	prgsyphoffer5_mz
Adult Individual	insida2021adultind	prgsyphpos2_mz
Adult Individual	insida2021adultind	prgsyphpos3_mz
Adult Individual	insida2021adultind	prgsyphtest2_mz
Adult Individual	insida2021adultind	prgsyphtest3_mz
Adult Individual	insida2021adultind	prgsyphtest4_mz
Adult Individual	insida2021adultind	prgsyphtest5_mz
Adult Individual	insida2021adultind	prgsyphtreat1_mz
Adult Individual	insida2021adultind	prgsyphtreat2_mz
Adult Individual	insida2021adultind	arvnrpg_mz
Adult Individual	insida2021adultind	arvsnotcurrsn_mz
Adult Individual	insida2021adultind	arvsnottake_mz
Adult Individual	insida2021adultind	chtsthivagelastm2
Adult Individual	insida2021adultind	chtsthivagem2
Adult Individual	insida2021adultind	chtsthivagem3
Adult Individual	insida2021adultind	hivcnotrsn_mz

The following variables had values collapsed into new categories because of small counts.

Table 10. Variables with new categories

Dataset(s)	Variable	Old Values	New Category
Adult individual	cerncnrslt	2, 3	2 – Abnormal/Positive/Suspect Cancer
		4, 5	5 – Inconclusive/Not received
Adult individual	cerncntrt	1, 2	1 – Yes, treated on the same or a different day
Adult individual	arvamtm	0, 1	1 – 0 or 1
Adult individual	schcom_mz	6, 7	7 – Elementary Technical Education/Basic Technical Education

6 Dataset specifications

Table 11. INSIDA 2021 dataset specifications

Dataset	Filename	Number of observations	Number of variables
Household	insida2021hh	11,375	403
Roster	insida2021roster	38,825	59
Adult individual	insida2021adultind	17,105	474
Adult biomarker	insida2021adultbio	14,488	205
Drug resistance[1]	insida2021drugresistance	TBD	TBD
PSU intermediary weights	Insida2021psuintermediarywts	324	324
Household intermediary weights	insida2021hhintermediarywts	11,375	485
Individual intermediary weights	insida2021indintermediarywts	38,825	649

Dataset specification	Description
Two-letter country code prefix for ID variables	MZ
Survey weighting variables	
No. of jackknife replicates	159
Survey weights provided (variable prefix)	PSU (psuwt) Dwelling (duwt) Household (hhwt) Individual interview (intwt) Blood test (btwt) Drug resistance (drwt)
Selected variable parameters	
Household characteristics used for wealth index construction	*See next section*
Mean duration recent infection (MDRI) used for HIV incidence estimation	130 days (95% CI 118-142 days, standard error 37.48575911)

[1]Drug resistance data will be provided once laboratory testing is complete and data has been cleaned and processed.

7 Wealth index

As described in the ***Manual***, a wealth index is constructed using principal component analysis (PCA) on household characteristics and asset ownership variables. The details of these variables vary by country. The table below lists the variables used to construct the wealth index for INSIDA 2021.

Table 12. Household characteristics used for wealth index construction in INSIDA 2021

Indicator variable	Type	Description
memsleep	Continuous	Number of household members per sleeping room[1]
matroof	Categorical	Dwelling roofing material
matexwalls	Categorical	Dwelling wall material
matfloor	Categorical	Dwelling floor material
toilettype	Categorical	Type of toilet used by the household
watersource	Categorical	Source of water used by the household
cookingfuel	Categorical	Type of cooking fuel used by the household
lightingfuel	Categorical	Type of fuel used for lighting by the household
econsup12	Binary	Any external economic support
For the remainder of the variables:		*Does this household have/own…?*
hhqitems (option A)	Binary	Electricity
hhqitems (option B)	Binary	A working radio
hhqitems (option C)	Binary	A working television
hhqitems (option D)	Binary	A working telephone/mobile telephone
hhqitems (option E)	Binary	A working refrigerator
hhqown (option A)	Binary	A bicycle
hhqown (option B)	Binary	A working motorcycle or motor scooter
hhqown (option C)	Binary	A working car or truck
hhqown (option D)	Binary	A working boat with a motor
		How many of the following does this household have/own? [2]
ownchiknnum	Continuous	Chicken
owncownum	Continuous	Cows
owndognum	Continuous	Dogs
owngoatnum	Continuous	Goats
ownpignum	Continuous	Pigs
ownhorsenum	Continuous	Horses

[1]Rounded to the nearest integer.

[2]For wealth index calculation, continuous variables for animal ownership have been changed into binary (yes/no). For example, the households that had any chickens were assigned “yes” and the households that had no chickens were assigned “no”. This was done to be consistent with the DHS computation of wealth index (see, for example, Moçambique Inquérito Demográfico e de Saúde, 2011).

Wealth scores and model performance

The first component of the PCA model is interpreted as an index of household wealth. However, it does not explain a large proportion of the total variance: it accounts for only around 7.33% of

the total variance in the common model, 4.02% for the urban model, and 5.18% for the rural model. Howe et al. note that this figure is “often less than 20%” (Howe et. Al., 2008). The results from INSIDA 2021 are consistent with those of other DHS studies in similar settings (Vyas and Kumaranayake, 2006; Filmer D and Pritchett, 2001; Moçambique Inquérito Demográfico e de Saúde, 2011).

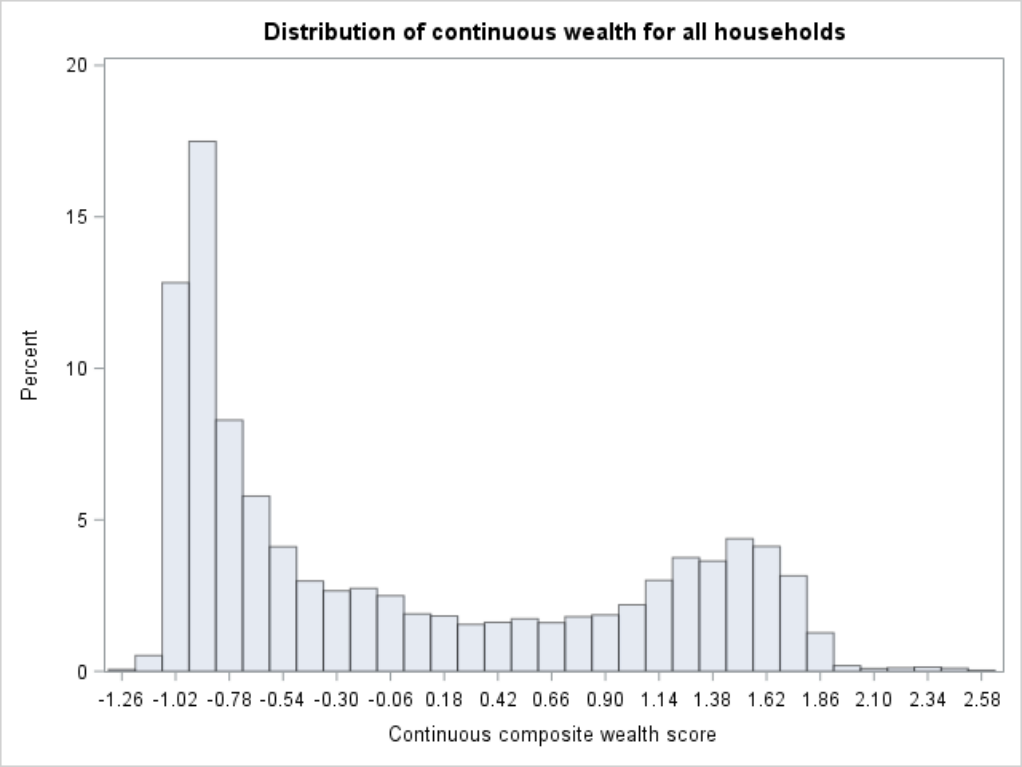
The PCA method does not guarantee the extraction of an index that is actually well-correlated with wealth but results from the PCA can be used to check whether the interpretation of the model makes sense. The component loading for each asset variable describes the association between that asset and the wealth index. The following table shows the most influential variables as measured by absolute value of their loading in each model.

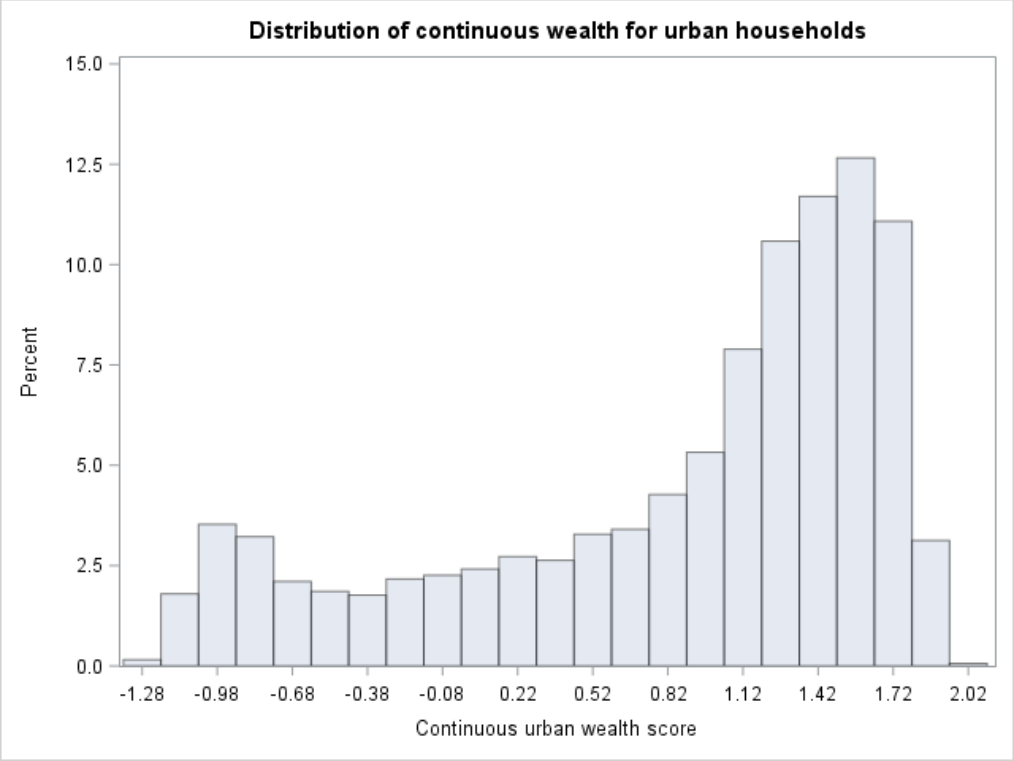
Table 13. PCA results for INSIDA 2021 wealth quintile

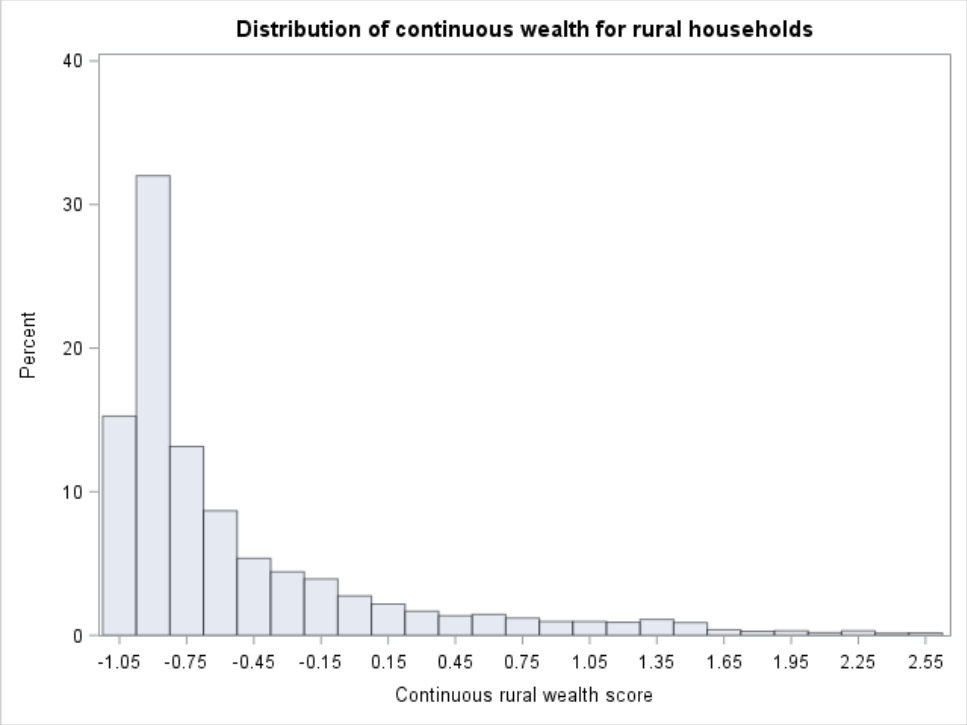
		Component loading		
Variable	**Category**	**Common model**	**Urban model**	**Rural model**
Electricity in the house	Yes	0.85957	0.79661	0.75001
Lighting fuel	Electricity	0.85686	0.79099	0.74447
Television	Yes	0.79295	0.71116	0.69184
Roof material	Thatch/Palm leaf	-0.75791	-0.72665	0.73507
Floor material	Cement/Terazo	0.72105	0.54975	0.73507
Refrigerator	Yes	0.71037	0.61505	0.67614
Cooking fuel	Firewood/Straw	-0.68384	-0.58067	-0.43463
Roof material	Corrugated Iron	0.64337	0.50437	0.62177
Floor material	Earth/Sand	-0.61757	-0.58748	-0.50887
Water source	Piped to yard/plot	0.61316	0.49196	0.57467

In INSIDA 2021, electricity and lighting fuel were particularly important for the determination of wealth score. Note that variables with negative component loadings are associated with lower wealth, while those with positive loadings indicate a wealthier household.

The distribution of wealth index values from the model is shown in the figures below, first the composite wealth index for all households, and then the urban and rural-specific wealth indices. The distribution for the composite wealth index is skewed towards households with lower wealth, with a smaller secondary peak towards the wealthier end of the score range.


Distribution of continuous wealth for all households
Percent
20
15
10
5
0
-1.26 -1.02 -0.78 -0.54 -0.30 -0.06 0.18 0.42 0.66 0.90 1.14 1.38 1.62 1.86 2.10 2.34 2.58
Continuous composite wealth score


Distribution of continuous wealth for urban households
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-1.28 -0.98 -0.68 -0.38 -0.08 0.22 0.52 0.82 1.12 1.42 1.72 2.02
Continuous urban wealth score


Distribution of continuous wealth for rural households
Percent
40
30
20
10
0
-1.05
-0.75
-0.45
-0.15
0.15
0.45
0.75
1.05
1.35
1.65
1.95
2.25
2.55
Continuous rural wealth score

8 References

Instituto Nacional de Saúde (INS). Mozambique Population-based HIV Impact Assessment 2021 (INSIDA 2021): Summary Sheet. Maputo: INS; December 2022.

Instituto Nacional de Saúde (INS). Mozambique Population-based HIV Impact Assessment 2021 (INSIDA 2021): Final Report. Maputo: INS; July 2023.

Filmer D, Pritchett LH. Estimating wealth effects without expenditure data - or tears: An application to educational enrollments in states of India. Demography. 2001;38(1):115-132.

Howe LD, Hargreaves JR, Huttly SR. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. Emerg Themes Epidemiol. 2008; 5:3.(doi):10.1186/1742-7622-1185-1183.

Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan. 2006;21(6):459-468. Epub 2006 Oct 2009.

Ministerio da Saude - MISAU/Moçambique, Instituto Nacional de Estatística - INE/Moçambique and ICF International. Moçambique Inquérito Demográfico e de Saúde 2011. Calverton, Maryland, USA: MISA/Moçambique, INE/Moçambique and ICF International.

9 Attachments

9.1 Questionnaires

Supplement Attachment 1 - INSIDA 2021 Questionnaires

9.2 Codebook with frequencies

Supplement Attachment 2 - INSIDA 2021 Codebook

9.3 Flow Diagrams for selected analytic variables

Supplement Attachment 3 - INSIDA 2021 Flow Diagrams for Analytic Variables.pdf

9.4 HIV Testing Methodology Diagram

Supplement Attachment 4 - INSIDA 2021 Testing Methodology Diagram.pdf

9.5 Sampling and Weighting Technical Report

Supplement Attachment 5 - INSIDA 2021 Sampling and Weighting Technical Report.docx

9.6 INSIDA 2021 Survey-specific table specifications

Supplement Attachment 6 - INSIDA 2021 Additional Table Specifications.xlsx

9.7 Requesting data

INSIDA 2021 data can be requested for use in research and analysis under the following conditions:

- Recipient will use this data only for the purpose of the research and analysis described in this data request. The recipient will submit a new request if they intend to use the data for another purpose.
- Recipient will not share this data with other researchers, with the exception of those listed in this data request as co-researchers for the project.
- Recipient will ensure that co-researchers are aware of and follow the terms of this data use agreement.
- Recipient will treat all data as confidential. Recipient will not use the data to deliberately compromise or otherwise infringe on the anonymity of participants' information and their right to privacy and will not attempt to identify any individual, household, or community in the survey based upon these data.
- Recipient will not publish any result in which participants, EAs or communities can be identified.
- Recipient will keep data in a secure location where it cannot be accessed by unauthorized users.
- Recipient will not use this data for any commercial venture.
- Recipient agrees that this agreement terminates immediately upon any breach by the recipient of the data or any co-researchers.

To see a demonstration of the data request process, watch the video here. The process is described in detail below.

To make a data request, first create an account at https://phia-data.icap.columbia.edu/ using the "Register" button and login using the button at the top right of the page. Once logged in, click "Data Sets" in the top menu to see the list of countries available. For INSIDA 2021, select "Mozambique" from the list.

The top part of the page shows the PHIA survey years and datasets available for request, and the lower part shows the available documentation. Documentation may be downloaded without submitting a request. To obtain access to datasets, select the datasets you require for your project and click "Request Access". Fill out the project title and project description, including the general aims of your research and a brief description of your planned analysis. Fill out any co-researcher details, then click "Next". Read the conditions of use and enter your name to agree to the conditions and submit your request. Requests will generally be reviewed and approved within 1-2 business days. You will receive an email confirmation of approval. Once access has been approved, the check marks beside the requested datasets will be replaced with clickable buttons which will begin downloads of the data.

Requests for PHIA geospatial data have a more rigorous approval process because of the additional privacy and confidentiality risks associated with location data. Requests for geospatial data must explain why geomasked cluster centroid data are essential to the proposed analysis and describe the specific spatial analytical methods that will be used. Refer to the PHIA Geospatial Data Use Manual, available freely on each country's data request page, for full information on the content of the geospatial datasets.

For assistance or for any questions about the data, you can use the help request section at the bottom of https://phia-data.icap.columbia.edu/help to submit a question.

9.8 Data explorer

The ICAP PHIA data site also includes data visualization tools which allow you to look up survey estimates for specific countries and to compare across countries. To access these, visit https://phia-data.icap.columbia.edu/visualization. To see a video demonstration of the data visualization tools, watch the video here. The main steps to create a data visualization are described below.

1. Choose Country

Select the country or countries you are interested in by clicking them on the map, then click "Next".

2. Choose Indicator

Use the "Indicator" drop down to choose the indicator of interest. Typing in the indicator box after clicking the drop down allows you to filter the indicators available. Many indicators include subindicators, which are selected using the subindicator drop down. For example, after selecting the "90-90-90 (self-reported ARV, Overall Percentages)" indicator, you can choose some or all of "Diagnosed", "On Treatment", and "Viral Load Suppression" as subindicators.

3. Specify Age and Gender

The age and gender drop downs allow you to subset the data visualization to include the age group and gender you are interested in.

4. Choose Stratification

Stratification categories allow you to obtain estimates broken down by a range of variables, such as age groups, education, marital status, and others. The available stratification options depend on the indicators selected.

5. Choose Visualization Type

Visualizations can be selected using the “Chart”, “Table”, and “Map” buttons in the top right of the display. The default is Chart, which typically displays a horizontal bar chart showing percentages with a 95% confidence interval, or for some indicators a count or median. The Table option shows the estimates in a tabular format, including columns for each selected option. The Map displays the estimates as a heat map for the selected countries.

6. Download

Chart and Table visuals can be saved by clicking the download button next to the question mark on the top right of the page. For a Chart, the download is a static image of the visual. For a table, a CSV file is generated for download.

For help with the data visualization tools, click the help button question mark in the top right of the page, or visit https://phia-data.icap.columbia.edu/help.