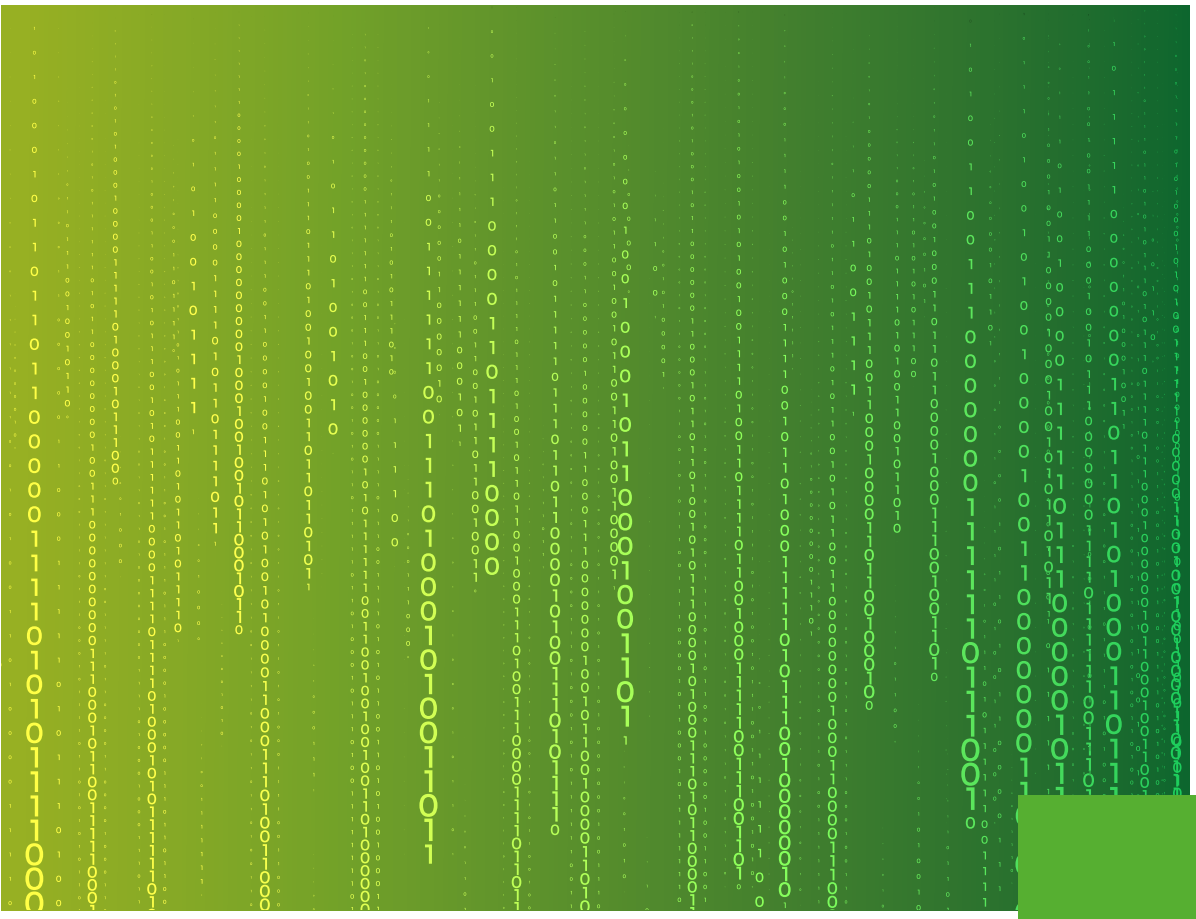
MPHIA 2020

Malawi
Population-based
HIV Impact Assessment



DATA USE MANUAL SUPPLEMENT

Malawi Population-based HIV Impact Assessment 2020-2021

MPHIA 2020-2021

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.

MPHIA 2020-2021 Collaborating Institutions

Malawi Ministry of Health (MOH)
Malawi National AIDS Commission (NAC)
The United States (US) President's Emergency Plan for AIDS Relief (PEPFAR)
Centers for Disease Control and Prevention (CDC), Malawi
Centers for Disease Control and Prevention (CDC), Atlanta
ICAP at Columbia University, New York
ICAP at Columbia University, Malawi
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

Malawi Population-based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021) Data Use Manual Supplement. New York, NY. November 2023.

Single dataset

Ministry of Health (MOH), Malawi. Malawi Population-Based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021) [Dataset]. <Insert dataset name used, e.g. mphia2020adultbio>. Lilongwe: MOH, National AIDS Commission, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2023.

Multiple datasets (more than 3)

Ministry of Health (MOH), Malawi. Malawi Population-Based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021) (various) [Datasets]. Lilongwe: MOH, National AIDS Commission, 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=3

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
MPHIA	Malawi 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 MPHIA 2020-2021?

The Malawi Population-based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021) is a cross-sectional household-based survey conducted in Malawi. MPHIA 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.

1.2 Purpose of the MPHIA 2020-2021 Data Manual Supplement

The purpose of the ***MPHIA 2020-2021 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 MPHIA 2020-2021 survey specifications, including survey-specific eligibility criteria, sampling approaches and measures, and survey-specific documentation such as codebooks and questionnaires. A summary of MPHIA 2020-2021 findings can be found in the ***MPHIA 2020-2021 Summary Sheet***, and a full report on findings is available in the ***MPHIA 2021-2022 Final Report***.

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 ***MPHIA 2020-2021 Summary Sheet*** and ***MPHIA 2020-2021 Final Report***. The ***MPHIA 2020-2021 Final Report*** contains detailed results from MPHIA 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.

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

For Stata, values have been labelled within each dataset. For SAS, there is a second statistical program containing code to label all values for variables on each of the data sets.

  - ***MPHIA 2020-2021 Formats.sas***: SAS program

2 Survey design and data collection

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

Table 1. MPHIA 2020-2021 survey design characteristics

Survey design characteristics	Description
Survey design	
Data source for survey weighting[1]	2018 Malawi Census
Sampling stratum	Zone
Primary sampling unit	National Statistics Office (NSO) Enumeration Areas (EAs)
Urban/rural categorization	Urban/rural
Survey administration	
Data collection dates	January 15, 2020 – April 26, 2021 Survey paused: March 31, 2020 – March 7, 2021
Languages	English, Chichewa, Tumbuka
Sample size[2]	
Number of selected EAs	438
Number of selected households	15,330
Number of rostered individuals (all ages)	59,021
Survey participation	
Number of completed household interviews	12,815
Number of completed individual interviews	26,519
Number of completed biomarker tests	22,662

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

Exceptions to the general PHIA design

The MPHIA 2020-2021 survey was paused due to the COVID-19 pandemic from late March 2020 until early March 2021. Prior to resuming data collection in 2021, questions were added for COVID-19-specific experiences around economic support, location and interruption of ARV access, and accessing general health care services during the pandemic. Additionally, response categories for COVID-19 were added to questions about reasons away from home, barriers to HIV testing, 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 in March 2021, which is a non-random subsample of the Malawi population.

Questionnaire Changes

There were several country-specific changes to the questionnaire in MPHIA 2020. Non-standard questions added to the questionnaire, and 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 "_mw" suffix. The full list of questions having country-specific changes is as follows:

Question	**Variable Name**
Language of questionnaire	lngvqx_lng_mw
Language of interview	lngvint_lng_mw
Language of respondent	lngnat_lng_mw
Highest class/form/year of education completed	schcom_mw
Before you moved here, which district/country did you live in?	outregionwhr_mw
In the last 12 months, have you ever lived away from home?	awayever_mw
How many times did you live away from home?	awaynum_mw
How long were you away? (months)	awaytimem_mw
The last time you were away from home for more than a month, where were you?	whereout_mw
What was the main reason you went away from home for more than a month?	reasonaway_mw
What method of cervical cancer testing did you receive?	cervcnmethod_mw
Has the corona (COVID-19) pandemic compromised your ability to access health care services?	servaccesscovid_mw
Which health care services were difficult to access due to corona (COVID-19)	servaccesstypecovid_[c,d,x]_mw
Where can you get condoms?	condomwhere_[a-h,x]_mw
Have you taken part in any of the following prevention or treatment programs?	adhivprev_[a-j,w,x]_mw
Which district or country is [NAME] in currently?	liveregionlivecountry_mw

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 Malawi as an individual aged 15-17 who is married or free from any legally competent representative as defined by law in Malawi.) 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 or visiting the selected households, who slept in the household the night before the survey, and who were willing and able to provide verbal consent
- Adolescents aged 15-17 years, living in or visiting the selected households, 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 MPHIA 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. MPHIA 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 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 MPHIA 2020-2021, 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 3. MPHIA 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.

MPHIA 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 two HIV rapid tests, see the attached ***HIV testing methodology diagram***. The Malawi HIV rapid testing algorithm applies two tests in sequence: Determine™ and Unigold™. 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 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 after 4 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.

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).

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
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
Roster	age	80
Adult individual	age	80
Adult biomarker	age	80
Household	ownchiknnum	20
Household	owncownum	7
Household	owndognum	5
Household	owngoatnum	10
Household	roomsleep	6
Household	diedagey_01	80
Household	diedagey_02	80
Household	diedagey_03	80
Adult individual	agemar	40
Adult individual	arvsmissdays	5
Adult individual	awaynum	10
Adult individual	childa2012	6
Adult individual	firstsxage	27
Adult individual	husnwif	4
Adult individual	lifetimesex	20
Adult individual	liveb	11
Adult individual	mcage	40
Adult individual	medinhmonths	24
Adult individual	monthtimes	6
Adult individual	numwif	3
Adult individual	part12monum	6

Adult individual	partage1	80
Adult individual	partage2	80
Adult individual	wifliveew	3
Adult individual	arvamtm	6
Adult individual	chtsthivagelastm1	24
Adult individual	chtsthivagem1	24
Adult individual	livetimey	61

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
Household	diedagey_01	5
Household	diedagey_02	5
Household	diedagey_03	5
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	2005
Adult individual	hivtfposy	2005
Adult individual	medinhmonths	2
Adult individual	monthwheny	1994
Adult individual	partage1	14
Adult individual	partage2	14
Adult individual	partage3	14
Adult individual	vltestlsty	2016
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)
Adult individual	hivtstlocation	6
Adult individual	outregionwhr_mw	12, 36
Adult individual	partrelation1	7
Adult individual	partrelation2	7
Adult individual	partrelation3	7
Adult individual	reasonaway_mw	3,4
Adult individual	whereout_mw	14, 30, 34, 38
Adult individual	workind	9
Adult individual	adhivprev_g_mw	1
Adult individual	adhivprev_k_mw	1
Adult individual	adhivprev_l_mw	1
Adult individual	chronicmed_d	1
Adult individual	cmethod_b	1
Adult individual	cmethod_h	1
Adult individual	condomnoteasyrsn_g	1
Adult individual	hivtstnvrrsn_j	1
Adult individual	hivtstnvrrsn_m	1
Adult individual	servaccesstypecovid_a_mw	1
Adult individual	servaccesstypecovid_b_mw	1
Adult individual	arvloc	3, 4, 5
Adult individual	arvswitchwhy	2, 3, 4
Adult individual	cerncntrt	1, 2, 3
Roster	liveregionlivecountry_mw	1, 10, 12, 13, 14, 15, 17, 18, 19, 2, 21
Roster	relattohh	9
Household	cookingfuel	2, 3, 4, 8, 95
Household	matfloor	12, 21, 22, 31, 32, 33, 35
Household	matroof	11, 13, 22, 31, 32, 33
Household	watersource	41, 42, 51, 91

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

Table 8. Variables that were redacted

Dataset(s)	Variable
Household	dieddatem_02-05
Household	hhqown_d
Household	ownhorsenum
Adult individual	arvnrpg
Adult individual	arvsnotcurrsn
Adult individual	arvsnottake
Adult individual	childalive2-5
Adult individual	childbrstfd2-5
Adult individual	childbrstfdnow2-5
Adult individual	chtsthivagem2
Adult individual	chtsthivagelastm2-4
Adult individual	chtsthivbirth2-5
Adult individual	chtsthivbrstfd2-5

Adult individual	chtsthivresultlast2-4
Adult individual	deathagemo1-3
Adult individual	deathageyr1-3

Table 9. Variables with new categories

Dataset(s)	Filename	Variable	Values	New Category
Adult individual	mphia2020adultind	cerncnrslt	2,3	2 – Abnormal/Positive/Suspect Cancer
			4,5	5 – Inconclusive/Not received
Adult individual	mphia2020adultind	arvamtm	0,1	1 – Zero or one month

6 Dataset specifications

Table 10. MPHIA 2020-2021 dataset specifications

Dataset (filename)		Number of observations	Number of variables
Household	mphia2020hh	15,330	291
Roster	mphia2020roster	59,021	58
Adult individual	mphia2020adultind	26,519	535
Adult biomarker	mphia2020adultbio	22,662	263
Drug resistance	mphia2020dr	*Forthcoming*	
Household intermediary weights	mphia2020hhintermediarywts	15,330	222
Individual intermediary weights	mphia2020indintermediarywts	59,021	879
Dataset specification		**Description**	
Two-letter country code prefix for ID variables		MW	
Survey weighting variables			
No. of jackknife replicates		217	
Survey weights provided (variable prefix)		Household (hhwt) Individual interview (intwt) Blood test (btwt) Drug Resistance (dr) *(forthcoming)*	
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, standard error 37.48575911)	

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 MPHIA 2020-2021.

Table 11. Household characteristics used for wealth index construction in MPHIA 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	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	None of the above
		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	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 MPHIA 2020-2021 are consistent with those of other DHS studies in similar settings.[2-4]

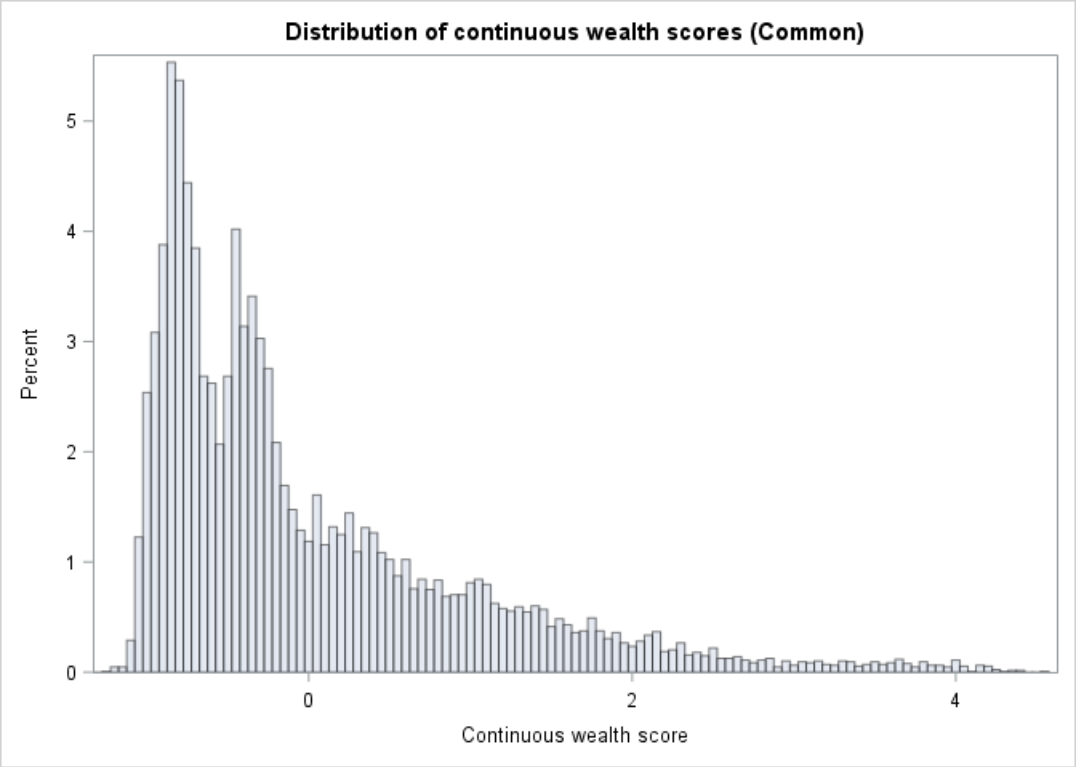
The PCA method does not guarantee the extraction of an index that is actually well-correlated with wealth but results from the PCA can be used to check whether the interpretation of the model makes sense. The component loading for each asset variable describes the association between that asset and the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

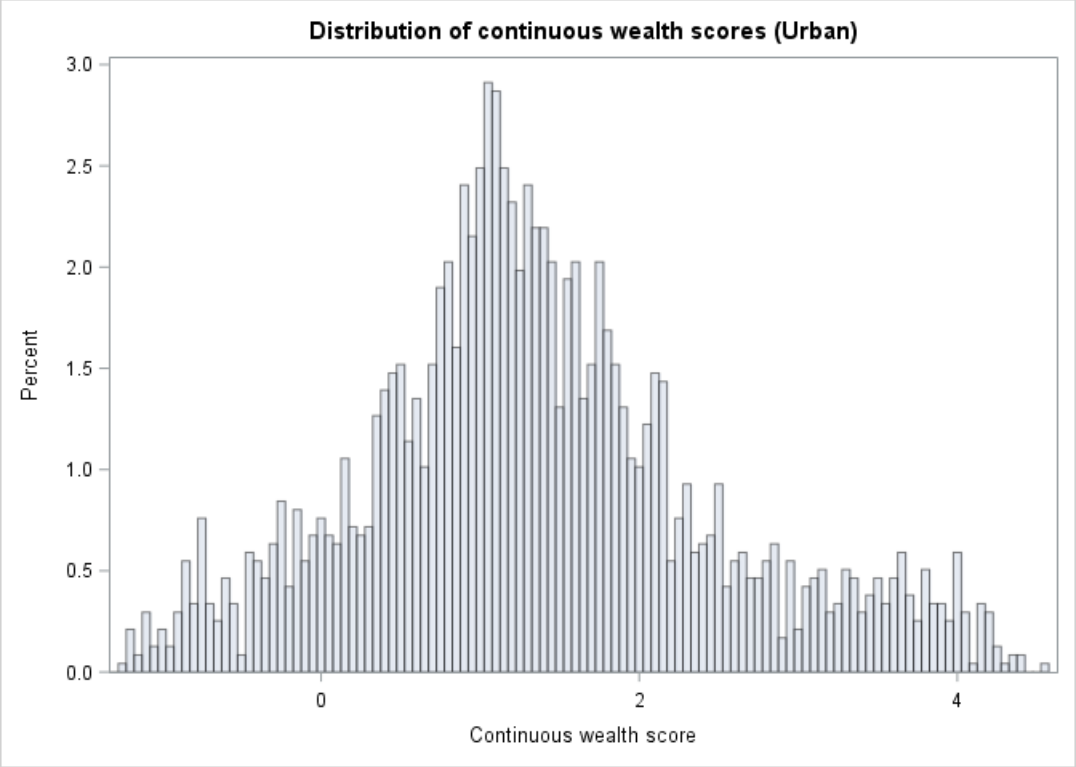
Table 12. PCA results for MPHIA 2020-2021 wealth quintile

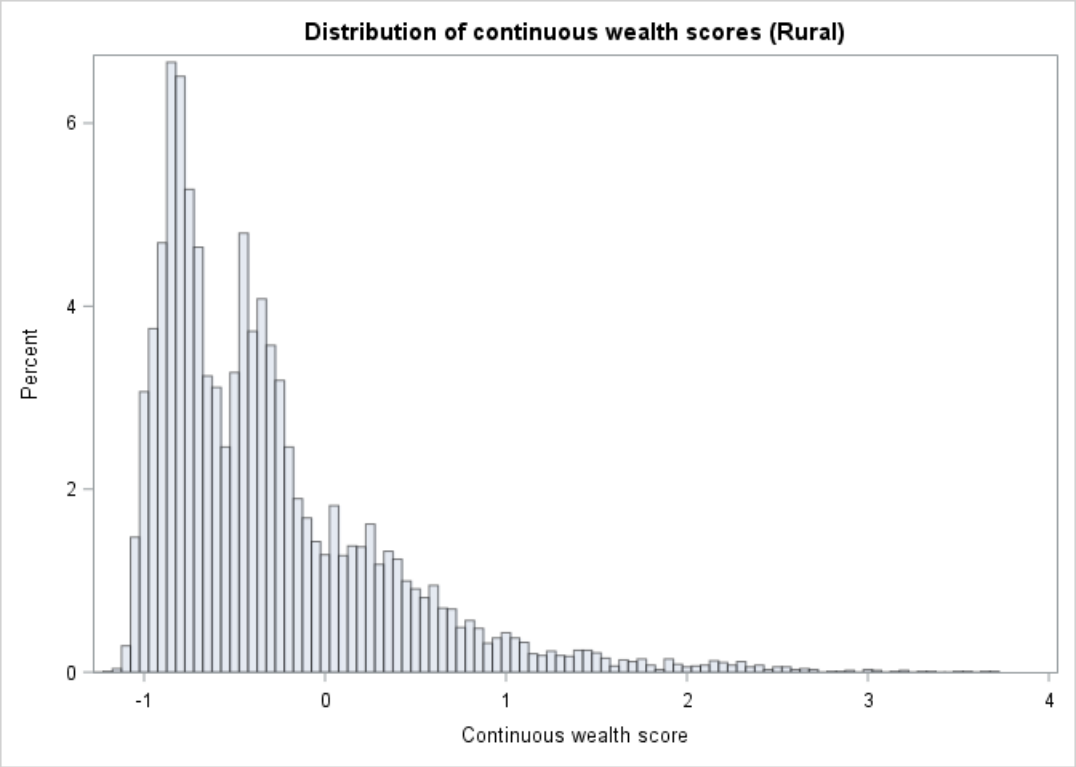
		Component loading		
Variable	**Category**	**Common model**	**Urban model**	**Rural model**
Dwelling flooring material	earth/sand	-0.77645	-0.62778	-0.76127
Electricity in the house	yes	0.75889	0.713	0.60472
Dwelling flooring material	cement/terazo	0.73942	0.44515	0.77285
Cooking fuel	firewood/straw	-0.72483	-0.46941	-0.51653
Television	yes	0.68985	0.69385	0.52382
Dwelling roofing material	thatch/palm leaf	-0.62075	-0.47701	-0.6786
Refrigerator	yes	0.60539	0.67518	0.41596

Dwelling flooring material and electricity in the house 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 higher wealth, with a smaller secondary peak towards the lower end of the score range.


Distribution of continuous wealth scores (Common)
Percent
0
1
2
3
4
5
0
2
4
Continuous wealth score


Distribution of continuous wealth scores (Urban)
Percent
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
2
4
Continuous wealth score


Distribution of continuous wealth scores (Rural)
Percent
0
2
4
6
-1
0
1
2
3
4
Continuous wealth score

8 References

1. Ministry of Health (MOH), Malawi. Malawi Population-based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021): Summary Sheet. Lilongwe: MOH, Malawi; March, 2022. Available from https://phia.icap.columbia.edu/malawi-summary-sheet-2/
2. Ministry of Health (MOH), Malawi. Malawi Population-based HIV Impact Assessment 2020-2021 (MPHIA 2020-2021): Final Report. Lilongwe: MOH, Malawi; November, 2022. Available from https://phia.icap.columbia.edu/malawi-final-report-2020-2021/
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/Malawi and ICF. 2017. Malawi Demographic and Health Survey 2015-16. Zomba, Malawi: NSO and ICF.
7. Ministry of Health, Malawi. Malawi Population-Based HIV Impact Assessment (MPHIA) 2015-2016: Final Report. Lilongwe, Ministry of Health. October 2018.

9 Attachments

9.1 Questionnaires

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

9.2 Codebook with frequencies

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

9.3 Flow Diagrams for selected analytic variables

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

9.4 HIV Testing Methodology Diagram

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

9.5 Sample design and weighting report

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

9.6 Requesting data

MPHIA 2020-2021 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 MPHIA 2020-2021, select "Malawi" 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.

9.7 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.