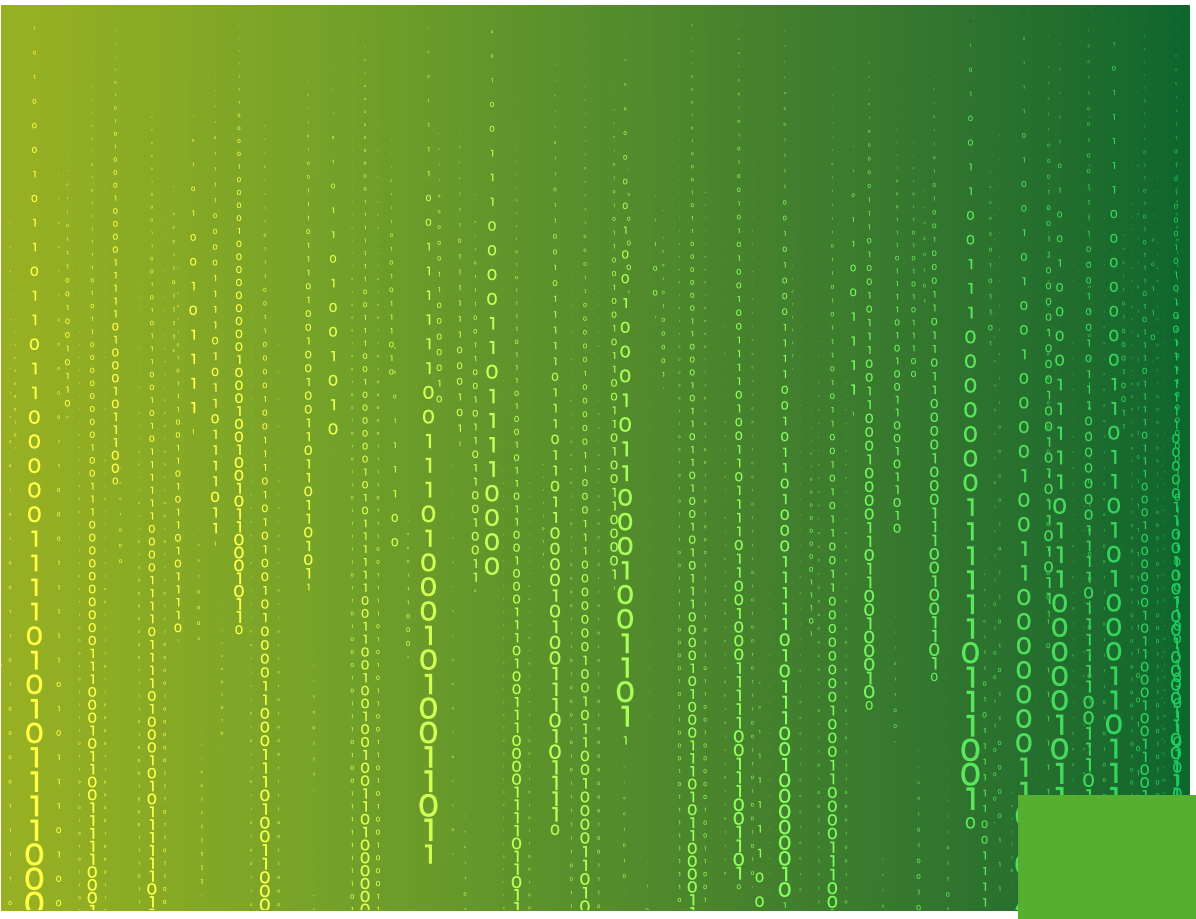
LePHIA 2020

LESOTHO
POPULATION-BASED
HIV IMPACT ASSESSMENT



DATA USE MANUAL SUPPLEMENT

BOPHELO
MINISTRY OF HEALTH

BOS
Bureau of Statistics

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

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Westat

icap Global Health

Lesotho Population-based HIV Impact Assessment 2020

LePHIA 2020

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.

LePHIA 2020 Collaborating Institutions

Lesotho Ministry of Health
Lesotho Bureau of Statistics
National Reference Laboratory
Centers for Disease Control and Prevention (CDC), Lesotho
Centers for Disease Control and Prevention (CDC), Atlanta
ICAP at Columbia University, New York
ICAP at Columbia University, Lesotho
Westat

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

Lesotho Population-based HIV Impact Assessment (LePHIA) 2020 Data Use Manual Supplement. New York, NY. November 2023.

Single dataset

Lesotho Ministry of Health (MOH). Lesotho Population-Based HIV Impact Assessment 2020 (LePHIA 2020) [Dataset]. <Insert dataset name used, e.g. lephia2020adultbio>. Maseru, Lesotho: Lesotho Ministry of Health, Lesotho Bureau of Statistics, National Microbiology Reference Laboratory, Lesotho, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2023.

Multiple datasets (more than 3)

Lesotho Ministry of Health (MOH). Lesotho Population-Based HIV Impact Assessment 2020 (LePHIA 2020) (various) [Datasets]. Maseru, Lesotho: Lesotho Ministry of Health, Lesotho Bureau of Statistics, National Microbiology Reference Laboratory, Lesotho, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2023.

Access this Manual Online

https://phia-data.icap.columbia.edu/datasets?country_id=8

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
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
LePHIA	Lesotho Population-based HIV Impact Assessment
OVC	Orphans and Vulnerable Children
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PHIA	Population-based HIV Impact Assessment

1 Background

1.1 What is LePHIA 2020?

The Lesotho Population-based HIV Impact Assessment 2020 (LePHIA 2020) is a cross-sectional household-based survey conducted in Lesotho. LePHIA 2020 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 LePHIA 2020 Data Manual Supplement

The purpose of the ***LePHIA 2020 Data Manual Supplement*** (hereafter, "***Supplement"***) is 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 packages and their contents, a guide to getting started with the PHIA data, and details on the files and variables included within the PHIA data. This ***Supplement*** contains LePHIA 2020 survey specifications, including survey-specific eligibility criteria, sampling approaches and measures, and survey-specific documentation such as codebooks and questionnaires. A summary of LePHIA 2020 findings can be found in the ***LePHIA 2020 Summary Sheet.***

1.3 Other documentation and resources

In addition to this ***Supplement***, users should refer to the ***Manual*** for general information on PHIA data and PHIA publications such as the ***LePHIA 2020 Summary Sheet*** and ***LePHIA 2020 Final Report***. The ***LePHIA 2020 Final Report*** contains detailed results from LePHIA 2020 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.

Several survey-specific pieces of documentation are provided as attachments to this ***Supplement***, including:

- ***Survey Questionnaires***: Three questionnaires are provided, the 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.
- ***Codebook with Frequencies***: Codebooks are provided for each dataset, indicating all variables contained within and frequencies of all categorical variables. These codebooks document each variable's name, category (i.e., the questionnaire module or source data of the variable), label (i.e., question wording or other label), type (e.g., integer, select one, select multiple, free text, and date/time) and coding values and labels.
- ***Analytic Variable Flow Diagrams***: These flow diagrams define key analytic variables that combine sets of source variables.
- ***Sampling and Weighting Technical Report***: Technical details of sampling and weighting procedures are provided in deeper detail.
- ***Survey-Specific Table Specifications****:* Containing tabulation detailed specifications for any final report tables outside of the general tabulation plan.

With each dataset download there are also statistical programs provided to help users get started with the PHIA data in three commonly used statistical packages: STATA, SAS, and R.

- ***LePHIA 2020 STATA Intro Code.do***: STATA do-file
- ***LePHIA 2020 SAS Intro Code.sas***: SAS program
- ***LePHIA 2020 R Intro Code.R***: R script

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

- ***LePHIA 2020 Formats.SAS***: SAS program

2 Survey design and data collection

LePHIA 2020 was a nationally representative, cross-sectional, two-stage, population-based survey of households across Lesotho. Its target population corresponded to adults, defined in this survey as those aged 15 years and older.

Table 1. LePHIA 2020 survey design characteristics

Survey design characteristics	Description
Survey design	
Data source for survey weighting[1]	2016 Lesotho Census
Sampling stratum	District
Primary sampling unit	Census Enumeration Areas (EA)
Urban/rural categorization	Urban/peri-urban/rural
Survey administration	
Data collection dates	December 2019 – March 2020
Languages	English, Sesotho
Sample size[2]	
Number of selected EAs	342
Number of selected households	11,968
Number of rostered individuals (all ages)	33,282
Survey participation	
Number of completed household interviews	9,665
Number of completed individual interviews	16,468
Number of completed biomarker tests	15,349

[1] See the ***Sampling and Weighting Technical Report*** for more details on survey weighting approach.
[2] See the ***LePHIA 2020 Summary Sheet*** for response rates.

Exceptions to the general PHIA design

As noted in the table above, Lesotho uses a unique three-category urban/rural designation, that contains an additional category of peri-urban. Urban areas were defined by the Bureau of Statistics, and were characterized by either high population density, or of a high level of economic activities or infrastructure. Peri-urban areas were defined as areas with moderate population density, or a lesser extent of economic activities or infrastructure. Rural areas were those with only minimal population density or little infrastructure or economic activities.[5]

Data users should note a mid-survey correction made to the skip logic for hivposprov, “Has a health care provider ever told you that you have HIV?” in February 2020. Individuals interviewed in December 2019 and January 2020 who reported ever testing for HIV (hivtstever = 1) and said that their HIV test result was something other than positive (hivtstrslt = 2, 3, 4, -8, -9) should have received the question hivposprov but did not. Individuals interviewed in February 2020 and later meeting these criteria did receive the question.

Questionnaire Changes

There were several country-specific changes to the questionnaire in LePHIA 2020. 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 an "_ls" suffix. The full list of questions having country-specific changes is as follows:

Question	Variable Name
Have you taken part in any of the following HIV prevention programs?	adhivprev_[option]_ls
Have you taken part in any of the following HIV prevention programs in the last 12 months?	adhivprev12_[option]_ls
Ever talked about HIV with peers and/or sexual partner(s)	adtpeershiv_ls
Ever talked with peers about sex	adtpeerssx_ls
In the last 12 months, taken ARVs from South Africa	arvlocsa_ls
In the last 12 months, paid money or gave gifts or favours for sex	buysx12mo_ls
Condom used last time paid money or gave gifts for sex	buysxcndm_ls
Ever paid money or given gifts/favours for sex	buysxever_ls
Ever tried drugs such as dagga, ex, cocaine, heroin, lean or others	everdr_ls
Drugs ever tried	everdrtype_[option]_ls
Language of respondent	lngnat_lng_ls
Language of interview	lngvint_lng_ls
Language of questionnaire	lngvqx_lng_ls
Province or country lived in before current	outregionwhr_ls
Age when started taking PrEP	prepage_ls
Length of time taking PrEP	preplength_ls
In the last 12 months, had sex for money, gifts, or favours	sellsx12mo_ls
Condom used last time had sex for money, gifts, or favours	sellsxcndm_ls
Ever had sex for money, gift, or favours	sellsxever_ls
Which province or country is [household member name] in currently?	liveregionlivecountry_ls
Has your household received any of the following forms of external economic support in the last 12 months?	econsup12_[option]_ls
In past 4 weeks, household had no food due to lack of resources	hhqfood_ls
In the past 4 weeks, frequency of no food in the household	hhqfoodfrq_ls
In the past 4 weeks, any household member went to sleep hungry	hhqhung_ls
In the past 4 weeks, frequency of household member going to sleep hungry	hhqhungfrq_ls

Items owned by household members	hhqown_[option]_ls
Main material of exterior walls	matexwalls_ls
Main material of roof	matroof_ls
Number of pigs owned by household members	ownpignum_ls

3 Overview of survey questionnaires

In participating households, a household questionnaire is administered to the designated household head. Household head is defined as an individual age 18 or older and emancipated minors (defined in Lesotho as an individual aged 15-17 who is married, the parent of a child, or has left home and is self-sufficient.) The household head provides consent for the household to participate in the survey, after which individual members are rostered during the household interview.

Then, adult individual questionnaires are administered to eligible and consenting individuals aged 15 and older in the household. Consent criteria are determined in each country. It should be noted that non-emancipated minors are consented 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 verbal 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 verbal assent, and whose parents or guardians were willing and able to provide verbal 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 LePHIA 2020 and other PHIA surveys. Users can refer to each PHIA survey's ***Survey Questionnaires*** and ***Codebooks*** provided as attachments to this document.

Table 2. LePHIA 2020 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 LePHIA 2020, biomarker testing is offered to all rostered and consenting adults (15+ years). Eligibility criteria for receiving tests for specific biomarkers are provided in the table below.

Table 3. LePHIA 2020 biomarker testing

Biomarker	Eligibility criteria
HIV serostatus[1]	All participants
Recency of HIV infection[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[3]	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 Manual***.
[3] ARV drug resistance data have been reported in some ***PHIA Publications*** but are not currently available for download. These data may be available with a future release.

LePHIA 2020 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, see the attached ***HIV testing methodology diagram***. The Lesotho HIV rapid testing algorithm applies two tests in sequence: Determine™ and Unigold™. As per the serial testing algorithm above, individuals with a non-reactive result on the screening test (Determine™) was reported as HIV-negative. Individuals with a reactive screening test underwent subsequent testing with Unigold™. Those with a reactive result on both screening and confirmatory tests were classified as HIV positive and were referred to the health facility for enrollment into care, as required by the national testing algorithm. Individuals with a reactive Determine™ test followed by a non-reactive Unigold™ test were classified as indeterminate and were immediately retested in parallel in the field. If during the parallel testing the results were repeatedly indeterminate, the individual was classified as indeterminate and was referred for testing within 2 weeks as per the national guidelines. For the purposes of the survey, samples with indeterminate results received further testing and evaluation to allow for final classification of HIV status.

For participants who self-report being HIV-positive but who test HIV-negative at the time of the survey, an additional HIV RT testing was conducted at the satellite/mobile lab (following the same national algorithm) as well as HIV total nucleic acid (TNA) polymerase chain reaction (PCR) for confirmation of HIV status. Survey staff was trained on how to interpret the initial RT results for these participants and provide counselling as appropriate. Survey staff returned to the HH after consultation with the MoH, for these participants to provide counseling on the final confirmed result. The survey team visiting these participants were trained to provide counseling on the interpretation of the laboratory results.

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 4. Date variables redacted for all PHIA surveys

Dataset(s)	Filename	Variable
Household	lephia2020hh	dieddated_01- dieddated_05
Adult individual	lephia2020adultind	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 5. Variables that underwent top-coding

Dataset(s)	Filename	Variable	Top-coding upper bound
Roster	lephia2020roster	age	80
Adult individual	lephia2020adultind	age	80
Adult biomarker	lephia2020adultbio	age	80
Household	lephia2020hh	ownchiknnum	30
Household	lephia2020hh	owncownum	10
Household	lephia2020hh	owndognum	6
Household	lephia2020hh	owngoatnum	70
Household	lephia2020hh	ownhorsenum	5
Household	lephia2020hh	ownpignum_ls	5
Household	lephia2020hh	roomsleep	5
Household	lephia2020hh	diedagey_01-04	80
Adult individual	lephia2020adultind	agemar	40
Adult individual	lephia2020adultind	arvsmissdays	5
Adult individual	lephia2020adultind	childa2012	6
Adult individual	lephia2020adultind	chtsthivagelastm1	36
Adult individual	lephia2020adultind	chtsthivagem1	36
Adult individual	lephia2020adultind	firstsxage	40
Adult individual	lephia2020adultind	husnwif	2
Adult individual	lephia2020adultind	lifetimesex	40
Adult individual	lephia2020adultind	liveb	10
Adult individual	lephia2020adultind	livetimey	65
Adult individual	lephia2020adultind	mcage	40
Adult individual	lephia2020adultind	medinhmonths	6

Adult individual	lephia2020adultind	monthtimes	10
Adult individual	lephia2020adultind	numwif	2
Adult individual	lephia2020adultind	part12monum	6
Adult individual	lephia2020adultind	partage1	80
Adult individual	lephia2020adultind	partage2	80
Adult individual	lephia2020adultind	wifliveew	2

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 6. Variables that underwent bottom-coding

Dataset(s)	Filename	Variable	Bottom-coding lower bound
Adult individual	lephia2020adultind	agemar	14
Adult individual	lephia2020adultind	arvfty	2004
Adult individual	lephia2020adultind	cervcntsy	2010
Adult individual	lephia2020adultind	firstsxage	14
Adult individual	lephia2020adultind	hivcly	2018
Adult individual	lephia2020adultind	hivlastnegy	2000
Adult individual	lephia2020adultind	hivtesty	2000
Adult individual	lephia2020adultind	hivtfposy	2000
Adult individual	lephia2020adultind	medinhmonths	1
Adult individual	lephia2020adultind	monthwheny	1990
Adult individual	lephia2020adultind	partage1	14
Adult individual	lephia2020adultind	partage2	14
Adult individual	lephia2020adultind	partage3	14
Adult individual	lephia2020adultind	vltestlsty	2016
Roster	lephia2020roster	liveawayy	2018

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

Table 7. Variables and values collapsed in to the “other” classification

Dataset(s)	Filename	Variable	Value(s)
Adult individual	lephia2020adultind	adhivprev12_aa_ls	1
Adult individual	lephia2020adultind	adhivprev12_ab_ls	1
Adult individual	lephia2020adultind	adhivprev12_ac_ls	1
Adult individual	lephia2020adultind	adhivprev12_ae_ls	1
Adult individual	lephia2020adultind	adhivprev12_af_ls	1
Adult individual	lephia2020adultind	adhivprev12_ag_ls	1
Adult individual	lephia2020adultind	adhivprev12_ah_ls	1
Adult individual	lephia2020adultind	adhivprev12_ai_ls	1
Adult individual	lephia2020adultind	adhivprev12_am_ls	1
Adult individual	lephia2020adultind	adhivprev12_an_ls	1
Adult individual	lephia2020adultind	adhivprev12_ao_ls	1
Adult individual	lephia2020adultind	adhivprev12_ap_ls	1
Adult individual	lephia2020adultind	adhivprev12_ar_ls	1
Adult individual	lephia2020adultind	adhivprev12_as_ls	1
Adult individual	lephia2020adultind	adhivprev12_at_ls	1
Adult individual	lephia2020adultind	adhivprev12_au_ls	1
Adult individual	lephia2020adultind	adhivprev12_av_ls	1
Adult individual	lephia2020adultind	adhivprev12_aw_ls	1
Adult individual	lephia2020adultind	adhivprev12_ba_ls	1
Adult individual	lephia2020adultind	adhivprev12_bc_ls	1
Adult individual	lephia2020adultind	adhivprev_ab_ls	1
Adult individual	lephia2020adultind	adhivprev_ac_ls	1
Adult individual	lephia2020adultind	adhivprev_ae_ls	1
Adult individual	lephia2020adultind	adhivprev_af_ls	1
Adult individual	lephia2020adultind	adhivprev_ag_ls	1
Adult individual	lephia2020adultind	adhivprev_ah_ls	1
Adult individual	lephia2020adultind	adhivprev_ai_ls	1
Adult individual	lephia2020adultind	adhivprev_aj_ls	1
Adult individual	lephia2020adultind	adhivprev_am_ls	1
Adult individual	lephia2020adultind	adhivprev_ao_ls	1
Adult individual	lephia2020adultind	adhivprev_ap_ls	1
Adult individual	lephia2020adultind	adhivprev_ar_ls	1
Adult individual	lephia2020adultind	adhivprev_at_ls	1
Adult individual	lephia2020adultind	adhivprev_au_ls	1
Adult individual	lephia2020adultind	adhivprev_av_ls	1
Adult individual	lephia2020adultind	adhivprev_aw_ls	1
Adult individual	lephia2020adultind	adhivprev_ba_ls	1
Adult individual	lephia2020adultind	arvloc	4, 5
Adult individual	lephia2020adultind	partlastcndmwhy1	2, 5, 9
Adult individual	lephia2020adultind	partlastcndmwhy2	2, 5, 9
Adult individual	lephia2020adultind	partlastcndmwhy3	2, 3, 6, 7, 8, 9, 10, 11
Roster	lephia2020roster	relattohh	9

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

Table 8. Variables that were redacted

Dataset(s)	Filename	Variable
Household	lephia2020hh	dieddatem_02-04
Adult individual	lephia2020adultind	arvnrpg
Adult individual	lephia2020adultind	arvsnotcurrsn
Adult individual	lephia2020adultind	birthmon
Adult individual	lephia2020adultind	childalive2-5
Adult individual	lephia2020adultind	childbrstfd2-5
Adult individual	lephia2020adultind	childbrstfdnow2-5
Adult individual	lephia2020adultind	chtsthivagelastm2-5
Adult individual	lephia2020adultind	chtsthivagem2-5
Adult individual	lephia2020adultind	chtsthivbirth2-5
Adult individual	lephia2020adultind	chtsthivbrstfd2-5
Adult individual	lephia2020adultind	chtsthivresult2-5
Adult individual	lephia2020adultind	chtsthivresultlast2-5
Adult individual	lephia2020adultind	deathagemo1-2
Adult individual	lephia2020adultind	deathageyr1-2
Adult individual	lephia2020adultind	firstsxagec
Adult individual	lephia2020adultind	livevisitor
Roster	lephia2020roster	hhcemanc
Roster	lephia2020roster	supportsocial3

The following variables had new categories created to ensure that the number of responses in each category met the minimum size requirements.

Table 9. Variables with new categories

Dataset(s)	Filename	Variable	New Categories
Adult individual	lephia2020adultind	arvamtm (number of months of ARV supply given last time)	1 – 0 or 1 2 – 2 3 – 3 4 – 4 or 5 5 – 6 or more
Adult individual	lephia2020adultind	arvsnottake (main reason never taken ARVs)	123 – Not eligible for treatment, health care provider did not prescribe, or HIV medicines not available 89 – Not attending HIV clinic or clinic is too far
Adult individual	lephia2020adultind	cerncrslt (Last cervical cancer test result)	2 – Abnormal/Positive/Suspect cancer 5 – Unclear/Inconclusive/Not received
Adult individual	lephia2020adultind	cerncntrt (Received treatment after last cervical cancer test)	1 – Yes, treated on same or different day
Adult individual	lephia2020adultind	schcom	7 – Level 3 (Master's or Doctoral Program)

6 Dataset specifications

Table 9. LePHIA 2020 dataset specifications

Dataset (filename)		**Number of observations**	**Number of variables**
Household	lephia2020hh	11,968	244
Roster	lephia2020roster	33,282	61
Adult individual	lephia2020adultind	16,468	487
Adult biomarker	lephia2020adultbio	15,349	215
Household intermediary weights	lephia2020hhintermediarywts	11,968	173
Individual intermediary weights	lephia2020indintermediarywts	33,282	683
Dataset specification		**Description**	
Two-letter country code prefix for ID variables		LS	
Survey weighting variables			
No. of jackknife replicates		168	
Survey weights provided (variable prefix)		Household (hhwt) Individual interview (intwt) Blood test (btwt)	
Selected variable parameters			
Household characteristics used for wealth index construction		*See next section*	
Mean duration recent infection used for HIV incidence estimation		130 days (95% CI 118-142 days)	

7 Wealth index

As described in the ***Manual***, a wealth index is constructed using principal components analysis (PCA) on household characteristics and asset ownership variables that can vary by country. The table below lists the variables used to construct the wealth index for LePHIA 2020.

Table 10. Household characteristics used for wealth index construction in LePHIA 2020

Indicator variable	Type	Description
memsleep	Numeric (count)	Number of household members per sleeping room
matroof_ls	Categorical	Dwelling roofing material
matexwalls_ls	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_ls	Categorical	Type of cooking fuel used 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
hhqown (option E)	Binary	A scotch cart
		How many of the following does this household have/own? [1]
ownchiknnum	Continuous	Chicken
owncownum	Continuous	Cows
owndognum	Continuous	Dogs
owngoatnum	Continuous	Goats
ownhorsenum	Continuous	Horses
ownpignum_ls	Continuous	Pigs

[1]For wealth index calculation, continuous variables have been changed into binary (yes/no). For example, the households that had any chickens will be assigned “yes”, and the households that and no chickens will be assigned “no”. This was done to be consistent with the DHS computation of wealth index.[4]

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 7.09% of the total variance in the common model, 4.81% for the urban model, 4.55% for the peri-urban model, and 5.85% for the rural model. Howe et al. note that this figure is “often less than 20%”.[1]

The results from LePHIA 2020 are consistent with those of other DHS studies in similar settings.[2-4]

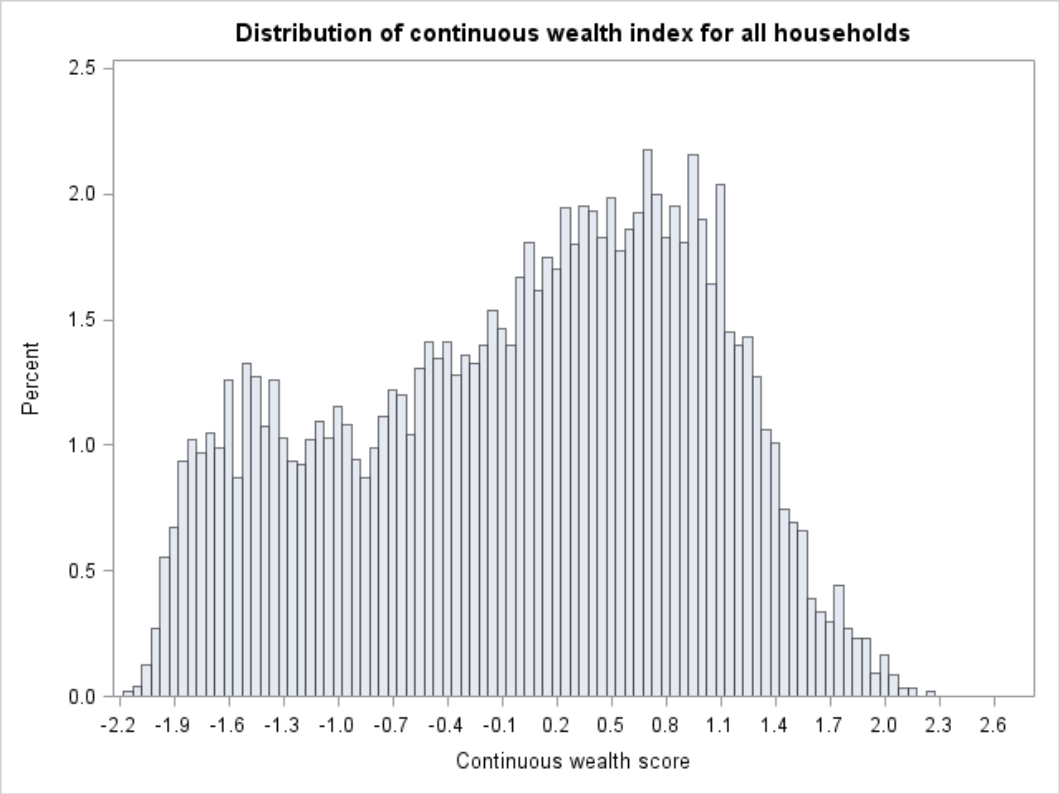
The PCA method does not guarantee the extraction of an index that is actually well-correlated with wealth but results from the PCA can be used to check whether the interpretation of the model makes sense. The component loading for each asset variable describes the association between that asset and the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

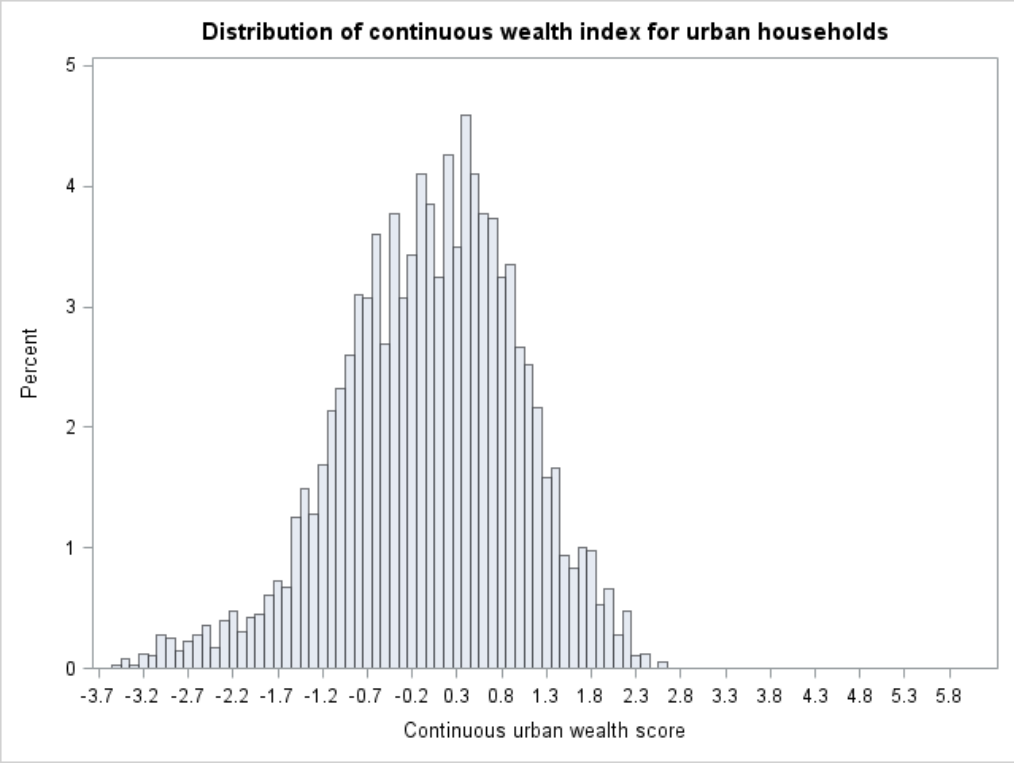
Table 11. PCA results for LePHIA 2020 wealth quintile

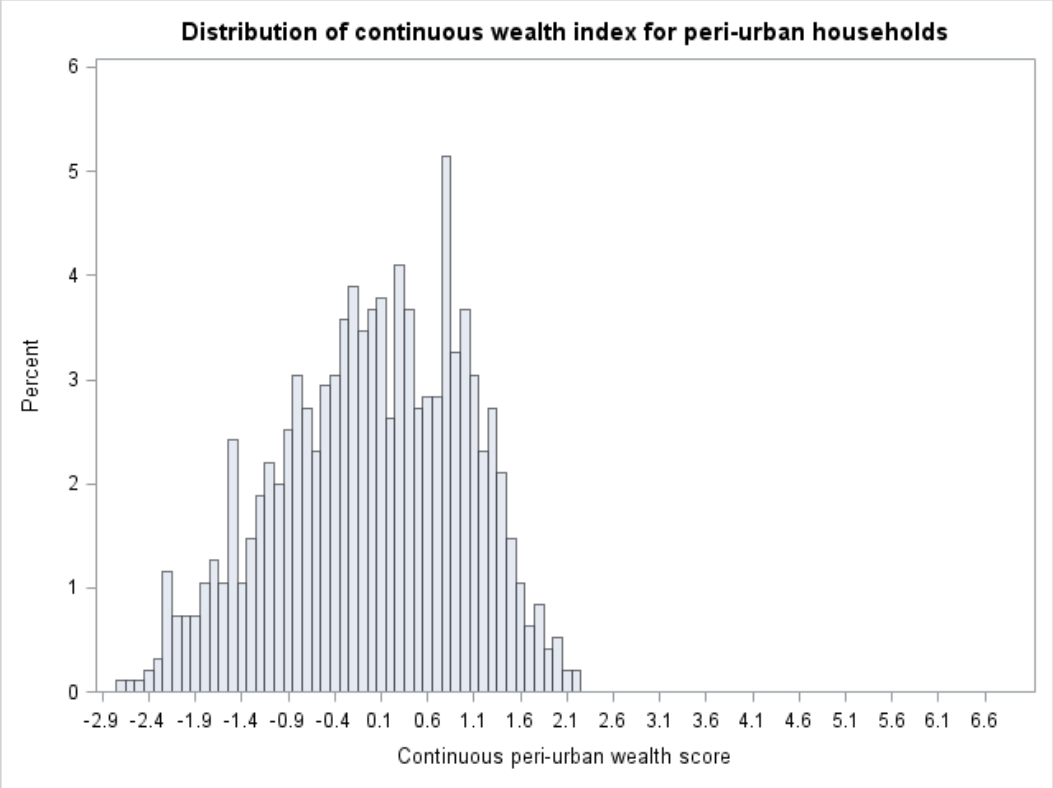
		Component loading			
Variable	**Category**	**Common model**	**Urban model**	**Peri-urban model**	**Rural model**
Electricity in the house	Yes	0.67276	0.65532	0.45594	0.6286
Water source	Piped to yard/plot	0.6279	0.3404	0.51121	0.30607
Television	Yes	0.62593	0.70964	0.56655	0.62359
Refrigerator	Yes	0.60502	0.69603	0.61795	0.61654
Cooking fuel	Firewood/straw/shrubs	-0.71937	-0.45375	-0.4997	-0.57598
Dwelling roofing material	Thatch/grass	-0.6905	-0.36762	-0.54755	-0.68951
Dwelling flooring material	Earth/sand/mud	-0.63354	-0.41098	-0.55721	-0.61824

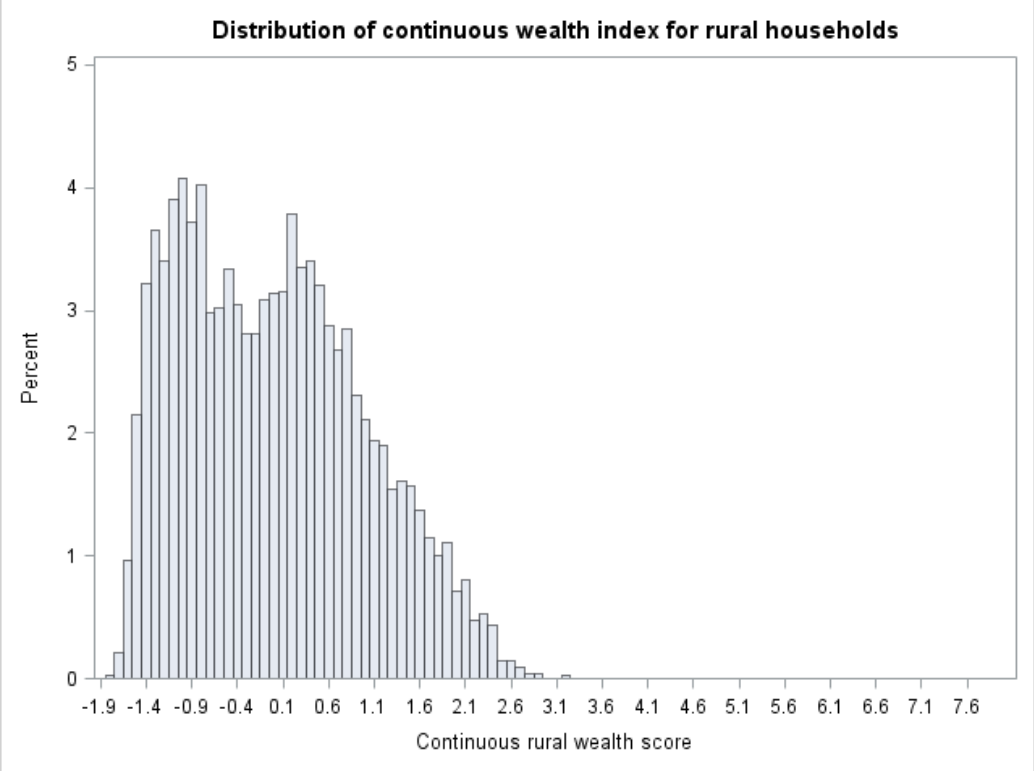
Electricity in the house and water source were particularly important for the determination of wealth score. Note that variables with negative component loadings are associated with lower wealth, while those with positive loadings indicate more wealthy households.

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, peri-urban, and rural-specific wealth indices. The distribution for the composite wealth index is skewed towards households with higher wealth, with a smaller secondary peak towards the lower end of the score range.


Distribution of continuous wealth index for all households
Percent
2.5
2.0
1.5
1.0
0.5
0.0
-2.2 -1.9 -1.6 -1.3 -1.0 -0.7 -0.4 -0.1 0.2 0.5 0.8 1.1 1.4 1.7 2.0 2.3 2.6
Continuous wealth score


Distribution of continuous wealth index for urban households
Percent
5
4
3
2
1
0
-3.7 -3.2 -2.7 -2.2 -1.7 -1.2 -0.7 -0.2 0.3 0.8 1.3 1.8 2.3 2.8 3.3 3.8 4.3 4.8 5.3 5.8
Continuous urban wealth score


Distribution of continuous wealth index for peri-urban households
Percent
0
1
2
3
4
5
6
-2.9 -2.4 -1.9 -1.4 -0.9 -0.4 0.1 0.6 1.1 1.6 2.1 2.6 3.1 3.6 4.1 4.6 5.1 5.6 6.1 6.6
Continuous peri-urban wealth score


Distribution of continuous wealth index for rural households
Percent
0
1
2
3
4
5
-1.9 -1.4 -0.9 -0.4 0.1 0.6 1.1 1.6 2.1 2.6 3.1 3.6 4.1 4.6 5.1 5.6 6.1 6.6 7.1 7.6
Continuous rural wealth score

8 References

1. Lesotho Ministry of Health (MOH). Lesotho Population-based HIV Impact Assessment 2020 (LePHIA 2020): Final Report. Maseru: MOH; June 2022. Available at https://phia.icap.columbia.edu/lesotho-final-report-2020/
2. Lesotho Ministry of Health (MOH). Lesotho Population-based HIV Impact Assessment 2020 (LePHIA 2020): Summary Sheet. Maseru: MOH; July 2021. Available at https://phia.icap.columbia.edu/lesotho-summary-sheet-2/
3. 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.
4. 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.
5. 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.
6. Ministry of Health/Lesotho and ICF International. 2016. Lesotho Demographic and Health Survey 2014. Maseru, Lesotho: Ministry of Health/Lesotho and ICF International. Available at http://dhsprogram.com/pubs/pdf/FR309/FR309.pdf
7. Ministry of Health, Lesotho, Centers for Disease Control and Prevention (CDC), and ICAP at Columbia University. Lesotho Population-based HIV Impact Assessment (LePHIA) 2016-2017: Final Report. Maseru, Lesotho, Atlanta, Georgia, and New York, New York, USA: Ministry of Health, CDC, and ICAP, September 2019.

9 Attachments

9.1 Questionnaires

LePHIA 2020 Data Manual Supplement Attachment 1 - Questionnaires.xlsx

9.2 Codebook with frequencies

LePHIA 2020 Data Manual Supplement Attachment 2 - Codebook.pdf

9.3 Flow Diagrams for selected analytic variables

LePHIA 2020 Data Manual Supplement Attachment 3 - Flow Diagrams for Analytic Variables.pdf

9.4 HIV Testing Methodology Diagram

LePHIA 2020 Data Manual Supplement Attachment 4 - Testing Methodology Diagram.pdf

9.5 Sample design and weighting report

LePHIA 2020 Data Manual Supplement Attachment 5 - Sampling and Weighting Technical Report.pdf

9.6 LePHIA 2020 Survey-specific table specifications

LePHIA 2020 Data Manual Supplement Attachment 6 - Additional Table Specifications.xlsx

9.7 Requesting data

LePHIA 2020 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 LePHIA 2020, select "Lesotho" 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.