UPHIA

2020

**Uganda
Population-based
HIV Impact Assessment**

DATA USE MANUAL SUPPLEMENT
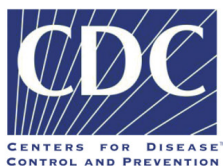
Uganda Population-based HIV Impact Assessment 2020-2021
UPHIA 2020-2021

UPHIA 2020-2021 Collaborating Institutions

Ministry of Health Uganda
Uganda Bureau of Statistics (UBOS)
Uganda Virus Research Institute (UVRI)
Centers for Disease Control and Prevention (CDC), Uganda
Centers for Disease Control and Prevention (CDC), Atlanta
ICAP at Columbia University, New York
ICAP at Columbia University, Uganda
Westat

Donor Support

Suggested Citation

Manual

Uganda Population-based HIV Impact Assessment (UPHIA) 2020-2021 Data Manual Supplement. New York, NY. September 2024.

Manual

Uganda Population-based HIV Impact Assessment (UPHIA) 2020-2021 Data Manual Supplement. New York, NY. September 2024.

Single dataset

Ministry of Health (MOH), Uganda. Uganda Population-Based HIV Impact Assessment 2020-2021 (UPHIA 2020-2021) [Dataset]. <Insert dataset name used, e.g. uphia2020adultbio>. Kampala: MOH, Uganda Bureau of Statistics, Uganda Virus Research Institute, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2024.

Multiple datasets (more than 3)

Ministry of Health (MOH), Uganda. Uganda Population-Based HIV Impact Assessment 2020-2021 (UPHIA 2020-2021) [Datasets].Kampala: MOH, Uganda Bureau of Statistics, Uganda Virus Research Institute, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2024.

Access this Manual Online

https://phia-data.icap.columbia.edu/datasets?country_id=7

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
UPHIA	Uganda 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 UPHIA 2020-2021?

The Uganda Population-based HIV Impact Assessment 2020-2021 (UPHIA 2020-2021) was a cross-sectional household-based survey conducted in Uganda. UPHIA 2020-2021 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.

Please note that UPHIA 2020-2021 was the second PHIA survey in Uganda. The previous PHIA in Uganda is referred to as UPHIA 2016-2017.

1.2 Purpose of the UPHIA 2020-2021 Data Manual Supplement

The purpose of the *UPHIA 2020-2021 Data Manual Supplement* (hereafter, "*Supplement"*) contains UPHIA 2020-2021 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 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 UPHIA 2020-2021 data.

1.3 Other documentation and resources

In addition to the *Manual* and this *Supplement*, users should refer to UPHIA 2020-2021 publications such as the *UPHIA 2020-2021 Summary Sheet* and *UPHIA 2020-2021 Final Report*. The *UPHIA 2020-2021 Summary Sheet* contains highlights and summary results from the survey. The *UPHIA 2020-2021 Final Report* contains detailed results from UPHIA 2020-2021 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.

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.

- o ***UPHIA 2020-2021 Stata Intro Code.do***: STATA do-file
- o ***UPHIA 2020-2021 SAS Intro Code.sas***: SAS program
- o ***UPHIA 2020-2021 R Intro Code.R***: R script

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

- o ***UPHIA 2020-2021 formats.sas***: SAS program

2 Survey design and data collection

UPHIA 2020-2021 was a nationally representative, cross-sectional, two-stage, population-based survey of households across Uganda. Its target population corresponded to adults, defined in this survey as those aged 15 years and older.

Table 1. UPHIA 2020-2021 survey design characteristics

Survey design characteristics	Description
Survey design	
Data source for survey weighting[1]	2014 National Population and Housing Census in Uganda
Sampling stratum	Region
Primary sampling unit	2014 National Population and Housing Census Enumeration Areas (EAs)
Urban/rural categorization	Urban/rural
Survey administration	
Data collection dates	February 20, 2020 – March 10, 2021

	Survey paused:
	    March 24, 2020 – October 26, 2020
	    December 17, 2020 – January 25, 2021

Languages	English, Ateso, Karamajong, Luganda, Lugbara, Luo, Runyankole, Rukiga, Runyoro-Rutoro
Sample size[2]	
Number of selected EAs	313
Number of selected households	11,130
Number of rostered individuals (all ages)	51,572
Survey participation	
Number of completed household interviews	10,164
Number of completed individual interviews	26,071
Number of completed biomarker tests	25,479

[1] See the ***Sampling and Weighting Technical Report*** for more details on survey weighting approach.
[2] See the ***UPHIA 2020-2021 Summary Sheet*** for response rates.

Exceptions to the general PHIA design

Data collection

The UPHIA 2020-2021 survey was paused due to the COVID-19 pandemic from late March 2020 until late October 2020. Prior to resuming data collection, questions were added asking about external economic support, and impact on HIV testing, care, and treatment services related to the COVID-19 pandemic. Additionally, response categories for COVID-19 were added to questions about reasons away from home, never testing for HIV or accessing HIV care and treatment, and reasons why it is not easy to get a condom. Data users should note that the COVID-19 questions

and responses were only included in interviews after data collection resumed, which is a non-random subsample of the Uganda population.

De jure individuals had data collected on them, which is not the typical design. These individuals data are not included in the individual interview and biomarker datasets.

Questionnaire Changes

There were several country-specific changes to the questionnaire in UPHIA 2020-2021. 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 "_ug" 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 prevention or treatment programs?	adhivprev_[a-g,w,x]_ug
The last time you picked up or received your ARVs, were you told that you were being given an extra supply because of the COVID-19 shutdown?	arvamtcovid_ug
Was there any period since March 2020 when your ARV treatment was interrupted due to the COVID-19 shutdown?	arvinterrcovid_ug
Was there any period since March 2020 when you obtained (or were told to obtain) your ARVs in a different way or place than where you usually receive them?	arvloccovid_ug
What was the main reason you did not take ARVs while you were pregnant with [CHILDNAME]?	arvnrpg_ug
Can you tell me the main reason why you stopped taking ARVs?	arvsnotcurrsn_ug
What is the main reason you have never taken ARVs?	arvsnottake_ug
Why were your ARVs changed?	arvswitchwhy_ug
Where did you give birth to [child name] during your last pregnancy? Please give your best guess.	brthwhr_ug
Have you ever been told by a doctor or health worker that you have any of the following chronic health conditions?	chroniccond_[a-j,x]_ug
Are you currently taking medication for any of the following chronic health conditions?	chronicmed_[a-j,x]_ug
Why is it not easy for you to get a condom?	condomnoteasyrsn_[a,c-g,v,w,x]_ug
Where can you get condoms?	condomwhere_[a-g,x]_ug
Was any of this external economic support related to COVID-19?	econsupcovid_ug
What is your ethnic group/tribe?	ethnic_ug
Does your household have any of the following items?	hhqitems_[a-g]_ug
What is the main reason why you have never received care or treatment for HIV from a doctor, clinical officer, or nurse?	hivcnotrsn_ug
If a self-test kit was available, would you use it?	hivselftsta_ug
Where was your last HIV test done?	hivtstlocation_ug
Is your husband or partner living with you now or is he staying elsewhere?	huslivew_ug
Which district or country does [NAME] currently live in?	livedistrict_ug/livecountry_ug
Native language of participant	lngnat_lng_ug
Language of interview	lngvint_lng_ug
Language of questionnaire	lngvqx_lng_ug
How old were you when you were circumcised by the medical provider? Please give your best guess.	mcage_ug
Are you circumcised?	mcstatus_ug
Just before you moved here, did you live in Kampala, in a town, or in a rural area?	outregiontype_ug
Just before you moved here, which district did you live in? If you lived outside of Uganda, which country did you live in?	outregionwhrd_ug/ outregionwhrc_ug
What is the main reason you did not visit a clinic for antenatal care when you were pregnant with [child name]?	pregncr_ug
What was the main reason you away from home for more than a month?	reasonaway_ug
What is your religion?	religion_ug

What is the highest class you completed?	schcom_ug
In the last 12 months, did you visit a clinic for TB diagnosis or treatment?	tbclinvisit_ug
In the last 12 months, were you told by a health care worker (such as a doctor, clinical officer, nurse or lab tech) that you had TB?	tbdiagn_ug
The last time you were away from home for more than a month, where were you?	whereout_ug
The last time you were away from home for more than a month, where were you?	whereoutd_ug/whereoutc_ug
You mentioned that you have [#] wife/wives who live elsewhere. Where are they?	wifewhere_ug
What is your occupation? That is, what kind of work do you mainly do?	workind_ug

Most of the changes correspond to additional or slightly modified response options. More substantial changes were made to the TB clinic attendance module and the male circumcision module and are described below. In addition, questions about pre-exposure prophylaxis (PrEP) were not included in the questionnaire for UPHIA 2020-2021.

In TB clinic attendance module, the question "In the last 12 months, did you visit a clinic for TB diagnosis or treatment?" (tbclinvisit) typically has response categories of 1 = Yes, 2 = No, -8 = Don't know and -9 = Refuse. However, in Uganda, the response category of 1 = Yes was replaced by 11 = Yes, visited a TB clinic and 12 = Yes, visited a non-TB clinic. The standard tables 13.3 and 13.4 were updated to indicate individuals who visited a TB or non-TB clinic for TB diagnosis and treatment and are provided in the attachments to this supplement as country-specific tables.

The male circumcision status question, "Are you circumcised" (mcstatus), was also modified to have categories of 1 = Yes, 2 = No, -8 = Don't know and -9 = Refuse, instead of 1 = Yes, fully circumcised, 2 = Yes, partially circumcised, 3 = Not circumcised, -8 = Don't know, and -9 = Refuse. The question "How old were you when you were circumcised?" (mcage) was modified to ""How old were you when you were circumcised by the medical provider?" and asked only of those who indicated that their circumcision was carried out by a medical provider. The corresponding analytic variable (mcdetail_ug) has also been modified to take these changes into account.

Erroneous Child Module Questions

Due to an error in tablet programming, variables in the child module relating to child HIV testing were only asked of the subset of all mothers who gave birth to more than one child in the three years prior to the survey. As a result, the following variables were dropped from the UPHIA 2020-2021 individual interview dataset:

Original Variables	Analytic Variables
deathageyr1-deathageyr5	lastborntestedbirthdetail
deathagemo1-deathagemo5	lastbornhivstatus
childbrstfd1-childbrstfd5	breastfedlastchild
childbrstfdnow1-childbrstfdnow5	
chtsthivbirth1-chtsthivbirth5	
chtsthivagenum1-chtsthivagenum5	
chtsthivage1-chtsthivage5	
chtsthivagedk1-chtsthivagedk5	
chtsthivresult1-chtsthivresult5	
chtsthivbrstfd1-chtsthivbrstfd5	

chtsthivagelastnum1- chtsthivagelastnum5	
chtsthivagelast1-chtsthivagelast5	
chtsthivagelastdk1-chtsthivagelastdk5	
chtsthivresultlast1-chtsthivresultlast5	

HIV Testing Procedures

After the completion of standard laboratory testing at the central lab an additional test was undertaken in a small subset of cases with a specific pattern of discrepancies. In cases with inconclusive rapid test or Geenius results, or with a positive rapid test result and negative Geenius result, the Western blot test was used to determine the final HIV status as directed by ILB. The final HIV status on data files is consistent with earlier countries, and considers information from all tests performed, although process data on each test outcome is not provided to data users.

Incidence Estimation

The relatively high proportion of HIV infections which are of subtype D in Uganda affects HIV incidence estimation. Approximately 10% of HIV positive samples from participants 15+ years were selected for testing to determine the subtype distribution in the study population. This testing found that 23% of infections were subtype D, and 77% were subtype A. Since the mean duration of recent infection (MDRI), a key parameter for incidence estimation, varies by subtype, a survey-specific MDRI was calculated for UPHIA 2020-2021. This MDRI is a weighted average of the MDRIs for subtypes A and D:

$$\text{MDRI}_{\text{Uganda}} = W_A * \text{MDRI}_A + W_D * \text{MDRI}_D$$

where the *W*s and MDRIs are the proportions and MDRIs for each HIV subtype. For subtype A an MDRI of 130 days (95% CI 118-142 days) was used, consistent with previous PHIA surveys. For subtype D, an MDRI of 244 days (95% CI 166-326 days) was used, based on the mean of estimates from several sources, including CEPHIA, JHU, and CDC (unpublished data). The resulting weighted average MDRI for UPHIA 2020-2021 incidence estimation is 156 days (95% CI: 130-182 days). Note that this differs slightly from the UPHIA 2016-2017 average MDRI of 153 days (95% CI: 127-179 days), derived using a similar procedure, and from the 130 day MDRI (95% CI: 118-142 days) used in other PHIA countries.

Region coding

UPHIA 2020-2021 has different coding for the region variable than was used in UPHIA 2016. Regions in the PHIA data are combinations of Ugandan districts, and a different division of the country was used for the later PHIA survey. Care should be taken when comparing data from the two PHIA surveys at the regional level to ensure that this change in coding is taken into account. Note in particular that the "North East" region in UPHIA 2016 now corresponds to two separate regions in UPHIA 2021: "Northeast: Karamoja" and "Northeast: Teso".

Region Code	UPHIA 2020-2021 Region	UPHIA 2016 Region
1	Central 1: South Buganda	Central 1: South Buganda
2	Central 2: North Buganda	Central 2: North Buganda

3	East Central	East Central
4	Kampala	Kampala
5	Mid Eastern	Mid Eastern
6	Mid North	North East
7	Mid Western	West Nile
8	Northeast: Karamoja	Mid-North
9	Northeast: Teso	Mid-West
10	South Western	South West
11	West Nile	N/A

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 Uganda as an individual aged 15-17 who is pregnant, 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, who were either living in the selected households or were 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, who were either living in the selected households or were visitors 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 UPHIA 2020-2021 and other PHIA surveys. Users can refer to each PHIA survey's **Survey Questionnaires** and **Codebooks** provided as attachments to this document.

Table 2. UPHIA 2020-2021 questionnaire

Questionnaire module	Eligibility criteria
Household questionnaire	Sample of households within selected EAs
Household roster	
Household roster for minors	
Orphans and Vulnerable Children (OVC) Support	
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	
HIV 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 a usual household member or confirmed to have slept in the household the night before the interview.

4 Biomarker testing

In UPHIA 2020-2021, 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. UPHIA 2020-2021 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[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.

UPHIA 2020-2021 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 three HIV rapid tests, see the attached **HIV testing methodology diagram**. The Uganda HIV rapid testing algorithm applies three tests in sequence: Determine™, Stat-Pak™, and SD Bioline™. As per the serial testing algorithm attached, individuals with a non-reactive result on the screening test (Determine™) were reported as HIV-negative. Individuals with a reactive screening test underwent subsequent testing with Stat-Pak™. 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 guidelines. Individuals with a reactive Determine™ test followed by a non-reactive Stat-Pak™ test were tested using SD Bioline™. Individuals with a non-reactive SD Bioline™ test were classified as HIV negative, while those with a reactive SD Bioline™ test were considered inconclusive. For the purposes of the survey, samples with indeterminate results received further testing and evaluation to allow for final classification of HIV status.

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 including DNA qualitative polymerase chain reaction (PCR) (Roche COBAS AmpliPrep/COBAS Taqman (CAP/CTM) HIV-1 Qualitative Test). In cases with inconclusive rapid test or Geenius results, or with a positive rapid test result and negative Geenius result, the Western blot test was used to determine the final HIV status as directed by ILB.

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)	Variable
Household	dieddated_01- dieddated_05
Household	dieddatem_01- dieddatem_05
Adult individual	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)	Variable	Top-coding upper bound
Adult individual	age	80
Adult individual	wifliveew	4
Adult individual	arvamtm	6
Adult individual	livetimey	61
Adult individual	agemar	40
Adult individual	arvsmissdays	5
Adult individual	childa2012	6
Adult individual	firstsxage	30
Adult individual	husnwif	6
Adult individual	lifetimesex	35
Adult individual	liveb	12
Adult individual	mcage_ug	40
Adult individual	medinhmonths	6
Adult individual	monthtimes	7
Adult individual	numwif	4
Adult individual	part12monum	6
Adult individual	partage1	80
Household	ownchiknnum	30

Household	owncownum	13
Household	owndognum	5
Household	owngoatnum	20
Household	ownhorsenum	3
Household	roomsleep	8
Household	diedagey_01	80
Household	diedagey_02	80
Household	diedagey_03	80
Household	diedagey_04	80

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)	Variable	Bottom-coding lower bound
Adult individual	agemar	14
Adult individual	arvfty	2004
Adult individual	cervcntsy	2005
Adult individual	firstsxage	14
Adult individual	hivcly	2018
Adult individual	hivlastnegy	2000
Adult individual	hivtesty	2006
Adult individual	hivtfposy	2006
Adult individual	medinhmonths	1
Adult individual	monthwheny	2002
Adult individual	partage1	14
Adult individual	partage2	14
Adult individual	partage3	14
Adult individual	vltestlsty	2018
Household	diedagey_01	5
Household	diedagey_02	5
Household	diedagey_03	5
Household	diedagey_04	5
Household	diedagey_05	5
Roster	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)	Variable	Value(s)
Roster	livedistrict	Many values with small counts coded to 996
Household	cookingfuel	5, 8
Household	matexwalls	11, 23, 24, 25
Household	matfloor	31
Household	matroof	11
Household	watersource	71
Adult individual	partrelation1	6, 7
Adult individual	partrelation2	7
Adult individual	partrelation3	6
Adult individual	pregncr_ug	2, 3, 4, 5, 6, 7, 9, 10
Adult individual	religion_ug	8, 9
Adult individual	whereoutc	139, 140
Adult individual	whereoutd	Many values with small counts coded to 996
Adult individual	workind_ug	9
Adult individual	cmethod_b	1
Adult individual	arvloc	4, 5
Adult individual	arvswitchwhy_ug	3
Adult individual	hivtstlocation_ug	6
Adult individual	outregionwhrc	140
Adult individual	outregionwhrd_ug	2, 49, 62, 80, 100, 104, 106, 107, 108, 117, 121
Adult individual	cmethod_b	1
Adult individual	condomnoteasyrsn_g	1
Adult individual	hivtstnvrrsn_j	1
Adult individual	hivtstnvrrsn_m	1

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

Table 8. Variables that were redacted

Dataset(s)	Variable
Roster	livecountry
Roster	hhcemanc
Household	hhqown_d
Household	diedagedk_01-04
Household	diednamedk_01-02,04
Household	dieddatem_01-05
Household	dieddated_01-05
Household	diedagenum_01-05
Household	diedage_01-05
Adult individual	arvamdk

Adult individual   livetimenum
Adult individual   livetime
Adult individual   livetimedk
Adult individual   arvnrpg
Adult individual   birthday
Adult individual   birthmon
Adult individual   childalive3
Adult individual   childalive4
Adult individual   hivposprovm
Adult individual   arvsnottake
Adult individual   arvsnotcurrsn_ug
Adult individual   arvamtnum
Adult individual   arvamt
Adult individual   surveystday

Table 9. Variables with new categories

Dataset(s)	Variable	New Category
Adult individual	arvamtm	Changed to categorical: 0 – less than one month, 1 – 1 month or more
Adult individual	arvloc	Values 4 and 5 collapsed to 96 - Other
Adult individual	cerncrslt	Suspect Cancer collapsed to 2 – Abnormal/Positive/Suspect Cancer; Unclear/Inconclusive collapsed to 5 – Inconclusive/Not Received
Adult individual	cerncntrt	Values 1 and 2 collapsed to 1 – Received treatment

6 Dataset specifications

Table 10. UPHIA 2020-2021 dataset specifications

Dataset (filename)		Number of observations	Number of variables
Household	uphia2020hh	11,130	256
Roster	uphia2020roster	51,572	64
Adult individual	uphia2020adultind	26,071	476
Adult biomarker	uphia2020adultbio	25,479	205
Household intermediary weights	uphia2020hhintermediarywts	11,130	164
Individual intermediary weights	uphia2020indintermediarywts	51,572	647
Dataset specification		**Description**	
Two-letter country code prefix for ID variables		UG	
Survey weighting variables			
No. of jackknife replicates		159	
Selected variable parameters			
Household characteristics used for wealth index construction		*See next section*	
Mean duration recent infection used for HIV incidence estimation		156 days (95% CI 130-182 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 UPHIA 2020-2021.

Table 11. Household characteristics used for wealth index construction in UPHIA 2020-2021

Indicator variable	Type	Description
memsleep	Numeric (count)	Number of household members per sleeping room
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
econsup12_a	Binary	Household did not receive any form of external economic support in the last 12 months
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
hhqitems (option G)	Binary	Solar panels
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? [1]
ownchiknnum	Continuous	Chicken
owncownum	Continuous	Cows
owndognum	Continuous	Dogs
owngoatnum	Continuous	Goats
ownhorsenum	Continuous	Horses

[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 UPHIA 2020-2021 are consistent with those of other DHS studies in similar settings.[2-4]
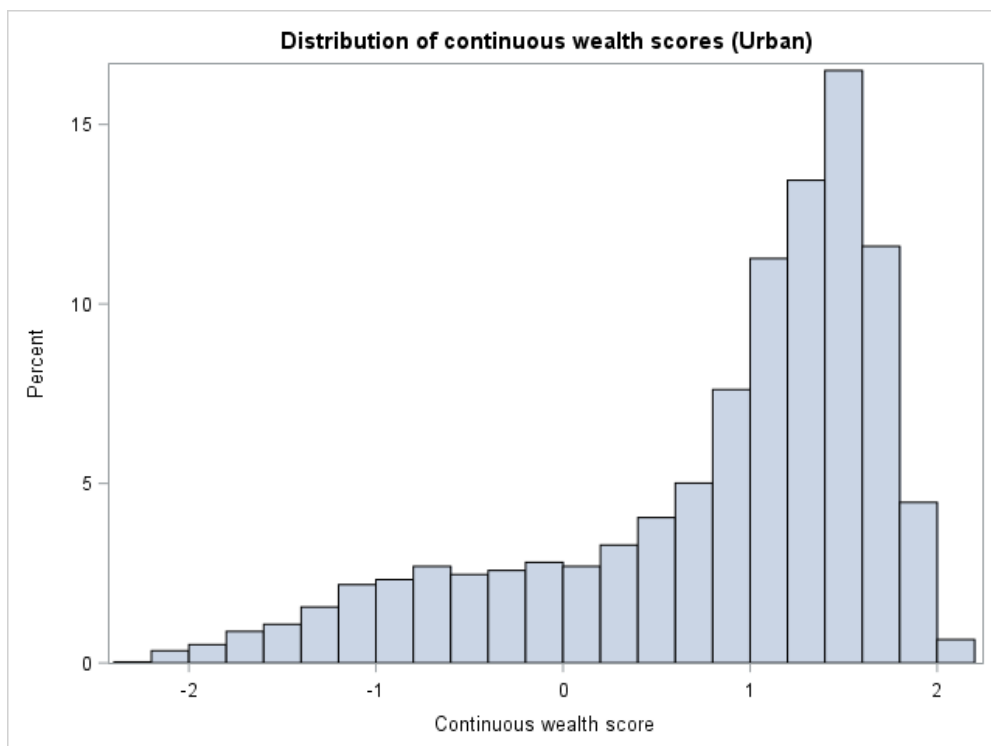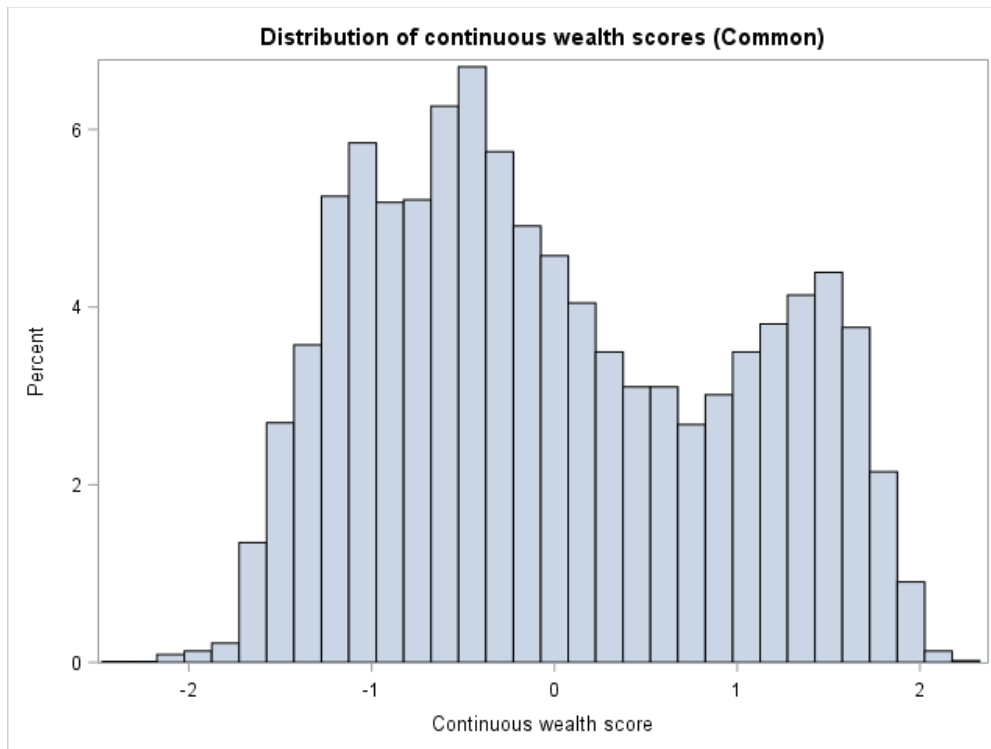
The PCA method does not guarantee the extraction of an index that is actually well-correlated with wealth but results from the PCA can be used to check whether the interpretation of the model makes sense. The component loading for each asset variable describes the association between that asset and the wealth index. The following table shows the most influential variables as measured by absolute value of their loading in each model:
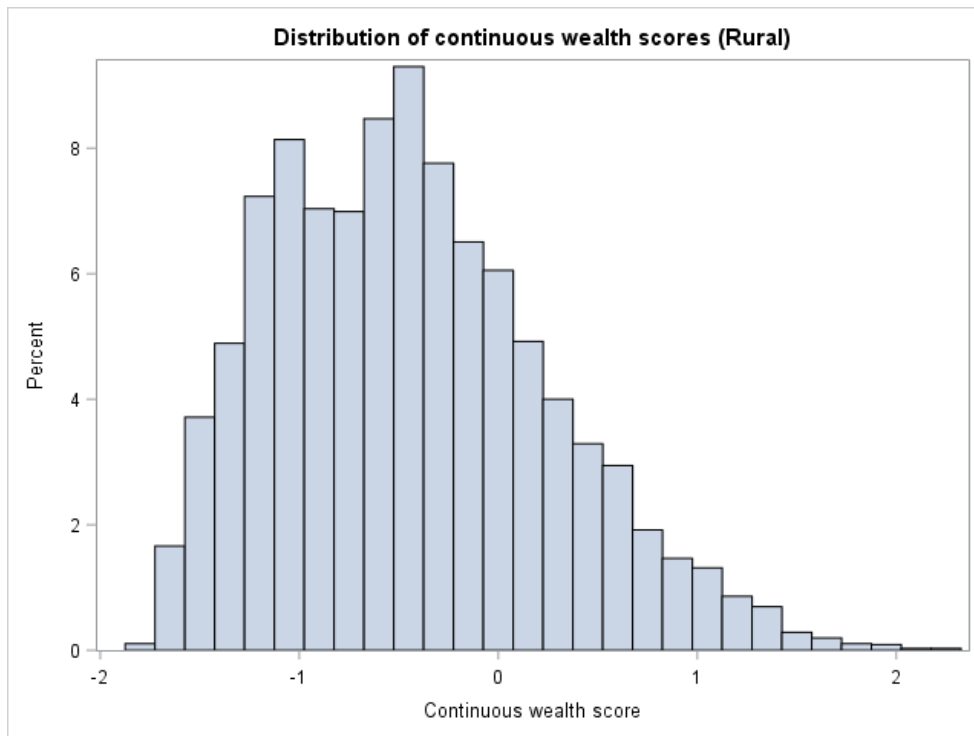
Table 12. PCA results for UPHIA 2020-2021 wealth quintile

		Component loading		
Variable	Category	Common model	Urban model	Rural model
Electricity in the house	yes	0.79083	-0.75566	0.55363
Cooking fuel	firewood/straw	-0.73918	0.71909	-0.51262
Dwelling flooring material	cement/terazo	0.68613	-0.57374	0.69910
Television	yes	0.68583	-0.61154	0.54149
Cooking fuel	charcoal from wood	0.66840	-0.58576	0.49900
Dwelling roofing material	thatch/palm leaf	-0.63953	0.70424	-0.69376
Dwelling wall material	cement	0.54866	-0.44126	0.46282

Electricity in the house and cooking 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 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 and rural-specific wealth indices. The distribution for the composite wealth index is skewed towards households with less wealth, with a smaller secondary peak towards the higher end of the score range.

Distribution of continuous wealth scores (Common)

Distribution of continuous wealth scores (Urban)

20

Distribution of continuous wealth scores (Rural)

8   References

1.   Ministry of Health, Uganda. Uganda Population-Based HIV Impact Assessment (UPHIA) 2020-2021: Summary Sheet. Kampala, Ministry of Health. August 2022.
2.   Ministry of Health (MOH), Uganda. Uganda Population-based HIV Impact Assessment 2020-2021 (UPHIA 2020-2021) Final Report. Kampala: MOH, Uganda; April 2024.
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.   National Statistical Office - NSO/Uganda and ICF. 2017. Uganda Demographic and Health Survey 2015-16. Zomba, Uganda: NSO and ICF.
7.   Ministry of Health, Uganda. Uganda Population-Based HIV Impact Assessment (UPHIA) 2015-2016: Final Report. Kampala, Ministry of Health. October 2018.

9 Attachments

9.1 Questionnaires

UPHIA 2020-2021 Data Manual Supplement Attachment 1 - Questionnaires.xlsx

9.2 Codebook with frequencies

UPHIA 2020-2021 Data Manual Supplement Attachment 2 - Codebook.docx

9.3 Flow Diagrams for selected analytic variables

UPHIA 2020-2021 Data Manual Supplement Attachment 3 - Flow Diagrams for Analytic Variables.pdf

9.4 HIV Testing Methodology Diagram

UPHIA 2020-2021 Data Manual Supplement Attachment 4 - Testing Methodology Diagram.pdf

9.5 Sample design and weighting report

UPHIA 2020-2021 Data Manual Supplement Attachment 5 - Sampling and Weighting Technical Report.docx

9.6 UPHIA 2020-2021 Survey-specific table specifications

UPHIA 2020-2021 Data Manual Supplement Attachment 6 - Additional Table Specifications.xlsx

9.7 Requesting data

UPHIA 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 UPHIA 2020, select "Uganda" 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.