

# Zimbabwe Population-based HIV Impact Assessment 2020

## ZIMPHIA 2020

### DATA USE MANUAL SUPPLEMENT



The mark "CDC" is owned by the US Dept. of Health and Human Services and is used with permission. Use of this logo is not an endorsement by HHS or CDC of any particular product, service or enterprise.

# Zimbabwe Population-based HIV Impact Assessment 2020

## ZIMPHIA 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.



## ZIMPHIA 2020 Collaborating Institutions

Ministry of Health and Child Care (MoHCC), Zimbabwe  
Centers for Disease Control and Prevention (CDC), Zimbabwe  
Centers for Disease Control and Prevention (CDC), Atlanta  
ICAP at Columbia University, New York  
ICAP at Columbia University, Zimbabwe  
Westat  
Lancet Clinical Laboratories Zimbabwe  
Zimbabwe National Statistics Agency (ZIMSTAT)  
National AIDS Council (NAC)

## 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*

Zimbabwe Population-based HIV Impact Assessment (ZIMPHIA) 2020 Data Use Manual Supplement. New York, NY. August 2023.

### *Single dataset*

Ministry of Health and Child Care (MoHCC). Zimbabwe Population-Based HIV Impact Assessment 2020 (ZIMPHIA 2020) [Dataset]. <Insert dataset name used, e.g. zimpbia2020adultbio>. Harare, Zimbabwe: MoHCC, National Microbiology Reference Laboratory, Zimbabwe National Statistics Agency, National AIDS Council, PEPFAR, CDC, Westat, ICAP at Columbia University [Producers]. ICAP at Columbia University [Distributor], 2023.

### *Multiple datasets (more than 3)*

Ministry of Health and Child Care (MoHCC). Zimbabwe Population-Based HIV Impact Assessment 2020 (ZIMPHIA 2020) (various) [Datasets]. Harare, Zimbabwe: MoHCC, National Microbiology Reference Laboratory, Zimbabwe National Statistics Agency, National AIDS Council, 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=6](https://phia-data.icap.columbia.edu/datasets?country_id=6)

## Contact Information

ICAP at Columbia University  
722 West 168<sup>th</sup> Street  
New York, NY 10032  
Website: [icap.columbia.edu](https://icap.columbia.edu)  
Email: [icap-communications@columbia.edu](mailto:icap-communications@columbia.edu)

## **Table of Contents**

<b>1</b>	<b>Background .....</b>	<b>4</b>
1.1	What is ZIMPHIA 2020? .....	4
1.2	Purpose of the ZIMPHIA 2020 Data Use Manual Supplement .....	4
1.3	Other documentation and resources .....	4
<b>2</b>	<b>Survey design and data collection .....</b>	<b>6</b>
<b>3</b>	<b>Overview of survey questionnaires.....</b>	<b>8</b>
<b>4</b>	<b>Biomarker testing .....</b>	<b>9</b>
<b>5</b>	<b>Data confidentiality.....</b>	<b>10</b>
<b>6</b>	<b>Dataset specifications .....</b>	<b>14</b>
<b>7</b>	<b>Wealth index.....</b>	<b>15</b>
<b>8</b>	<b>References .....</b>	<b>19</b>
<b>9</b>	<b>Attachments .....</b>	<b>20</b>
9.1	Questionnaires .....	20
9.2	Codebook with frequencies .....	20
9.3	Flow Diagrams for selected analytic variables .....	20
9.4	HIV Testing Methodology Diagram .....	21
9.5	Sampling and Weighting Technical Report .....	21
9.6	ZIMPHIA 2020 Survey-specific table specifications .....	22
9.7	Requesting data .....	22
9.8	Data explorer .....	23

## Abbreviations

ART	Antiretroviral Therapy
ARV	Antiretroviral
CAP/CTM	COBAS AmpliPrep/COBAS Taqman HIV-1 Qualitative Test
CD4	CD4+ T-Cell
CI	Confidence Interval
DHS	Demographic and Health Surveys
DNA	Deoxyribonucleic Acid
EA	Enumeration Area
HIV	Human Immunodeficiency Virus
ID	Identification
LAgi-EIA	Limiting-Antigen Avidity Enzyme Immunoassay
ZIMPHIA	Zimbabwe 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 ZIMPHIA 2020?

The Zimbabwe Population-based HIV Impact Assessment 2020 (ZIMPHIA 2020) was a cross-sectional household-based survey conducted in Zimbabwe. ZIMPHIA 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.

Please note that ZIMPHIA 2020 was the second PHIA survey in Zimbabwe. The previous PHIA in Zimbabwe is referred to as ZIMPHIA 2015-2016.

## 1.2 Purpose of the ZIMPHIA 2020 Data Use Manual Supplement

The **ZIMPHIA 2020 Data Manual Supplement** (hereafter, **Supplement**) contains ZIMPHIA 2020 survey specifications, including survey-specific eligibility criteria, sampling approaches and measures, and survey-specific technical documentation such as codebooks and questionnaires. It is intended to accompany the **PHIA Data Manual** (hereafter, **Manual**), which contains information on PHIA data generally applicable to all PHIA surveys, including general information on the data and documentation packages and their contents, a guide to getting started with the PHIA data, and details on the files and variables available for all PHIA. Users should refer to both the **Manual** and this **Supplement** for a complete discussion of the ZIMPHIA 2020 data.

## 1.3 Other documentation and resources

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

Important ZIMPHIA 2020-specific documentation attached to this **Supplement** includes:

- **ZIMPHIA 2020 Survey Questionnaires:** The ZIMPHIA 2020 household, roster, and adult questionnaires. These questionnaires illustrate the questionnaire's structure, including the order that the questions were asked, each question's wording, variable names and labels, value coding and labels, and skip patterns. The question number on the questionnaire is referenced in the variable label on the datasets and in the "variable label" of the codebook, where applicable.
- **ZIMPHIA 2020 Codebook with Frequencies:** Codebooks are provided for each dataset and document each variable's name, category (i.e., the questionnaire module or source data of the variable), full question text or variable description, variable label (i.e., a condensed label used on the datasets), type and width (e.g., numeric, text), coding values and labels, and the frequency and percent of records containing each value. Summary statistics are provided in the coding values and labels for selected numeric variables, such as counts.

- **ZIMPHIA 2020 Analytic Variable Flow Diagrams:** Flow diagrams illustrate the logic used to create key analytic variables.
- **ZIMPHIA 2020 Testing Methodology Diagram:** Flow diagram illustrating household-based HIV testing algorithm.
- **ZIMPHIA 2020 Sampling and Weighting Technical Report:** Details of ZIMPHIA 2020 sampling and weighting procedures.
- **ZIMPHIA 2020 Survey-Specific Table Specifications:** Table shells and technical specifications for ZIMPHIA 2020-specific final report tabulations. (Shells and specifications for all standard final report tabulations are included in the **Manual**.)

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

- **ZIMPHIA 2020 STATA Intro Code.DO:** Stata do-file
- **ZIMPHIA 2020 SAS Intro Code.SAS:** SAS program
- **ZIMPHIA 2020 R Intro Code.R:** R script

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

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

## 2 Survey design and data collection

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

**Table 1. ZIMPHIA 2020 survey design characteristics**

Survey design characteristics	Description
Survey design	
Data source for survey weighting <sup>1</sup>	2012 Zimbabwe Census
Sampling stratum	Province
Primary sampling unit	Census Enumeration Areas (EA)
Urban/rural categorization	Urban/rural
Survey administration	
Data collection dates	November 2019 – March 2020
Languages	English, Shona, Ndebele
Sample size <sup>1</sup>	
Number of selected EAs	356
Number of selected households	12,460
Number of rostered individuals	43,403
Survey participation	
Number of completed household interviews	10,499
Number of completed individual interviews	20,793
Number of completed biomarker tests	19,535

<sup>1</sup> See **Attachment 9.5: Sampling and Weighting Technical Report** for more details on the survey weighting approach and for response rates.

### Exceptions to the general PHIA design

One EA in ZIMPHIA 2020 was inadvertently missed during data collection. For completeness of the dataset, all households from the missed EA are included in the household dataset. These households are assigned “unknown eligibility” (hhstatus=4). to the **ZIMPHIA 2020 Sampling and Weighting Technical Report** for more details.

The general PHIA design designates the first individual on the roster (line number 1) as the household head. In ZIMPHIA 2020, there was one household in which the interviewer deleted the first individual on the roster (householdid = ZW200000006977). The second individual on the roster (line number 2) proceeded to provide consent for the household to participate and answered the household interview. For this one household, the individual with line number 2 is considered the household head for analysis, and their information is used in analytic variables pertaining to the household head.

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 ZIMPHIA 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 a “\_zw” suffix. The full list of questions having country-specific changes is as follows:

<b>Question</b>	<b>Variable Name</b>
Language of interview	lngvint_lng_zw
Language of interview (other options)	lngvint_lngoht_zw
Native language of participant	lngnat_lngoht_zw
What is the highest level of school you have completed?	schcom_zw
Before you moved here, which province did you live in? If you lived outside of Zimbabwe, which country did you live in?	outregionwhr_zw
The last time you were away from home for more than a month, where were you?	whereout_zw
Why were your ARVs changed?	arvswitchwhy_zw
How do you get your ARVs when you are away from home for more than a month?	arvinterraway_zw
Have you taken part in any of the following HIV prevention programs?	adhivprev_[option]_zw
In the past 12 months, how many times have you participated in a school meeting or class period where they talked about HIV/AIDS? If you are not certain, give your best guess.	adhivschmtg_zw
Which province or country is [household member name] in currently?	liveregionlivecountry_zw
Has your household received any of the following forms of external economic support in the last 12 months?	econsup12_[option]_zw

### 3 Overview of survey questionnaires

---

In participating households, a household questionnaire was administered to the designated household head. Household head is defined as an individual age 18 or older and emancipated minors (defined in Zimbabwe as an individual ages 15-17 who is married or pregnant, has a child, lives alone, is head of one or more households, or is free from any legally competent representative as defined by law in Zimbabwe) The household head provided consent for the household to participate in the survey, after which individual members were rostered during the household interview.

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

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

Modules included in each questionnaire and their associated eligibility criteria are listed in the table below. The content and order of each module may differ between ZIMPHIA 2020 and other PHIA surveys. Users should refer to the **Supplement** for any other surveys they intend to analyze to familiarize themselves with surveys differences.

**Table 2. ZIMPHIA 2020 questionnaire**

Questionnaire module	Eligibility criteria
<i>Household questionnaire</i>	Sample of households within selected EAs
Household roster	
Household roster for minors	
Deaths in the household	
Household characteristics	
Economic support	
<i>Individual questionnaire – adults (15 years and older)</i>	All eligible <sup>1</sup> 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

<sup>1</sup> Household members are eligible if they were confirmed to have slept in the household the night before the interview.

## 4 Biomarker testing

In ZIMPHIA 2020, 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. ZIMPHIA 2020 biomarker testing**

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

<sup>1</sup> See HIV testing algorithm below.

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

### ZIMPHIA 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 three HIV rapid tests, see **Attachment 9.4: HIV testing methodology diagram**. Individuals with a nonreactive result on an initial HIV rapid screening test (Alere Determine™ HIV-1/2 Ag/Ab) were classified as HIV-negative. Then, individuals with a reactive screening test underwent confirmatory testing using the First Response HIV 1-2.0 rapid HIV test (StatPak). Individuals with a reactive screening test result followed by a non-reactive confirmatory test results were retested in parallel in the field. Individuals with both tests non-reactive were classified as HIV-negative. Those with both tests reactive were classified as HIV-positive. When both tests were discordant, the INSTI® HIV-1/HIV-2 Antibody Test (INSTI) was introduced. Individuals with a non-reactive INSTI test were classified as HIV-negative. Individuals with a reactive INSTI test were considered to have inconclusive results and referred to a health facility for retesting in 14 days as per the national guidelines.

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

## 5 Data confidentiality

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

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

**Table 4. Date variables redacted for all PHIA surveys**

Dataset(s)	Filename	Variable
Household	zimphia2020hh	dieddated_01- dieddated_05
Household	zimphia2020hh	dieddate_01-dieddate_05
Adult individual	zimphia2020adultind	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	zimphia2020roster	age	80
Adult individual	zimphia2020adultind	age	80
Adult biomarker	zimphia2020adultbio	age	80
Household	zimphia2020hh	ownchiknum	20
Household	zimphia2020hh	owncownum	15
Household	zimphia2020hh	owndognum	6
Household	zimphia2020hh	owngoatnum	16
Household	zimphia2020hh	ownhorsenum	6
Household	zimphia2020hh	roomsleep	6
Household	zimphia2020hh	diedagey_01	80
Household	zimphia2020hh	diedagey_02	80
Household	zimphia2020hh	diedagey_04	80
Adult individual	zimphia2020adultind	agemar	40
Adult individual	zimphia2020adultind	arvsmisdays	4
Adult individual	zimphia2020adultind	childa2012	6
Adult individual	zimphia2020adultind	chtsthivagelastm1	36
Adult individual	zimphia2020adultind	chtsthivagem1	36
Adult individual	zimphia2020adultind	firstsxage	37
Adult individual	zimphia2020adultind	husnwif	4
Adult individual	zimphia2020adultind	lifesex	35
Adult individual	zimphia2020adultind	liveb	10
Adult individual	zimphia2020adultind	livetimey	62
Adult individual	zimphia2020adultind	mcage	40

<b>Dataset(s)</b>	<b>Filename</b>	<b>Variable</b>	<b>Top-coding upper bound</b>
Adult individual	zimphia2020adultind	medinhmonths	5
Adult individual	zimphia2020adultind	monthtimes	7
Adult individual	zimphia2020adultind	numwif	3
Adult individual	zimphia2020adultind	part12monum	6
Adult individual	zimphia2020adultind	partage1	80
Adult individual	zimphia2020adultind	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**

<b>Dataset(s)</b>	<b>Filename</b>	<b>Variable</b>	<b>Bottom-coding lower bound</b>
Household	zimphia2020hh	diedagey_01	5
Household	zimphia2020hh	diedagey_02	5
Adult individual	zimphia2020adultind	agemar	14
Adult individual	zimphia2020adultind	arvfty	2004
Adult individual	zimphia2020adultind	cervcntsy	2010
Adult individual	zimphia2020adultind	chtsthivagem1	2
Adult individual	zimphia2020adultind	firstsxage	14
Adult individual	zimphia2020adultind	hivcly	2017
Adult individual	zimphia2020adultind	hivlastnegy	2004
Adult individual	zimphia2020adultind	hivtesty	2005
Adult individual	zimphia2020adultind	hivtfposy	2005
Adult individual	zimphia2020adultind	medinhmonths	1
Adult individual	zimphia2020adultind	monthwheny	2002
Adult individual	zimphia2020adultind	partage1	14
Adult individual	zimphia2020adultind	partage2	14
Adult individual	zimphia2020adultind	partage3	14
Adult individual	zimphia2020adultind	vltestlsty	2014
Roster	zimphia2020roster	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)
Household	zimphia2020hh	cookingfuel	5,8
Household	zimphia2020hh	matexwalls	11,23,24,25
Household	zimphia2020hh	matfloor	22,32
Household	zimphia2020hh	matroof	11
Household	zimphia2020hh	watersource	61,71, 91
Household	zimphia2020hh	econsup12_j_zw	1
Adult individual	zimphia2020adultind	arvsinterraway_zw	4
Adult individual	zimphia2020adultind	arvswitchwhy_zw	52
Adult individual	zimphia2020adultind	chronicmed_d	1
Adult individual	zimphia2020adultind	chronicmed_e	1
Adult individual	zimphia2020adultind	cmethod_b	1
Adult individual	zimphia2020adultind	cmethod_h	1
Adult individual	zimphia2020adultind	cmethod_i	1
Adult individual	zimphia2020adultind	hivstlocation	6,7
Adult individual	zimphia2020adultind	hivstnvrnsn_b	1
Adult individual	zimphia2020adultind	hivstnvrnsn_h	1
Adult individual	zimphia2020adultind	hivstnvrnsn_j	1
Adult individual	zimphia2020adultind	hivstnvrnsn_k	1
Adult individual	zimphia2020adultind	lngnat_lngoth_zw	1,10,12,14,2,6,7,9
Adult individual	zimphia2020adultind	outregionwhr_zw	13
Adult individual	zimphia2020adultind	whereout_zw	15
Adult individual	zimphia2020adultind	workind	9
Roster	zimphia2020roster	liveregionlivecountry_zw	11,13

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

**Table 8. Variables that were redacted**

Dataset(s)	Filename	Variable
Adult individual	zimphia2020adultind	arvnprg
Adult individual	zimphia2020adultind	arvsnotcurrsn
Adult individual	zimphia2020adultind	arvsnottake
Adult individual	zimphia2020adultind	arvtpkg
Adult individual	zimphia2020adultind	childalive2-childalive5
Adult individual	zimphia2020adultind	childbrstfd2-childbrstfd5
Adult individual	zimphia2020adultind	childbrstfdnow2-childbrstfdnow5
Adult individual	zimphia2020adultind	chtsthivagelastm2
Adult individual	zimphia2020adultind	chtsthivagem2
Adult individual	zimphia2020adultind	chtsthivbirth2-chtsthivbirth5
Adult individual	zimphia2020adultind	chtsthivbrstfd2
Adult individual	zimphia2020adultind	chtsthivresult2, chtsthivresult5
Adult individual	zimphia2020adultind	chtsthivresultlast2
Adult individual	zimphia2020adultind	deathagemo1-deathagemo3

<b>Dataset(s)</b>	<b>Filename</b>	<b>Variable</b>
Adult individual	zimphia2020adultind	deathageyr1-deathageyr3
Adult individual	zimphia2020adultind	firstsxagec
Adult individual	zimphia2020adultind	hivposprovm
Adult individual	zimphia2020adultind	lngvint_lngoth_zw
Roster	zimphia2020roster	hhcemanc
Roster	zimphia2020roster	supportsocial3

**Table 9. Variables with new categories**

<b>Dataset(s)</b>	<b>Filename</b>	<b>Variable</b>	<b>New Category</b>
Adult individual	zimphia2020adultind	arvamt	1 – 0 or 1 2 – 2 3 – 3 5 – 4 or 5 6 – 6 or more
Adult individual	zimphia2020adultind	arvloc	96 – Other
Adult individual	zimphia2020adultind	cerncnrslt	96 – Other
Adult individual	zimphia2020adultind	schcom_zw	18 – Level 3 (Master's or Doctoral program)

## 6 Dataset specifications

**Table 10. ZIMPHIA 2020 dataset specifications**

Dataset	Filename	Number of observations	Number of variables
Household	zimphia2020hh	12,460	258
Roster	zimphia2020roster	43,403	60
Adult individual	zimphia2020adultind	20,793	458
Adult biomarker	zimphia2020adultbio	19,535	221
Drug resistance	zimphia2020dr	673	360
Household intermediary weights	zimphia2020hhintermediarywts	12,460	180
Individual intermediary weights	zimphia2020indintermediarywts	43,403	711

Dataset specification	Description
Two-letter country code prefix for ID variables	ZW
Survey weighting variables	
No. of jackknife replicates	175
Survey weights provided (variable prefix)	Household (hhwt) Individual interview (intwt) Blood test (btwt) Drug resistance (drwt)
Selected variable parameters	
Household characteristics used for wealth index construction	<i>See next section</i>
Mean duration recent infection (MDRI) 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 component analysis (PCA) on household characteristics and asset ownership variables. The details of these variables vary by country. The table below lists the variables used to construct the wealth index for ZIMPHIA 2020.

**Table 11. Household characteristics used for wealth index construction in ZIMPHIA 2020**

Indicator variable	Type	Description
memsleep	Continuous	Number of household members per sleeping room <sup>1</sup>
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_zw	Binary	Any external economic support
<i>For the remainder of the variables:</i>		
		<i>Does this household have/own...?</i>
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
		<i>How many of the following does this household have/own?<sup>2</sup></i>
ownchiknnum	Continuous	Chicken
owncownum	Continuous	Cows
owndognum	Continuous	Dogs
owngoatnum	Continuous	Goats
ownhorsenum	Continuous	Horses

<sup>1</sup>Rounded to the nearest integer.

<sup>2</sup>For wealth index calculation, continuous variables for animal ownership have been changed into binary (yes/no). For example, the households that had any chickens were assigned “yes” and the households that had no chickens were assigned “no”. This was done to be consistent with the DHS computation of wealth index (ZIMSTAT, 2016).

### Wealth scores and model performance

The first component of the PCA model is interpreted as an index of household wealth. However, it does not explain a large proportion of the total variance: it accounts for only around 7.33% of the total variance in the common model, 4.02% for the urban model, and 5.18% for the rural model. Howe et al. note that this figure is “often less than 20%” (Howe et. Al., 2008). The results

from ZIMPHIA 2020 are consistent with those of other DHS studies in similar settings (Vyas and Kumaranayake, 2006; Filmer D and Pritchett, 2001; ZIMSTAT, 2016).

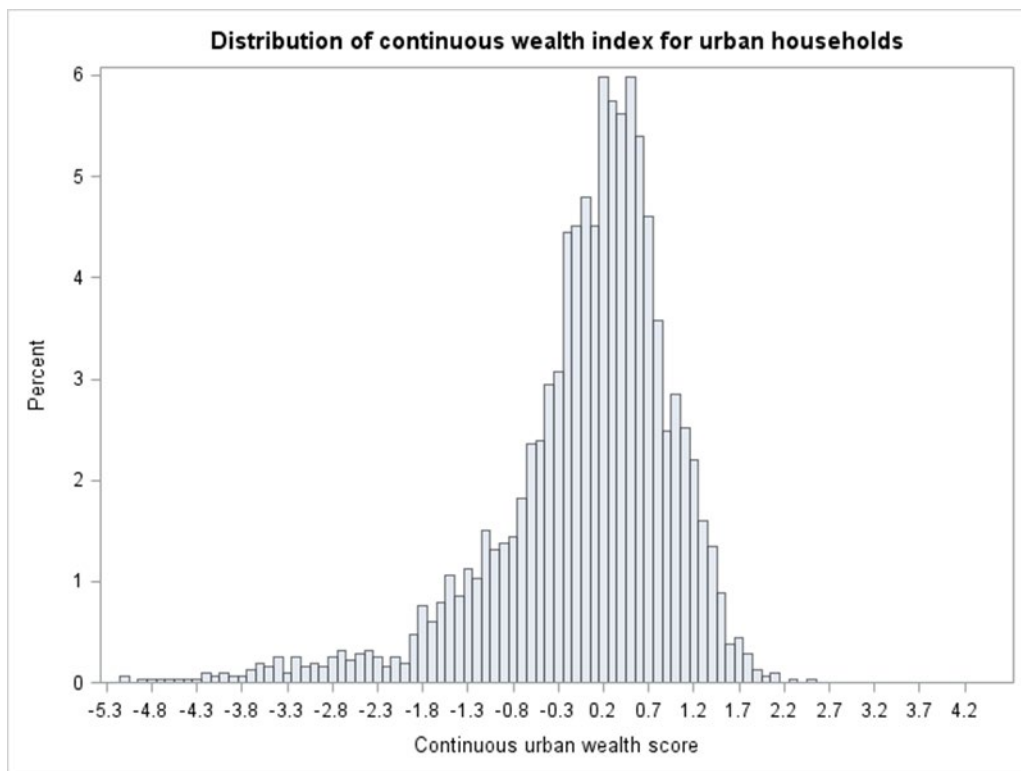
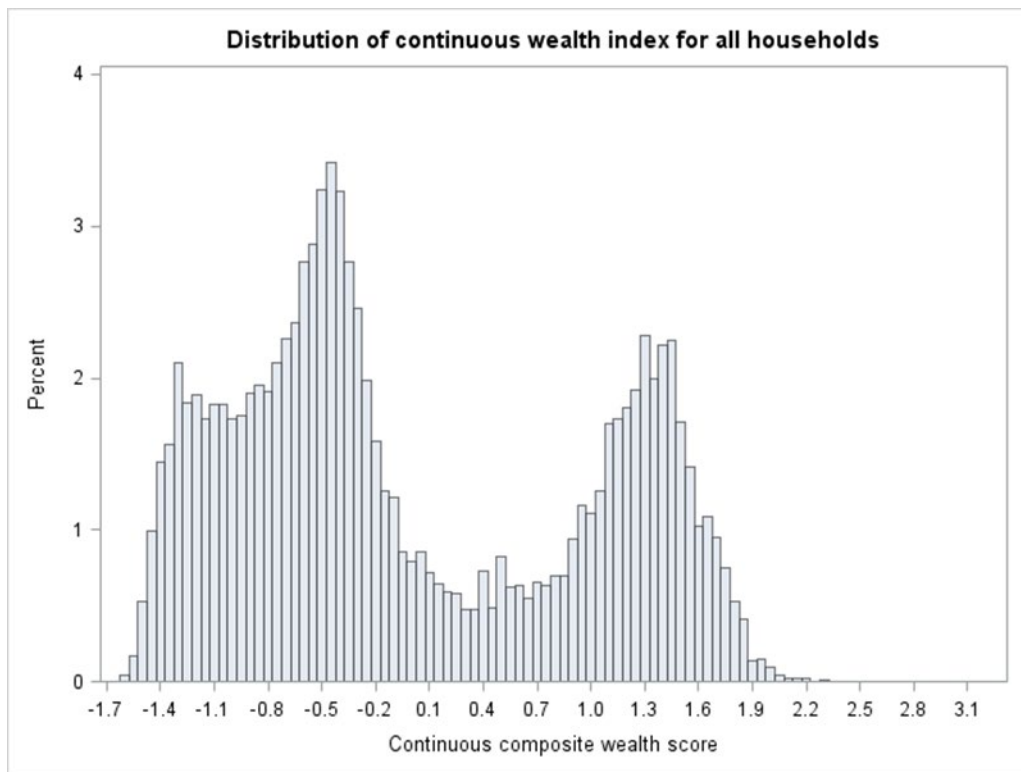
The PCA method does not guarantee the extraction of an index that is actually well-correlated with wealth but results from the PCA can be used to check whether the interpretation of the model makes sense. The component loading for each asset variable describes the association between that asset and the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

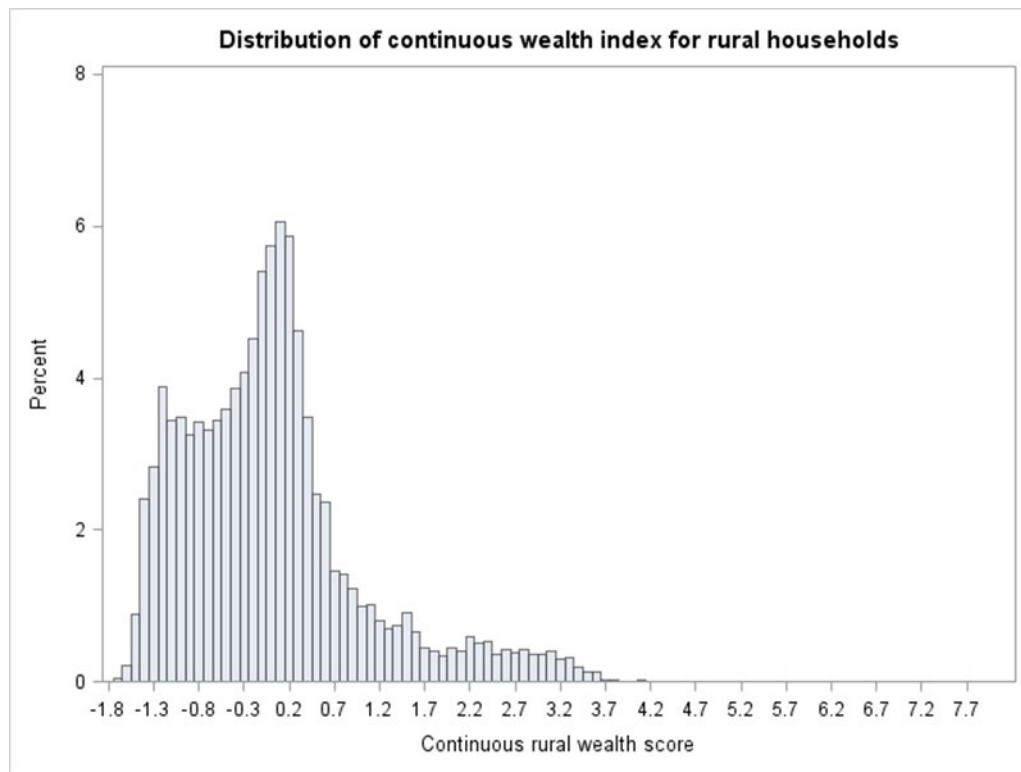
**Table 12. PCA results for ZIMPHIA 2020 wealth quintile**

Variable	Category	Component loading		
		Common model	Urban model	Rural model
Toilet type	flush or pour flush toilet	0.85188	0.62156	0.60652
Electricity in the house	yes	0.84975	0.66823	0.68635
Cooking fuel	firewood / straw	-0.78075	-0.44676	-0.60766
Refrigerator	yes	0.73164	0.60649	0.61871
Television	yes	0.71934	0.61096	0.55615
Cooking fuel	electricity	0.60752	0.42750	0.55142
Water source	pipied into dwelling	0.56053	0.32465	0.41757
Owns chicken	yes	-0.51051	0.16398	-0.23409
Toilet type	no facility/bush/field	-0.46986	-0.25998	-0.47198

In ZIMPHIA 2020, toilet facilities 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 a wealthier household.

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





## 8 References

---

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

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

Ministry of Health and Child Care (MOHCC), Zimbabwe. Zimbabwe Population-based HIV Impact Assessment (ZIMPHIA) 2015-2016: Final Report. Harare: MOHCC; August 2019.

Ministry of Health and Child Care (MoHCC). Zimbabwe Population-based HIV Impact Assessment 2020 (ZIMPHIA 2020): Final Report. [Harare]: MoHCC; December 2021.

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.

Zimbabwe National Statistics Agency (ZIMSTAT), ICF International. Zimbabwe Demographic and Health Survey 2015: Final Report. Rockville, Maryland, USA: Zimbabwe National Statistics Agency and ICF International; 2016.

## 9 Attachments

---

### **9.1 Questionnaires**

See Supplement Attachment 1 - ZIMPHIA 2020 Questionnaires.pdf

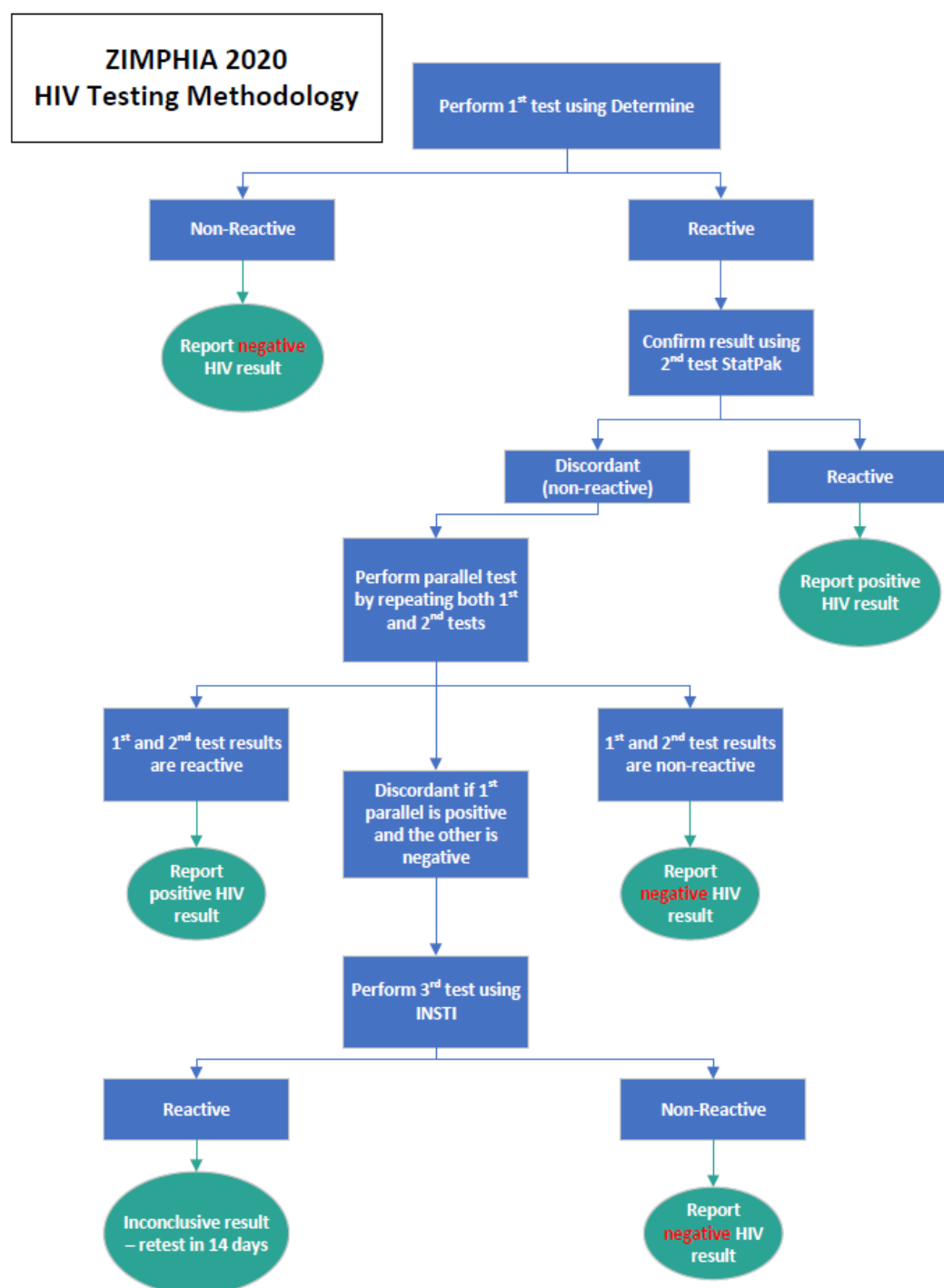
### **9.2 Codebook with frequencies**

See Supplement Attachment 2 - ZIMPHIA 2020 Codebook.pdf

### **9.3 Flow Diagrams for selected analytic variables**

See Supplement Attachment 3 - ZIMPHIA 2020 Flow Diagrams for Analytic Variables.pdf

## 9.4 HIV Testing Methodology Diagram



## 9.5 Sampling and Weighting Technical Report

See Supplement Attachment 5 - ZIMPHIA 2020 Sampling and Weighting Technical Report.pdf

## 9.6 ZIMPHIA 2020 Survey-specific table specifications

See Supplement Attachment 6 - ZIMPHIA 2020 Additional Table Specifications.xlsx

## 9.7 Requesting data

ZIMPHIA 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 ZIMPHIA 2020, select “Zimbabwe” 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>.