ZIMPHIA
ZIMBABWE POPULATION-BASED HIV IMPACT ASSESSMENT

# SAMPLING AND WEIGHTING TECHNICAL REPORT
# ZIMPHIA 2015-2016



This project is supported by the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) through CDC under the terms of cooperative agreement #U2GGH001226.
The contents of this document do not necessarily represent the official position of the funding agencies.

# Table of Contents

## Contents Continued

# Acronyms

| | |
|---|---|
| CASI | Computer Assisted Self Interview |
| CDC | US Centers for Disease Control and Prevention |
| CHAID | Chi-square Automatic Interaction Detector |
| CI | Confidence Interval |
| CV | Coefficient of Variation |
| DEFF | Design Effect |
| DHS | Demographic and Health Survey |
| DU | Dwelling Unit |
| EA | Enumeration Area |
| FTP | File Transfer Protocol |
| HH | Household |
| HIV | Human Immunodeficiency Virus |
| HIVK | HIV Knowledge |
| ICC | Intra Cluster Correlation |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| MDRI | Mean Duration of Recent Infection |
| MOS | Measure of Size |
| PHIA | Population-based HIV Impact Assessment |
| PEPFAR | President's Emergency Plan for AIDS Relief |
| PSU | Primary Sampling Unit |
| RSE | Relative Standard Error |
| SAS | Statistical Analysis System |
| UEW | Unequal Weighting |
| UNAIDS | Joint United Nations Programme on HIV and AIDS |
| USAID | United States Agency for International Development |
| VLS | Viral Load Suppression |
| VM | Violence Module |
| WHO | World Health Organization |
| WLM | Weighted Log-linear Modeling |
| ZIMPHIA | Zimbabwe Population-based HIV Impact Assessment |
| ZIMSTAT | Zimbabwe National Statistics Agency |

# Introduction 1

The 2015 Zimbabwe Population-based HIV Impact Assessment (ZIMPHIA) is a cross-sectional sample survey designed to assess the prevalence of key human immunodeficiency virus (HIV)-related health indicators. The ZIMPHIA was conducted between October 2015 and April 2016, and included over 29,000 individuals in approximately 12,000 households. The purpose of this report is to document the procedures used to select the households and individuals for the study and the subsequent weighting of the respondent sample.

## 1.1 Overview of Sample Design

The sample design for the ZIMPHIA is a stratified multistage probability sample design, with strata defined by the 10 provinces of Zimbabwe, first-stage sampling units defined by enumeration areas (EAs) within strata, second-stage sampling units defined by households within EAs, and finally eligible persons within households.

The first-stage sampling units (also referred to as the "primary sampling units" or PSUs) were stratified by the ten provinces of the country, and then within each province were selected with probabilities proportionate to the number of households in the PSU based on the 2012 census. The allocation of the sample PSUs to the ten provinces was made in a manner designed to achieve specified precision levels for a national estimate of HIV incidence rate, and provincial estimates of viral load suppression (VLS) rates.

The second-stage sampling units were selected from lists of dwelling units/households compiled by trained staff for each of the sampled PSUs. Upon completion of the listing process, a random systematic sample of dwelling units/households was selected from each PSU at rates designed to yield a self-weighting (i.e., equal probability) sample within each province to the extent feasible.

Within the sampled households, all eligible adults 15 years of age or older were included in the study sample for data collection. All eligible children 0-14 years of age in a randomly designated subset of one-half of the selected households were included in the study for data collection.

Details of the sample design employed for the ZIMPHIA are provided in Section 2.

## 1.2 Overview of Weighting Process

The purpose of weighting survey data from a complex sample design is to (1) compensate for variable probabilities of selection, (2) account for differential nonresponse rates within relevant subsets of the sample, and (3) adjust for possible undercoverage of certain population groups. Weighting is accomplished by assigning an appropriate sampling weight to each responding sampled unit (e.g., a household or person), and using that weight to calculate weighted estimates from the sample.

The main steps of the weighting process are:

- Initial checks to confirm that the probabilities of selection associated with the sampled units are computed correctly.
- Creation of jackknife replicates to be used for variance estimation.
- Calculation of PSU base weights to reflect the overall PSU probabilities of selection.
- Adjustment for PSU nonresponse to compensate for PSUs for which no household data were collected.
- Calculation of household weights to reflect the probabilities of selecting households within PSUs, and to compensate for household nonresponse.
- Calculation of person-level interview weights to reflect the differential probabilities of selecting individuals within households, and to compensate for nonresponse to the interview.
- Poststratification of the person-level interview weights to calibrate the weighted counts of persons completing the interview so that they match external population counts.
- Calculation of person-level blood test weights to reflect the differential probabilities of selecting individual within households, compensate for nonresponse to the blood test, and adjust for possible undercoverage through poststratification.

Technical details of the weighting procedures employed in ZIMPHIA are provided in Section 3.

# Sample Design 2

## 2.1 Population of Inference

The population of inference for the PHIA is comprised of individuals who were present in households (i.e., "slept in the household") on the night prior to the date of interview. This population is referred to as the *de facto* population. In contrast, those individuals who are usual residents of the household regardless of whether they were present in the household during the previous night comprise the *de jure* population. All individuals belonging to either the *de facto* or *de jure* populations were included for PHIA data collection; however, as discussed later in Section 2.5, only members of the *de facto* population are included in the PHIA study population. Table 2-1 summarizes projections of the 2016 Zimbabwe *de facto* population by gender and age group that the ZIMPHIA is designed to represent.

**Table 2-1 Summary of 2016 de facto population projections for Zimbabwe by gender and age group**

| Age group | Gender | | Total |
|---|---|---|---|
| | Male | Female | |
| 14 years or younger | 2,871,652 | 2,916,267 | 5,787,919 |
| 15 to 49 years | 3,492,746 | 3,753,791 | 7,246,537 |
| 50 years or older | 607,267 | 837,860 | 1,445,127 |
| Total | 6,971,665 | 7,507,918 | 14,479,583 |

Source: Tables A-4.1 to A-4.10 in Appendix A of the *Population Projections Thematic Report*, Zimbabwe National Statistics Agency (http://www.zimstat.co.zw/sites/default/files/img/publications/Census/population_projection.pdf)

## 2.2 Precision Specifications and Assumptions

The following specifications and assumptions were used to develop the sample design for the ZIMPHIA.

- The overall sample size is 15,000 (i.e., the number of dwelling units to be selected prior to losses due to vacancy);

- The number of first-stage sampling units (EAs) to be selected is 500, with an average of 30 sampled dwelling units per EA;
- The sample size for each of the 10 strata (provinces) are to be determined so that 95% confidence bounds around the estimated viral load suppression (VLS) rate among HIV+ persons aged 15-49 for each province are approximately equal and no greater than ±10%.
- The total sample size must also be sufficient to produce a national annual HIV incidence rate for persons aged 15-49 with a relative standard error (RSE) of 30% or less.
- An overall HIV prevalence rate of 0.152 (15.2%) that varies by stratum (see Table 2-2).
- An annual HIV incidence rate for adults aged 15-49 of $P_a = 0.0096$ (0.96%).
- A mean duration of recent infection (MDRI) of 130 days, yielding an annualization rate of 365/130= 2.8077. Hence, the estimated incidence rate for MDRI = 130 days is $P_m = 0.0096/2.8077 = 0.0034$ (0.34%).
- A viral load suppression (VLS) rate among HIV+ adults aged 15-49 in stratum *h* of $P_{vh} = 50\%$.
- An intra-cluster correlation (ICC) of $\rho = 0.05$ for both prevalence and incidence. The ICC provides an average measure of the homogeneity of responses within the first-stage sampling units.
- An occupancy rate of 92.7% for sampled dwellings. Note that this is not nonresponse but does factor in the calculation of the numbers of dwelling units to be sampled. A sample of about 15,000 dwelling units will yield a sample of about 14,000 occupied dwelling units (households).
- The average number of persons aged 15 to 49 in a household is 1.85 (source: 2010-11 Demographic and Health Survey).
- The percentage of persons in households who are 0-14 is 42.8% source: 2010-11 Demographic and Health Survey).
- The percentage of persons in households who are 50+ is 12.1% source: 2010-11 Demographic and Health Survey).
- Among the individuals in the eligible responding households, a biomarker response rate of 67% for persons 15 year olds or older.
- Among the children in the eligible responding households, a biomarker response rate of 61% for persons 0-14 years of age.

Based on the above assumptions, the sample of 500 clusters was allocated to the 10 strata (provinces) as shown in Table 2-2. Also shown in the table are the corresponding projected numbers of respondents by three broad age groups (15 to 49 years; 50 years and older, and 0 to 14 years). Because a relatively large fixed sample size of 15,000 households had been specified for the ZIMPHIA, sampling precision was expected to be better than the targets indicated in (c) and (d) above. For example, the RSE of the estimated national incidence rate was expected to be 25% under the proposed design, while the 95% confidence bounds around provincial estimates of VLS rates were expected to range from around ±6% to ±7%. Given the uncertainty about many of the assumptions used in the sample design, the actual sample sizes achieved in the study differed from the expectations shown in Table 2-2. Sections 2.4 and 2.5 summarize the actual numbers of households and individuals that participated in the ZIMPHIA.

**Table 2-2. Allocation of sample clusters (EAs) and dwelling units and projected sample sizes (number of respondents) by stratum**

| Stratum (Province) | Est. HIV prevalence rate [1] | Sample clusters (EAs) | Target no. of dwelling units (DUs) to sample | House-holds [2] | Projected number of respondents [3] | | |
|---|---|---|---|---|---|---|---|
| | | | | | 15-49 | 50+ | 0-14 [4] |
| Bulawayo | 0.191 | 43 | 1,292 | 1,213 | 1,434 | 397 | 643 |
| Harare | 0.134 | 57 | 1,696 | 1,592 | 1,882 | 521 | 844 |
| Manicaland | 0.141 | 54 | 1,629 | 1,529 | 1,808 | 500 | 811 |
| Mashonaland Central | 0.137 | 56 | 1,666 | 1,564 | 1,849 | 512 | 829 |
| Mashonaland East | 0.157 | 50 | 1,497 | 1,406 | 1,662 | 460 | 745 |
| Mashonaland West | 0.148 | 52 | 1,568 | 1,472 | 1,740 | 481 | 780 |
| Masvingo | 0.144 | 53 | 1,602 | 1,504 | 1,778 | 492 | 797 |
| Matabeleland North | 0.183 | 44 | 1,333 | 1,252 | 1,480 | 409 | 664 |
| Matabeleland South | 0.212 | 40 | 1,198 | 1,125 | 1,329 | 368 | 596 |
| Midlands | 0.154 | 51 | 1,520 | 1,427 | 1,687 | 467 | 757 |
| TOTAL | 0.152 | 500 | 15,000 | 14,085 | 16,650 | 4,607 | 7,467 |

[1] Source: 2010-11 Zimbabwe Demographic and Health Survey (DHS).
[2] Assumes occupancy rate of 0.939 (source: 2010-11 Zimbabwe DHS).
[3] Entries are projected counts based on the assumptions used to develop the sample design. See Section 2.5 for actual sample sizes achieved.
[4] All responding children in 50% of the participating households.

## 2.3 Selection of the Primary Sampling Units (PSUs)

### 2.3.1 Definition of PSUs

The first-stage or primary sampling units (PSUs) for the ZIMPHIA were defined to be the Enumeration Areas (EAs) created for the 2012 Zimbabwe Population Census. The sampling frame consisted of 29,365 EAs, stratified by province, containing an estimated 3,059,016 households and 12,927,301 persons, with an average number of households and persons per EA of 104 and 440, respectively.

### 2.3.2 Selection of the PSU sample

A stratified sample of 500 EAs was selected from the EA sampling frame in accordance with the sample allocation given in Table 2-1. The samples were selected systematically and with probabilities proportionate to a measure of size (MOS) within each province. The MOS used for sampling was equal to the number of households in the EA based on the 2012 Population Census. The first step of the sampling process was to divide the sampling frame of EAs into strata corresponding to the 10 provinces of the country. Next, the EAs were sorted by urban/rural status, district within each urban/rural status, and finally by ward within district. The sorting of the EAs prior to sample selection induces an implicit stratification of the sampling frame designed to ensure that a representative mix of EAs with respect to urban/rural status and geography are included in the sample. To select the sample from a particular province, the cumulative MOS was determined for each EA in the ordered list of EAs, and the sample selections were designated using a sampling interval equal to the total MOS of the EAs in the province divided by the number of EAs to be selected and a random starting point. The resulting sample has the property that the probability of selecting an EA for the study is proportional to the MOS of the EA within the province.

### 2.3.3 Segmentation

Of the 500 sampled EAs, three were deemed to be very large with an estimated 250 households or more in each. Possible strategies for listing them included (a) listing them entirely or (b) dividing them into smaller subareas referred to as segments, randomly selecting a segment, and listing the selected segment. Of the three large EAs, two were listed entirely, and one underwent the segmentation process in which (a) the EA was subdivided into two segments, (b) a rough measure of size was assigned to each defined segment, and (c) one segment was randomly selected with

probability proportionate to the rough measure of size for listing. The segmentation procedures used in PHIA are described in **Zimbabwe HIV Impact Assessment: Manual for Household Mapping and Listing**, prepared by the Zimbabwe National Statistics Agency.

### 2.3.4 Substitution

Four of the PSUs (enumeration areas) originally selected for the study were replaced. Three of them were inaccessible for various reasons, and one was a secure Army camp where listers were not allowed to enter. All four of these EAs are considered to be eligible for PHIA because they contained occupied dwelling units. The replacement EAs were identified by locating the position of the originally-selected EA in the ordered sampling frame, and then selecting the EA immediately preceding it on the list within the same substratum defined by the sorting variables used in sample selection. If there were no EAs preceding the original EA, the EA immediately following it was chosen. In this way, the substitute EA will have characteristics broadly similar to the originally-sampled EA. For subsequent sampling and weighting purposes, the probability of selecting the substitute EA was set equal to the probability of selection it would have had if it had originally been selected.

### 2.3.5 Results of PSU Sampling

Table 2-3 summarizes the distribution of the 500 PSUs selected for the study by the 10 provinces of Zimbabwe, and the corresponding numbers of EAs that were substituted or segmented.

Table 2-3 Distribution of the sampled EAs by sampling status

| Stratum (Province) | Number of sample EAs | Number of replaced EAs | Number of segmented EAs | Number of inscope EAs (clusters) included in study |
|---|---|---|---|---|
| Bulawayo | 43 | 0 | 0 | 43 |
| Harare | 57 | 1 | 0 | 57 |
| Manicaland | 54 | 0 | 1 | 54 |
| Mashonaland Central | 56 | 2 | 0 | 56 |
| Mashonaland East | 50 | 0 | 0 | 50 |
| Mashonaland West | 52 | 1 | 0 | 52 |
| Masvingo | 53 | 0 | 0 | 53 |
| Matabeleland North | 44 | 0 | 0 | 44 |
| Matabeleland South | 40 | 0 | 0 | 40 |
| Midlands | 51 | 0 | 0 | 51 |
| Total | 500 | 4 | 1 | 500 |

## 2.4 Selection of Households

The selection of households for the ZIMPHIA involved the following steps: (1) listing the dwelling units/households within the sampled EAs, (2) assigning eligibility codes to the listed dwelling unit/household records, (3) selecting the samples of dwelling units/households, and (4) designating a subsample of households for collection of child data.

### 2.4.1 Definition of Second-Stage Sampling Units

For both sampling and analysis purposes, a household is defined to be a group of individuals who reside in a physical structure such as a house, apartment, compound, or homestead, and share in housekeeping arrangements. The physical structure in which people reside is referred to as the "dwelling unit" which may contain more than one household meeting the above definition. Households are eligible for participation in the study if they are located within the sampled enumeration area (EA).

### 2.4.2 Listing

In essence, the listing process involves compiling complete, up-to-date, and accurate lists of all dwelling units and households for each sampled EA through a field operation using trained staff referred to as "listers." For each of the 500 EAs selected for the study, listers were provided with a Census sketch map (made in 2010 for the 2012 Population and Housing Census) from which to delineate the boundaries of the EA, and to record the general locations of the dwelling units/households that are found by the listers in the field. Information about the listed dwelling units/households matching the information on the sketch maps was also recorded on paper forms. The paper forms included information about the head of household, household size, and other information to assist the interviewer in locating the dwelling unit/household during data collection. The information on the paper forms was transferred to electronic format for subsequent data cleaning and sampling. Over 55,000 dwelling units/households were listed for the ZIMPHIA.

### 2.4.3 Determination of Eligibility for Sampling

Because of confidentiality concerns, only the bare minimum information required for sampling and linking back to the paper forms was included in the electronic files, e.g., province and EA codes, a unique line number to enable linking to the source information on the paper forms, and a variable

that indicated whether the listed unit was an inhabited or habitable dwelling (coded "Y") or not (coded "N"). The "N" category included not only structures that would not normally serve as dwellings (such as shops, churches, schools, etc.) but also some mainly rural, former dwellings in run-down conditions whose inhabitants had moved away, and which were very unlikely to be inhabited in the future. On the other hand, if there was evidence that part of a shop, school or similar building was also used as residential living quarters, the listed unit would be indicated as a "Y".

On the basis of this information, a formal decision was taken to consider those lines on the listing form with code "Y" to represent (potential) households, and all others to be excluded from consideration in sampling. That is, a "Y" could be currently inhabited (an actual household), or vacant but potentially inhabited at the time of the survey fieldwork. Lines indicated as "vacant" but with code "N" were made out of scope for sampling. Subsequent quality control checks identified 36 records with no household line number, and these were also classified as not eligible for sampling.

Table 2-4 summarizes the number of listings (households or dwelling units) identified by the listers, the number of discarded listings, the number of unoccupied and occupied dwelling units based on the information collected during listing, and the total number of dwelling units that were eligible for sampling.

Table 2-4 Distribution of records in listing file by type of record and eligibility status

| Stratum (Province) | Number of listings (dwelling units/households) | Number of listings discarded | Number of unoccupied dwelling units | Number of occupied dwelling units | Number of dwelling units eligible for sampling |
|---|---|---|---|---|---|
| Bulawayo | 4,139 | 2 | 133 | 4,006 | 4,137 |
| Harare | 6,129 | 0 | 226 | 5,903 | 6,129 |
| Manicaland | 5,900 | 28 | 367 | 5,533 | 5,872 |
| Mashonaland Central | 6,458 | 1 | 385 | 6,073 | 6,457 |
| Mashonaland East | 5,740 | 2 | 309 | 5,431 | 5,738 |
| Mashonaland West | 6,157 | 0 | 247 | 5,910 | 6,157 |
| Masvingo | 5,597 | 1 | 118 | 5,479 | 5,596 |
| Matabeleland North | 4,795 | 0 | 171 | 4,624 | 4,795 |
| Matabeleland South | 4,430 | 1 | 74 | 4,356 | 4,429 |
| Midlands | 5,713 | 1 | 310 | 5,403 | 5,712 |
| Total | 55,058 | 36 | 2,340 | 52,718 | 55,022 |

### 2.4.4 Selection of Dwelling Units

A goal of sampling for the ZIMPHIA was to select an average of 30 dwelling units per EA. In order to achieve an equal probability sample of dwelling units within each province, the sampling rates required to select dwelling units within an EA depended on the difference between the size measure used in sampling (i.e., the number of households in the EA based on the 2012 census) and the actual number of households found at the time of listing. Thus, application of these within-EA sampling rates can yield more than 30 households in EAs that have experienced growth in population since the 2012 census, and fewer than 30 households in EAs that have declined in population.

The calculation of the required within-EA sampling rates proceeded as follows. First, the target overall sampling rate for stratum (province) $h$ = 1, 2, ..., 10, was computed as:

$$F_h^{overall} = T_h \,/\, \sum_{i=1}^{m_h} (N_{hi} \,/\, P_{hi}) ,$$

where

$T_h$ = target sample size for stratum $h$ given in Table 2-2;

$m_h$ = number of sample EAs in stratum $h$ given in Table 2-2;

$N_{hi}$ = number of eligible dwelling units in PSU $i$ in stratum $h$ based on listing counts;

$P_{hi}$ = probability of selecting PSU $i$ in stratum $h$ .

The total *targeted* number of listings to be selected across all 10 strata is $\sum_{h=1}^{10} T_h = 15{,}000$ (see Table 2-2). The probabilities of selection, $P_{hi}$, for the four substitute EAs (see Section 2.3.4) were set to the probabilities they would have had if they had originally been selected for the sample. The probability of selection of the segmented EA was set to $P_{hs} = P_{hi}^{EA} P_{s|hi}^{seg}$, where $P_{hi}^{EA}$ = the selection probability of EA *hi*, and $P_{s|hi}^{seg}$ = the conditional probability of selecting segment *s* in EA *hi*.

To obtain an equal probability sample within stratum *h*, the required within-EA sampling rate for EA *i* in stratum *h* was then computed as:

$$f_{hi}^{within} = F_h^{overall} / P_{hi}.$$

and the corresponding expected sample size for EA i in stratum h was computed as:

$$\mathrm{E}(n_{hi}) = N_{hi} f_{hi}^{within}.$$

Inspection of the values of $\mathrm{E}(n_{hi})$ indicated that there would be unduly large workloads in some EAs. To maintain acceptable workloads in EAs that experienced considerable growth, the maximum number of dwelling units to be selected in any EA was capped at no more than 60. The difference between the number of dwelling units that would have been selected and the capped number was then re-distributed to the other EAs in the same stratum so as to maintain the desired total sample size. The within-EA sampling rates, $f_{hi}^{within}$, were thus adjusted to reflect the capping and the redistribution of the sample within the stratum. The adjusted within-EA sampling rate used to select the sample of dwelling units, $f_{hi}^{adj(w)}$, was calculated as:

$$f_{hi}^{adj(w)} = A_{hi} f_{hi}^{within},$$

where the adjustment factors, $A_{hi}$, were determined such that $A_{hi} f_{hi}^{within} \leq 60$ and $\sum_{i=1}^{m_h} A_{hi} f_{hi}^{within} = T_h$.

To preserve the geographical order in which they were listed, the eligible dwelling units in each EA were sorted by the line number assigned during listing. A total of 15,009 dwelling units was then selected systematically from the ordered lists at the rates, $f_{hi}^{adj(w)}$, specified above. In addition, a random subsample of 7,510 of the 15,009 selected dwelling units were designated (flagged) for child data collection.

### 2.4.5 Results of Second-Stage Sampling

Table 2-5 summarizes the numbers of dwelling units/households selected for the study, the number designated for child data collection, and the minimum and maximum EA sample size by stratum (province). The last column shows the unequal weighting (UEW) design effects to be expected for the selected sample. The UEW design effect provides a measure of the increase in the variance of a sample-based estimate resulting from the application of variable overall sampling fractions within a stratum (e.g., see Kish, 1965, page 403). With an equal probability sample within a stratum, the design effects would ordinarily equal 1.0. However, with the capping and redistribution of the sample described previously, the overall sampling rates (and, hence, household weights) will vary within a stratum. Despite the variation in weights, the UEW design effects are all very close to 1.0 (indicating minimal increase in variance due to unequal weighting) for all strata.

Table 2-6 provides a summary of the number of dwelling units/households selected for PHIA by final survey response status. Of the 15,009 sampled dwelling units, 1,038 (6.9%) were determined during data collection to be ineligible (vacant, destroyed, nonresidential), 178 (1.2%) for which eligibility for the survey (i.e., occupancy status) could not be established, 2,076 (13.8%) were determined to be eligible for the study (i.e., contained eligible household members) but did not complete the household roster, and 11,717 (78.0%) completed the household roster. Excluding the known 1,038 ineligible cases, the unweighted response rate (i.e., the percent of sampled households completing the household roster) was 83.9%.

Table 2-5 Number of sampled dwelling units/households and expected unequal weighting design effects by stratum

| Stratum (Province) | Number of sample EAs (clusters) | Number of sampled dwelling units/house-holds | Number of dwelling units/house-holds flagged for child data collection | Minimum EA sample size | Maximum EA sample size | UEW design effect for PHIA sample after capping |
|---|---|---|---|---|---|---|
| Bulawayo | 43 | 1,292 | 649 | 17 | 44 | 1.00 |
| Harare | 57 | 1,710 | 856 | 11 | 50 | 1.01 |
| Manicaland | 54 | 1,623 | 812 | 10 | 51 | 1.00 |
| Mashonaland Central | 56 | 1,683 | 841 | 13 | 53 | 1.00 |
| Mashonaland East | 50 | 1,499 | 753 | 20 | 51 | 1.00 |
| Mashonaland West | 52 | 1,561 | 782 | 15 | 55 | 1.00 |
| Masvingo | 53 | 1,590 | 793 | 18 | 54 | 1.00 |
| Matabeleland North | 44 | 1,322 | 657 | 14 | 57 | 1.00 |
| Matabeleland South | 40 | 1,201 | 603 | 13 | 52 | 1.02 |
| Midlands | 51 | 1,528 | 764 | 18 | 47 | 1.00 |
| Total | 500 | 15,009 [1] | 7,510 | 10 | 57 | 1.08 [2] |

[1] Counts of sampled dwelling units differ slightly from targets given in Table 2-1.

[2] Reflects variation in weights across and within EAs.

Table 2-6 Distribution of dwelling unit sample by response status

| Stratum (Province) | Number of sampled dwelling units (DUs) | Number of ineligible DUs [1] | Number of DUs with unknown eligibility [2] | Number of households completing roster | Number of eligible nonresponding households | Unweighted response rate [3] |
|---|---|---|---|---|---|---|
| Bulawayo | 1,292 | 22 | 9 | 1,032 | 229 | 0.813 |
| Harare | 1,710 | 61 | 53 [4] | 1,219 | 377 | 0.740 |
| Manicaland | 1,623 | 184 | 11 | 1,278 | 150 | 0.889 |
| Mashonaland Central | 1,683 | 136 | 37 | 1,215 | 295 | 0.787 |
| Mashonaland East | 1,499 | 161 | 20 | 1,168 | 150 | 0.874 |
| Mashonaland West | 1,561 | 99 | 11 | 1,289 | 162 | 0.882 |
| Masvingo | 1,590 | 131 | 14 | 1,262 | 183 | 0.866 |
| Matabeleland North | 1,322 | 72 | 7 | 1,102 | 141 | 0.882 |
| Matabeleland South | 1,201 | 78 | 8 | 958 | 157 | 0.853 |
| Midlands | 1,528 | 94 | 8 | 1,194 | 232 | 0.833 |
| Total | 15,009 | 1,038 | 178 | 11,717 | 2,076 | 0.839 |

[1] Vacant, destroyed, non-residential, households with no persons eligible for PHIA.

[2] Dwelling units for which occupancy status could not be determined.

[3] Computed as $R/ [ R + N + U*\{ (R + N)/( R + N + I ) \} ]$, where R = number of households completing roster; N = number of eligible nonresponding households; I = number of ineligible DUs, and U = number of DUs with unknown eligibility.

[4] Includes 38 dwelling units in one EA for which eligibility for the study could not be determined.

## 2.5 Selection of Individuals within Households

The selection of individuals for the ZIMPHIA involved the following steps: (1) compiling a list of all individuals known to reside in the household or who slept in the household during the night prior to data collection; (2) identifying those rostered individuals who are eligible for data collection; (3) selecting those individuals meeting the age and residency requirements of the study. However, as noted below, only those individuals who were present in the household the night before the interview (i.e., the *de facto* population) are retained for subsequent weighting and analysis.

### 2.5.1 Household Rosters

A comprehensive list (roster) of all household members was compiled during the administration of the household interview. The rosters included all persons who were present in the household during the night prior to the interview, along with other individuals who are usual residents of the household but were away during that time. The information recorded for each rostered individual included sex, age, relationship to head of household, residency status (i.e., whether a usual resident), and physical presence in household (i.e., slept in household the night prior to interview). Table 2-7 summarizes the number of households completing the roster and the corresponding number of rostered individuals by stratum and resident status.

Table 2-7 Number of households completing rosters and number of persons by resident status

| Stratum (Province) | Number of households completing rosters | Usual resident but did not sleep here | Usual resident and slept here | Nonresident but slept here | Total |
|---|---|---|---|---|---|
| Bulawayo | 1,032 | 25 | 3,834 | 72 | 3,931 |
| Harare | 1,219 | 68 | 4,406 | 61 | 4,535 |
| Manicaland | 1,278 | 109 | 5,022 | 115 | 5,246 |
| Mashonaland Central | 1,215 | 236 | 4,946 | 139 | 5,321 |
| Mashonaland East | 1,168 | 173 | 4,273 | 127 | 4,573 |
| Mashonaland West | 1,289 | 183 | 5,240 | 196 | 5,619 |
| Masvingo | 1,262 | 118 | 5,052 | 111 | 5,281 |
| Matabeleland North | 1,102 | 95 | 4,539 | 114 | 4,748 |
| Matabeleland South | 958 | 29 | 3,780 | 84 | 3,893 |
| Midlands | 1,194 | 44 | 4,603 | 73 | 4,720 |
| Total | 11,717 | 1,080 | 45,695 | 1,092 | 47,867 |

### 2.5.2 Selecting Individuals for Data Collection

All of the individuals listed in the household rosters who were 15 years of age or older and were either usual residents of the household or who slept in the household were eligible for data collection. Basic information about all children was obtained from parents or guardians in the child module of the adult questionnaire, but children 0-14 years of age were eligible for additional data collection only if the household in which they resided had been randomly designated for child biomarker data collection (see Section 2.4.5). Table 2-8 summarizes the number of individuals eligible for data collection by stratum, age group, and resident status.

Although data collection was attempted for all of 28,516 adults and 9,731 children indicated in Table 2-8, only those individuals in the *de facto* population will be weighted (see Section 3) and included in analysis. The *de facto* population is represented by the 27,741 adults and 9,563 children who slept in the household during the night prior to the interview.

Table 2-8 Number of individuals eligible for data collection

| Stratum (Province) | Adults 15 or older [1] | | | | Children 0-14 [1][2] | | | |
|---|---|---|---|---|---|---|---|---|
| | Usual resident but did not sleep here | Usual resident and slept here | Non-resident but slept here | Total | Usual resident but did not sleep here | Usual resident and slept here | Non-resident but slept here | Total |
| Bulawayo | 19 | 2517 | 56 | 2,592 | 0 | 706 | 7 | 713 |
| Harare | 53 | 2902 | 48 | 3,003 | 8 | 772 | 8 | 788 |
| Manicaland | 78 | 2839 | 84 | 3,001 | 17 | 1085 | 18 | 1,120 |
| Mashonaland Central | 150 | 2818 | 86 | 3,054 | 47 | 1103 | 33 | 1,183 |
| Mashonaland East | 125 | 2513 | 77 | 2,715 | 23 | 867 | 32 | 922 |
| Mashonaland West | 137 | 3108 | 131 | 3,376 | 32 | 1067 | 43 | 1,142 |
| Masvingo | 91 | 2841 | 82 | 3,014 | 16 | 1035 | 15 | 1,066 |
| Matabeleland North | 61 | 2560 | 77 | 2,698 | 19 | 996 | 16 | 1,031 |
| Matabeleland South | 26 | 2171 | 47 | 2,244 | 2 | 808 | 12 | 822 |
| Midlands | 35 | 2736 | 48 | 2,819 | 4 | 922 | 18 | 944 |
| Total | 775 | 27,005 | 736 | 28,516 | 168 | 9,361 | 202 | 9,731 |

[1] Age recorded in roster. In a small number of cases, the actual age at interview may be different. See Section 3.4.3.

[2] Includes only those children in households selected for child blood draw.

### 2.5.3 Distribution of Person Samples

Tables 2-9A through 2-9C summarize the number of individuals selected for data collection and the corresponding numbers completing the interview and blood test, for adults 15 years and over, adolescents 10-14 years, and children 0-9 years, respectively, where the age classification is based on the rostered age. The numbers of completed interviews and blood tests that can be weighted to represent the PHIA study population are shown under the *de facto* heading in these tables. Note that counts of children in these tables include only children in households selected for child blood draw, and that for children 0-9 years in Table 2-9C the counts of completed "interviews" refer to the number of children for whom a parent completed the child questionnaire module for that particular child.

**Table 2-9A Distribution of completed interviews and blood tests for adults 15 years or older**

| Stratum (Province) | *De facto* [1] | | | *De jure but not de facto* [2] | | |
|---|---|---|---|---|---|---|
| | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] |
| Bulawayo | 2,573 | 2,245 | 2,071 | 19 | 10 | 8 |
| Harare | 2,950 | 2,415 | 2,172 | 53 | 30 | 28 |
| Manicaland | 2,923 | 2,694 | 2,505 | 78 | 35 | 33 |
| Mashonaland Central | 2,904 | 2,563 | 2,264 | 150 | 69 | 64 |
| Mashonaland East | 2,590 | 2,381 | 2,179 | 125 | 75 | 70 |
| Mashonaland West | 3,239 | 2,935 | 2,688 | 137 | 63 | 60 |
| Masvingo | 2,923 | 2,648 | 2,430 | 91 | 57 | 49 |
| Matabeleland North | 2,637 | 2,371 | 2,188 | 61 | 32 | 30 |
| Matabeleland South | 2,218 | 2,008 | 1,853 | 26 | 16 | 14 |
| Midlands | 2,784 | 2,463 | 2,215 | 35 | 20 | 19 |
| Total | 27,741 | 24,723 | 22,565 | 775 | 407 | 375 |

[1] Persons who were reported to have slept in the household last night.
[2] Usual residents of the household who did not sleep in the household last night.
[3] Persons who completed the blood test but not the interview are treated as interview respondents for weighting purposes. See Appendix B for more information about the response status categories defined for the individual interview.
[4] These are cases that provided an analyzable blood sample, regardless of whether the individual interview was completed. Of the 22,565 de facto cases completing the blood test, one did not complete the interview but is treated as an interview respondent for weighting purposes. See Appendix B for more information about the response status categories defined for the blood tests.

Table 2-9B Distribution of completed interviews and blood tests for adolescents 10-14 years in households selected for child biomarker collection

| Stratum (Province) | *De facto* [1] | | | *De jure but not de facto* [2] | | |
|---|---|---|---|---|---|---|
| | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] |
| Bulawayo | 195 | 148 | 138 | 0 | 0 | 0 |
| Harare | 204 | 156 | 148 | 1 | 0 | 0 |
| Manicaland | 355 | 313 | 300 | 3 | 0 | 0 |
| Mashonaland Central | 361 | 243 | 214 | 14 | 2 | 2 |
| Mashonaland East | 316 | 268 | 251 | 8 | 5 | 5 |
| Mashonaland West | 336 | 279 | 266 | 8 | 4 | 4 |
| Masvingo | 351 | 304 | 294 | 1 | 0 | 0 |
| Matabeleland North | 310 | 250 | 237 | 2 | 2 | 2 |
| Matabeleland South | 260 | 201 | 191 | 0 | 0 | 0 |
| Midlands | 288 | 180 | 170 | 1 | 0 | 0 |
| Total | 2,976 | 2,342 | 2,209 | 38 | 13 | 13 |

[1] Persons who were reported to have slept in the household last night.

[2] Usual residents of the household who did not sleep in the household last night.

[3] Persons who completed the blood test but not the interview are treated as interview respondents for weighting purposes. See Appendix B for more information about the response status categories defined for the individual interview.

[4] These are cases that provided an analyzable blood sample, regardless of whether the individual interview was completed. Of the 2,209 de facto cases completing the blood test, two did not complete the interview but are treated as interview respondents for weighting purposes. See Appendix B for more information about the response status categories defined for the blood tests.

Table 2-9C Distribution of completed interviews and blood tests for children 0-9 years in households selected for child biomarker collection

| Stratum (Province) | *De facto* [1] | | | *De jure but not de facto* [2] | | |
|---|---|---|---|---|---|---|
| | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] | Number selected for data collection | Number completing interview[3] | Number completing blood test[4] |
| Bulawayo | 518 | 473 | 357 | 0 | 0 | 0 |
| Harare | 576 | 538 | 384 | 7 | 4 | 0 |
| Manicaland | 748 | 710 | 586 | 14 | 8 | 3 |
| Mashonaland Central | 775 | 710 | 435 | 33 | 23 | 6 |
| Mashonaland East | 583 | 549 | 445 | 15 | 14 | 4 |
| Mashonaland West | 774 | 732 | 595 | 24 | 18 | 7 |
| Masvingo | 699 | 659 | 575 | 15 | 8 | 4 |
| Matabeleland North | 702 | 649 | 529 | 17 | 17 | 10 |
| Matabeleland South | 560 | 503 | 412 | 2 | 0 | 0 |
| Midlands | 652 | 568 | 441 | 3 | 1 | 1 |
| Total | 6,587 | 6,091 | 4,759 | 130 | 93 | 35 |

[1] Persons who were reported to have slept in the household last night.

[2] Usual residents of the household who did not sleep in the household last night.

[3] Persons who completed the blood test but not the interview are treated as interview respondents for weighting purposes. See Appendix B for more information about the response status categories defined for the individual interview.

[4] These are cases that provided an analyzable blood sample, regardless of whether the individual interview was completed. Of the 4,759 de facto cases completing the blood test, 44 did not complete the interview but are treated as interview respondents for weighting purposes. For children ages 0-9, "interview" is defined as "data provided by the linked adult interview". See Appendix B for more information about the response status categories defined for the blood tests.

# 3 Weighting and Estimation

In general, the purpose of weighting survey data from a complex sample design is to (1) compensate for variable probabilities of selection, (2) account for differential nonresponse rates within relevant subsets of the sample, and (3) adjust for possible undercoverage of certain population groups. Weighting is accomplished by assigning an appropriate sampling weight to each responding sampled unit (e.g., a household or person), and using that weight to calculate weighted estimates from the sample. The critical component of the sampling weight is the base weight which is defined to be the reciprocal of the probability of including a household or person in the sample. The base weights are used to inflate the responses of the sampled units to population levels and are generally unbiased (or consistent) if there is no nonresponse or noncoverage in the sample (e.g., see Kish, 1965, page 67). When nonresponse or noncoverage occurs in the survey, weighting adjustments are applied to the base weights to compensate for both types of sample omissions.

Nonresponse is unavoidable in virtually all surveys of human populations. For PHIA, nonresponse can occur at different stages of data collection, for example, (1) before the enumeration of individuals in the household, (2) after household enumeration and selection of persons but before completion of the individual interview, and (3) after completion of the interview but before collection of a usable blood sample. The procedures used to compensate for nonresponse at each of the relevant stages of data collection are described in Section 3.4.

Noncoverage arises when some members of the survey population have no chance of being selected for the sample. For example, noncoverage can occur if the field operations fail to enumerate all dwelling units during the listing process, or if certain household members are omitted from the household rosters. To compensate for such omissions, the poststratification procedures described in Sections 3.4.3 and 3.4.4 are used to calibrate the weighted sample counts to available population projections.

## 3.1 Overview of the Weighting Process

The overall weighting approach for PHIA Zimbabwe includes several steps.

**Initial checks:** Checks of the data files are carried out as part of the survey and data quality control, and the probabilities of selection for PSUs and households are calculated and checked.

**Creation of Jackknife Replicates**: The variables needed to create the jackknife replicates for variance estimation are established at this point. This step can be implemented immediately after the PSU sample has been selected. All of the subsequent weighting steps described below are applied to the full sample, and to each of the jackknife replicates

**Calculation of PSU Base Weights:** The weighting process begins with the calculation and checking of the sample PSU (EA) base weights as the reciprocals of the overall PSU probabilities of selection.

**Adjustment for PSU Nonresponse**: Since one EA with 38 sampled dwelling units in one of the provinces had no household data collected, an EA nonresponse adjustment is made for the remaining "responding" EAs in this province.

**Calculation of Household Weights:** The next step is to calculate household weights. The household base weights are calculated as the nonresponse adjusted EA weights times the reciprocal of the within-EA household selection probabilities. The household base weights are adjusted first to account for dwelling units for which it could not be determined whether the dwelling unit contained an eligible household (as shown in Table 2-6 above, this only happened for 1.2% of the listings) and then the responding households have their weights adjusted to account for nonresponding eligible households. This adjustment is made based on the EA the households are in, and the resulting weight is the final household weight.

**Calculation of Person-Level Interview Weights:** Once the household weights are determined, they are used to calculate the individual base weights. The individual base weights are then adjusted for nonresponse among the eligible individuals, with a final adjustment for the individual weights to compensate for undercoverage in the sampling process by weighting up to 2016 population projections produced by the Zimbabwe National Statistics Agency (ZIMSTAT). For children in households not selected for child blood draws (see Section 2.4.5), data was collected from eligible

parents or guardians, but the children were not assigned interview weights. For analysis of this full set of children, child module weights were generated after all other weighting was completed. See Appendix H for details.

**Calculation of Person-Level Blood Test Weights:** The individual weights adjusted for nonresponse are in turn the base weights for the blood data sample, with a further adjustment for nonresponse to the blood draw, and a final poststratification adjustment to compensate for undercoverage.

**Application of Weighting Adjustments to Jackknife Replicates**: All of the adjustment processes are applied to the full sample and the replicate samples so that the final set of full sample and replicate weights can be used for variance estimation that takes into account the complex sample design and every step of the weighting process.

## 3.2 Preparation for Weighting

Five basic data files are used as input to the weighting process. In this section we discuss these files from the perspective of the weighting process.

### 3.2.1 Data Files for Weighting

The PHIA survey data that are used to construct the sampling weights are contained in the following data files. These are work files created and used during the weighting process and are not included in the public-use data.

- **phiazim_cff_hhqx_20161201**: A household (HH) file that contains the majority of household data collected in the HH questionnaire.
- **phiazim_cff_hhdeath_20161201**: A household (HH) file that contains data collected in the HH questionnaire regarding any deaths that have occurred in the household since 2013.
- **phiazim_cff_hhroster_20161201**: A file that contains the roster of household members collected in the HH questionnaire with a record for each rostered person.
- **phiazim_cff_indiv_20161201**: An individual level file that includes data collected on individual questionnaire tablets. This file contains data from the appropriate questionnaire modules for each person, with "null" values for those modules that do

not apply to that person. So variables for individual questionnaire data collected from persons aged 15 and over, for individual questionnaire data collected from persons aged 10 to 14, for children under 10 for data collected from the child's parent or guardian are all included in every record, with values only for the applicable variables.

- **ZimBiomarker20161220**: A biomarker file containing identifying information and results for lab analyses of blood samples for individuals whose blood was drawn and analyzed in the lab.

For weighting purposes, each of these files except the biomarker file contains records for all sampled cases, irrespective of response and eligibility status.

### 3.2.2 Checks of Data Files

Prior to the start of the weighting process, the survey data files are checked and compared against information available in the sampling files. These checks include:

- Check IDs, merging household survey files with sampling files, and account for records found in one file and not the other. (This type of check for the EAs occurs as part of the HH selection process.)
- Check counts of sampled and responding HHs against what was expected, overall and by province.
- Acknowledge/adjust for substitution, missed HH procedures, if applicable. Check that guidelines have been followed and selection probabilities are consistent with guidelines.
- Set disposition codes (respondent, eligible nonrespondent, ineligible, unknown eligibility) to be used for weighting purposes based on data elements received for (a) all sampled households, (b) all sampled individuals, and (b) all sampled individuals for blood draws.
- Verify that the survey data, for all three components, have passed data cleaning.

## 3.3 Creation of Variables for Variance Estimation

Two general methods can be used for estimating the sampling errors of survey-based estimates derived from PHIA: the jackknife replication and Taylor's Series methods. The jackknife replication variance estimation method is a widely used method for producing variance estimates using data from a complex survey. This method can correctly account for the stratification, clustering, and sample weighting, including nonresponse and poststratification weighting adjustments, from the

PHIA complex sample design. The Taylor's Series is another widely used method that uses linear approximations to calculate the variance of a sample-derived estimate.

In order to implement either method, certain variables required for variance estimation must be included in the weighted data files. In the case of jackknife replication, the required variables are a series of weights that correspond to each of the jackknife replicates. In the case of the Taylor's Series method, the required variables are variables that indicate the "variance stratum" and the "variance unit" to which each sampled respondent belongs.

### 3.3.1 Jackknife Replication

In order to calculate variance estimates from the survey data, a series of weights, referred to as jackknife replicate weights, are attached to each record in the data file, along with the corresponding final full-sample weight. Calculation of the replicate weights first requires the construction of a set of subsamples of the full sample referred to as "jackknife replicates." Since these replicates depend only on the selected PSUs, they can be created immediately after the selection of PSUs.

As described in Section 2.3, within each province, the stratified sample of PSUs was selected systematically from a list of PSUs that had been ordered by urban/rural status, district within each urban/rural status, and finally by ward within district. To take account of the precision benefits of implicit stratification as fully as possible, the sampled PSUs were paired off in the systematic order in which they were selected, treating each pair as a variance-estimation stratum. When there was an odd number of sampled PSUs in a province, one of the variance-estimation strata was defined to contain three sampled PSUs. To fully reflect the sample design, the formation of the substrata was applied to all of the sampled PSUs, including those that may later have become a "nonresponse" (e.g., a sampled PSU containing households that was found to be inaccessible at the time of data collection) or ineligible (e.g., the PSU was found to contain no households).

For the ZIMPHIA, a total of 248 variance-estimation strata were formed. A jackknife replicate was then formed by randomly deleting a PSU from a particular variance-estimation stratum $k$, say, and retaining all of the PSUs in the remaining variance-estimation strata. The weight of the retained PSU within variance-estimation stratum $k$ was then doubled. This process was repeated for all $r = 1, 2, ...,$ 248 variance-estimation strata, resulting in a total of 248 jackknife replicates. In the case where a variance-estimation stratum consisted of three PSUs, the replicate was formed by randomly deleting one PSU in the variance-estimation stratum. In this case, the other two PSUs within the variance-

estimation stratum had their weights increased by 1.5 (see Section 3.4.1). Table 3-1 summarizes the number of jackknife replicates that were created for variance estimation.

Table 3-1 Number of PSUs and variance-estimation strata constructed for variance estimation

| Sampling Stratum (Province) | No. PSUs | No. of variance strata consisting of pairs | No. of variance strata consisting of triplets | Number of jackknife replicates |
|---|---|---|---|---|
| Bulawayo | 43 | 20 | 1 | 21 |
| Harare | 57 | 27 | 1 | 28 |
| Manicaland | 54 | 27 | 0 | 27 |
| Mashonaland Central | 56 | 28 | 0 | 28 |
| Mashonaland East | 50 | 25 | 0 | 25 |
| Mashonaland West | 52 | 26 | 0 | 26 |
| Masvingo | 53 | 25 | 1 | 26 |
| Matabeleland North | 44 | 22 | 0 | 22 |
| Matabeleland South | 40 | 20 | 0 | 20 |
| Midlands | 51 | 24 | 1 | 25 |
| Total | 500 | 244 | 4 | 248 |

### 3.3.2 Taylor's Series

Even though jackknife replication is the recommended method for variance estimation, not all software packages have a replication option to produce variance estimates. For example, SPSS has built-in options for estimating variance using Taylor's Series methods, but the end user has to write a program within SPSS to produce replicate estimates of variance. Therefore, information for producing Taylor's Series estimates of variance is included in the PHIA data files.

The full-sample weight (see Section 3.4) is used as the weight to compute Taylor's Series variance estimates. The variable **VarStrat** indicates the 248 variance-estimation strata and the variable **VarUnit** indicates the primary sampling unit (PSU) or cluster within the variance-estimation stratum. This pair of variables allows the analyst to produce variance estimates if their software does not easily accommodate replication methods, but does have a Taylor's Series capability. Note that the variance-estimation strata and the sampling strata are not equivalent: as shown in Table 3-1, the sampling strata are defined by the province and urban/rural areas, while the variance-estimation strata are based on groupings of PSUs within each sampling stratum.

## 3.4 Development of Weights

### 3.4.1 PSU Weights

The initial weighting step after the jackknife replicates were defined was to calculate PSU weights for the full sample and the replicates. Note that for convenience, we use the term PSU (primary sampling unit) to refer to either the originally-sampled EA, or the selected segment within the EA if the segmentation process was applied to the PSU.
The full-sample PSU weight was computed from the formula:

$$W_{hi}^{(1)} = 1/P_{hi}^{PSU},$$

where $P_{hi}^{PSU}$ = probability of selecting PSU *i* from province *h*. Note that if the PSU was segmented, then $P_{hi}^{PSU}$ is the product of the probability of selecting the EA and the conditional probability of selecting the segment within the EA (e.g., see Section 2.4.4). If the PSU was a replacement PSU, then $P_{hi}^{PSU}$ is the probability that the substitute PSU would have had if it had originally been selected for the sample.

Using the PSU weights defined above, the sampled PSUs (i.e., whole EAs or segments) weight up to the numbers shown in the second column of Table 3-2. However, one of the PSUs in Harare was a "nonresponding" PSU because none of its sampled dwelling units completed the household roster (see Table 2-6). To compensate for the missing PSU, the weights of the remaining PSUs in the province were adjusted by the ratio of the sum of the base weights for all sampled PSUs in Harare to the sum of the base weights for PSUs with responding households in Harare; i.e., the adjusted PSU weight was computed as

$$W_{hi}^{(1A)} = A_{hi}^{(1)} W_{hi}^{(1)},$$

where *h* denotes the province with the nonresponding PSU, $m_h$ is the number of sample PSUs in the province, $m_h^r$ is the number of responding PSUs in the province, and

$$A_{hi}^{(1)} = \sum_{i=1}^{m_h} W_{hi}^{(1)} / \sum_{i=1}^{m_h^r} W_{hi}^{(1)}$$

is the PSU weight adjustment factor. The values of $A_{hi}^{(1)}$ are shown in the next-to-last column of Table 3-2, which is equal to 1.00 for every province except Harare, where the factor is 1.019. After

applying this factor to the weights for the 56 remaining PSUs in Harare, their adjusted weights sum to the original sum of the PSU weights. The adjusted PSU weights, $W_{hi}^{(1A)}$, are passed to the household weighting process described in the next section.

As indicated in Table 3-1, 248 jackknife replicates were formed from the 500 sampled PSUs. For variance estimation, replicate-specific PSU weights, $W_{(r)hi}^{(1)}$, $r$ = 1, 2, ..., 248 were created to provide the basis for calculating the required replicate weights in subsequent stages of the weighting process. Let $h$ denote one of the 248 variance-estimation strata created for jackknife replication (Section 3.3.1) and let $i$ denote the PSU within variance-estimation stratum $h$. For a given jackknife replicate, $r$ = 1, 2, ..., 248, the corresponding replicate-specific PSU base weight was computed as

$$W_{(r)hi}^{(1)} = a\,W_{hi}^{(1)} \quad \text{if } h = r \text{ and PSU } i \text{ in variance-estimation stratum } h \text{ is included in replicate } r$$

$$= 0 \quad \text{if } h = r \text{ and PSU } i \text{ in variance-estimation stratum } h \text{ is not included in replicate } r$$

$$= W_{hi}^{(1)} \quad \text{if } h \neq r$$

where the coefficient $a$ = 2 or 1.5 depending on whether the variance-estimation stratum consisted of 2 or 3 PSUs, respectively.

The corresponding replicate-specific nonresponse-adjusted PSU weights, $W_{(r)hi}^{(1A)}$, were obtained by applying the PSU nonresponse adjustment factors in Table 3-2 to each of the replicate-specific PSU base weights, $W_{(r)hi}^{(1)}$ .

**Table 3-2 Number of PSUs and weighted sums by province, before and after adjusting for PSU nonresponse, with nonresponse adjustment factors**

| Stratum (Province) | Number of sampled PSUs (EAs) | PSUs weighted by base PSU weights [1] | Number of PSUs with responding households | PSU nonresponse adjustment factor | PSUs with resp. households weighted by adjusted PSU weights [2] |
|---|---|---|---|---|---|
| Bulawayo | 43 | 1,707.2 | 43 | 1.000 | 1,707.2 |
| Harare | 57 | 5,047.3 | 56 | 1.019 | 5,047.3 |
| Manicaland | 54 | 4,023.6 | 54 | 1.000 | 4,023.6 |
| Mashonaland Central | 56 | 2,544.9 | 56 | 1.000 | 2,544.9 |
| Mashonaland East | 50 | 3,354.4 | 50 | 1.000 | 3,354.4 |
| Mashonaland West | 52 | 3,292.0 | 52 | 1.000 | 3,292.0 |

| | | | | | |
|---|---|---|---|---|---|
| Masvingo | 53 | 3,326.1 | 53 | 1.000 | 3,326.1 |
| Matabeleland North | 44 | 1,521.1 | 44 | 1.000 | 1,521.1 |
| Matabeleland South | 40 | 1,539.6 | 40 | 1.000 | 1,539.6 |
| Midlands | 51 | 3,200.5 | 51 | 1.000 | 3,200.5 |
| Total | 500 | 29,556.6 | 499 | -- | 29,556.6 |

[1] Weights are the PSU base weights, $W_{hi}^{(1)}$.
[2] Weights are the adjusted PSU weights, $W_{hi}^{(1A)}$ .

### 3.4.2 Household Weights

#### 3.4.2.1 Household Base Weights

The household weighting process starts by calculating the household-level base weights. These are the product of the PSU weight adjusted for PSU nonresponse (described in Section 3.4.1) and the reciprocal of the within-PSU household selection probability. Thus, the household base weight for sampled dwelling unit/household *j* in PSU *i* in province *h* was computed as:

$$W_{hij}^{(2)} = W_{hi}^{(1A)} / P_{j|hi}^{HH}$$

where

$W_{hi}^{(1A)}$ = the final weight for PSU *i* in province *h* (adjusted for PSUs with no responding households)

$P_{j|hi}^{HH}$ = the conditional probability of selecting household *j* in PSU *i* in province *h*

The corresponding weights for jackknife replicate $r = 1, 2, \ldots, 248$, were computed as:

$$W_{(r)hij}^{(2)} = W_{(r)hi}^{(1A)} / P_{j|hi}^{HH},$$

where $W_{(r)hi}^{(1A)}$ is the adjusted PSU weight for PSU *i* in province *h* in replicate *r* described in Section 4.4.1.

Next, the sampled dwelling units/households were assigned to one of the four response status groups specified in Table 3-3. In Table 3-4, we show the corresponding weighted sums by response status and province using the household base weights calculated as just described. The characteristics of the household base weight were checked by examining statistical summaries of the

weights such as the mean weight, CV (coefficient of variation) of the weights, sum of the weights, minimum and maximum values of the weights, both overall and by province.

Table 3-3 Response-status groups specified for household weighting

| Household response status group [1] | Description | Number of dwelling units/households |
|---|---|---|
| 1 | Eligible respondent | 11,717 |
| 2 | Eligible nonrespondent | 2,076 |
| 3 | Ineligible/out-of-scope | 1,038 |
| 4 | Unknown eligibility status | 178 |

[1] See Appendix B for definitions.

Table 3-4. Weighted sums of household base weights by response status

| Stratum (Province) | Household Response Status | | | | |
|---|---|---|---|---|---|
| | Group 1: Eligible Respondents | Group 2: Eligible Nonresp-ondents | Group 3: Not Eligible (Vacant, Destroyed, not a DU, etc.) | Group 4: Could not determine eligibility | Weighted Count of Households [1] |
| Bulawayo | 129,134 | 28,655 | 2,753 | 1,126 | 161,668 |
| Harare | 381,842 | 119,960 | 19,061 | 5,737 | 526,599 |
| Manicaland | 341,862 | 40,125 | 49,220 | 2,942 | 434,149 |
| Mashonaland Central | 209,243 | 50,804 | 23,421 | 6,372 | 289,840 |
| Mashonaland East | 295,245 | 37,917 | 40,697 | 5,056 | 378,915 |
| Mashonaland West | 308,833 | 39,057 | 23,536 | 2,606 | 374,031 |
| Masvingo | 274,041 | 39,507 | 28,417 | 3,014 | 344,979 |
| Matabeleland North | 136,043 | 17,407 | 8,889 | 864 | 163,203 |
| Matabeleland South | 135,665 | 22,501 | 10,925 | 1,099 | 170,191 |
| Midlands | 276,912 | 53,805 | 21,800 | 1,855 | 354,373 |
| Total | 2,488,820 | 449,738 | 228,718 | 30,672 | 3,197,948 |

[1] Weights are the household base weights, $W_{hij}^{(2)}$ specified in Section 3.4.2.1.

### 3.4.2.2 Adjustment for Household Nonresponse

The general approach for handling household nonresponse was to increase the weights of responding households so that they represent the nonresponding households in the same PSU. Because such nonresponse could occur before establishing whether or not a sampled dwelling unit is eligible for the study (i.e., whether or not the household contains persons eligible for PHIA), the household nonresponse adjustment was implemented in two phases. In the first phase of

adjustment, the weights were adjusted to compensate for sampled dwelling units for which eligibility for the survey (e.g., occupancy status) was not ascertained. In the second phase of adjustment, the first-phase adjusted weights were further adjusted to compensate for the nonresponding households among those households known to be eligible for the study.

To account for variation in response rates across different types of PSUs, it is desirable to make the household nonresponse adjustments within weighting cells defined by the individual PSUs. However, if a PSU has a very low household response rate, such PSU-level adjustments can result in very large adjusted weights that would lead to increases in the variances of the survey estimates. To avoid this problem, such PSUs can be collapsed with a similar PSU to form a single non-response adjustment cell comprised of two or more PSUs. For the ZIMPHIA, a total of six PSUs were found to have response rates at or below 50% which translates to an adjustment factor at or above 2.00. To dampen the effect of the adjustment for these PSUs, each was paired with the nearest PSU on the sorted list of sample PSUs to form the final weighting cell for nonresponse adjustment. Without such collapsing, the adjustment factors would have ranged from 1.00 (for PSUs with 100% response rate) to 2.75 (for a PSU with a response rate of 36.4%). After the grouping the highest adjustment factor was reduced to 1.92.

The procedures used to compute the nonresponse-adjusted household weights are described below.

### *Phase 1 Adjustment*

As indicated above, the weighting cells for the household nonresponse adjustments are generally individual PSUs or a group of PSUs. We refer to these as "PSU weighting cells."

Let $n_{hi}^{samp}$ denote the number of sampled dwelling units in PSU weighting cell *i* in province *h*. Note that $n_{hi}^{samp}$ is the sum of the sample sizes in each of the four response status groups defined in Table 3-3, i.e.,

$$n_{hi}^{samp} = n_{hi}^{(1)} + n_{hi}^{(2)} + n_{hi}^{(3)} + n_{hi}^{(4)}$$

where

$n_{hi}^{(1)}$ = the number of responding households (i.e., households completing the roster) in PSU weighting cell *i* in province *h*

$n_{hi}^{(2)}$ = the number of eligible nonresponding households (i.e., households known to contain eligible persons but did not complete the roster) in PSU weighting cell *i* in province *h*

$n_{hi}^{(3)}$ = the number of known ineligible dwelling units (i.e., sampled dwelling units known to contain no persons eligible for the study) in PSU weighting cell *i* in province *h*

$n_{hi}^{(4)}$ = the number of sampled dwelling units for which eligibility for the study could not be ascertained in PSU weighting cell *i* in province *h*

The first-phase household nonresponse adjustment factor for PSU weighting cell *i* in province *h* was computed as the ratio:

$$A_{hi}^{(HH1)} = \sum_{j=1}^{n_{hi}^{samp}} W_{hij}^{(2)} \; / \; \sum_{j=1}^{n_{hi}^{(1)}+n_{hi}^{(2)}+n_{hi}^{(3)}} W_{hij}^{(2)}$$

where $W_{hij}^{(2)}$ is the base weight for dwelling unit/household j in PSU weighting cell *i* in province *h*, and where the sum in the numerator extends over the entire sample of dwelling units/households in PSU weighting cell *i* in province *h*, while the sum in the denominator extends over the three groups of dwelling units/households for which eligibility for the study is known.

For the sampled dwelling units/households in response-status groups 1, 2 or 3, the first-phase adjusted weight for dwelling unit/household j in PSU weighting cell *i* in province *h* was then computed as:

$$W_{hij}^{HH1} = A_{hi}^{(HH1)} W_{hij}^{(2)}$$

The corresponding replicate weights for replicate *r* = 1, 2, …, 248 were computed in similar fashion as:

$$W_{(r)hij}^{HH1} = A_{(r)hi}^{(HH1)} W_{(r)hij}^{(2)},$$

where

$$A_{(r)hi}^{(HH1)} = \sum_{j=1}^{n_{(r)hi}^{samp}} W_{(r)hij}^{(2)} \; / \; \sum_{j=1}^{n_{(r)hi}^{(1)}+n_{(r)hi}^{(2)}+n_{(r)hi}^{(3)}} W_{(r)hij}^{(2)} .$$

Note that for the sampled dwelling units/households in response-status group 4, $W_{hij}^{HH1} = W_{(r)hij}^{HH1} = 0$ for $r = 1, 2, \ldots, 248$.

The effect of this adjustment is to distribute the total weight of the undetermined-eligibility cases (i.e., the estimated 30,672 dwelling units shown in the next-to-last column of Table 3-4) to the combined weight of the remaining three groups of sampled dwelling units/households. The resulting weighted counts using $W_{hij}^{HH1}$ as computed above are given in Table 3-5.

**Table 3-5 Weighted sums of household weights adjusted for unknown eligibility**

| Province | Household Response Status | | | | |
|---|---|---|---|---|---|
| | Group 1: Eligible responding households | Group 2: Eligible nonresponding households | Group 3: Ineligible dwellings | Total dwelling units/house-holds | Total eligible households |
| Bulawayo | 129,967 | 28,924 | 2,777 | 161,668 | 158,891 |
| Harare | 385,238 | 122,037 | 19,325 | 526,599 | 507,275 |
| Manicaland | 344,174 | 40,418 | 49,557 | 434,149 | 384,592 |
| Mashonaland Central | 213,828 | 51,948 | 24,065 | 289,840 | 265,776 |
| Mashonaland East | 299,166 | 38,717 | 41,031 | 378,915 | 337,883 |
| Mashonaland West | 310,948 | 39,408 | 23,676 | 374,031 | 350,356 |
| Masvingo | 276,283 | 39,972 | 28,724 | 344,979 | 316,255 |
| Matabeleland North | 136,782 | 17,486 | 8,935 | 163,203 | 154,268 |
| Matabeleland South | 136,409 | 22,620 | 11,162 | 170,191 | 159,029 |
| Midlands | 278,345 | 54,056 | 21,972 | 354,373 | 332,401 |
| Total | 2,511,139 | 455,585 | 231,224 | 3,197,948 | 2,966,724 |

Note: Counts in table are weighted counts using first-phase adjusted household weights, $W_{hij}^{HH1}$.

### *Phase 2 Adjustment*

In the second phase of adjustment, the weights of the responding households (response status group 1) were inflated by the inverse of the (weighted) response rate in the PSU weighting cell after eliminating the known ineligible dwelling units (i.e., response-status group 3). The second-phase household nonresponse adjustment factor for PSU weighting cell *i* in province *h* was computed as the ratio:

$$A_{hi}^{(HH2)} = \sum_{j=1}^{n_{hi}^{(1)}+n_{hi}^{(2)}} W_{hij}^{HH1} \; / \; \sum_{j=1}^{n_{hi}^{(1)}} W_{hij}^{HH1}$$

where $W_{hij}^{HH1}$ is the first-phase adjusted weight for dwelling unit/household *j* in PSU weighting cell *i* in province *h*, and where the sum in the number extends over the sample of responding and nonresponding households in PSU weighting cell *i* in province *h*, while the sum in the denominator extends over the responding households.

The final nonresponse-adjusted weight for *responding* household *j* in PSU weighting cell *i* in province *h* was then computed as:

$$W_{hij}^{(2A)} = A_{hi}^{(HH2)} W_{hij}^{HH1}.$$

The corresponding replicate weights for replicate $r = 1, 2, \ldots, 248$ were computed in similar fashion as:

$$W_{(r)hij}^{(2A)} = A_{(r)hi}^{(HH2)} W_{(r)hij}^{HH1},$$

where

$$A_{(r)hi}^{(HH2)} = \sum_{j=1}^{n_{(r)hi}^{(1)}+n_{(r)hi}^{(2)}} W_{(r)hij}^{HH1} \; / \; \sum_{j=1}^{n_{(r)hi}^{(1)}} W_{(r)hij}^{HH1} .$$

The sum of the final nonresponse-adjusted household weights, $W_{hij}^{(2A)}$, summed across the responding households (response status group 1), is equal to the weighted count shown in the last column of Table 3-5.

### 3.4.3 Person-Level Interview Weights

Below, we detail the calculation of person-level base weights and nonresponse-adjusted person-level weights for analyzing the ZIMPHIA data files. Specifically, we first define the initial person-level (interview) base weights for adults, adolescents, and children in Section 3.4.3.1. Interview nonresponse adjustment using the LASSO and CHAID algorithms for variable selection is addressed in Section 3.4.3.2.

The samples for PHIA are categorized into three age groups for which different data elements are collected: (1) adults aged 15 and over, with data collected using the adult questionnaire; (2) adolescents, aged 10-14, with survey responses collected from the adolescent using an adolescent questionnaire; and (3) children aged 0-9, with survey responses provided by a parent or guardian in

the children's module of the adult questionnaire. Furthermore, some different questions are asked within the various age groups depending on the sex of the individual. All of the persons in sampled households are enumerated and placed into one of the three age categories based on the data collected in the household roster. Although all rostered adults are asked to participate in the study, only those individuals who are considered part of the *de facto* population are included in the weighting process. Adolescents and children are included in the study if they belong to the one-half subsample of households designated for child data collection.

#### 3.4.3.1 Person Base Weights

The sampled individuals were classified into three groups as indicated in Table 3-6 based on the age reported in the household roster. As discussed in Section 3.4.2.2, the starting point for developing the interview nonresponse adjustments is the final nonresponse-adjusted household weight, $W_{hij}^{(2A)}$. The sample person's base weight is the same as the nonresponse-adjusted household weight for adults (persons age 15 and over), but it is twice the nonresponse-adjusted household weight for eligible adolescents (10-14) and children (0-9) in households designated for child data collection. That is, the base weight for sample person *k* in household *j* in PSU *i* in province *h* was computed from the formula

$$W_{hijk}^{(3)} = K_k \, W_{hij}^{(2A)} ,$$

where $K_k = 1$ if the roster age of person *k* is 15 years or older, or $K_k = 2$ if the roster age of person *k* is 14 years or younger in households designated for child data collection.

The corresponding replicate base weights, $W_{(r)hijk}^{(3)}$, $r = 1, 2, \ldots, 248$, were computed in an analogous manner, with $W_{hij}^{(2A)}$ replaced by $W_{(r)hij}^{(2A)}$ in the above formula.

Table 3-6 summarizes the counts of eligible individuals by age group and interview response status, and the corresponding weighted counts using the person-level base weights, $W_{hijk}^{(3)}$. As indicated earlier in Section 2.5.3, the counts of eligible interview respondents shown in Table 3-6 includes a small number of persons who did not complete the interview but did provide an analyzable blood test.

Table 3-6 Distribution of eligible sample persons by age group and interview response status

| Group | Age [1] | Interview Status [2] | Count | Weighted count [3] |
|---|---|---|---|---|
| Adults | 15+ | Eligible Respondent | 24,723 | 6,201,864 |
| | | Eligible Nonrespondent | 2,900 | 779,681 |
| Adolescents | 10-14 | Eligible Respondent | 2,342 | 1,150,595 |
| | | Eligible Nonrespondent | 622 | 306,040 |
| Children | 0-9 | Eligible Respondent | 6,091 | 3,023,892 |
| | | Eligible Nonrespondent | 477 | 230,598 |

[1] Based on age reported in interview.
[2] Eligible respondents include cases that completed the individual interview or the blood test. See Appendix B for definitions of response status categories.
[3] Weighted by the person-level base weight, $W_{hijk}^{(3)}$.

### 3.4.3.2 Adjustment of Person Weights for Interview Nonresponse

To compensate for interview nonresponse, the person base weights were adjusted within cells defined by variables available for both the responding and nonresponding individuals. These variables included data from the household roster and other information collected in the household questionnaire, and selected PSU characteristics such as region (province) and urban/rural status. The age and sex variables used to make the nonresponse adjustments are those reported in the household roster and not the interview-reported age and sex, because the latter values are not known for the nonrespondents.

#### *The Least Absolute Shrinkage and Selection Operator (LASSO) for Initial Variable Selection*

There are approximately 50 variables from the household questionnaire and EA sampling frame that could potentially be used for nonresponse adjustment. The LASSO procedure was used for initial variable selection to reduce the number of variables to a manageable subset of the most important and relevant predictors. The LASSO is a restrictive procedure similar to linear regression that shrinks regression coefficient estimates to zero. In other words, predictors that are found to be nonsignificant have their regression coefficients set to 0 (Hastie, Tibshirani, and Friedman, 2009). The role of the LASSO is used to reduce the number of variables that would subsequently be entered into the CHAID algorithm to define the final nonresponse adjustment weighting cells.

In the final model produced by the LASSO, only the most significant variables predictive of the response variable were identified and kept. The HPGENSELECT procedure (Johnston and Rodriguez, 2015) with selection method=lasso in SAS 9.4 was used to select the variables, with the weight set to the person base weight, $W_{hijk}^{(3)}$. Separate models were fitted for the three age groups

indicated in Table 3-6. The models were selected on the basis of cross validation with observations in the input data set partitioned into disjoint subsets for model, reserving 25% for training, 50% for validation, and 25% for testing. As there is some randomness in how the LASSO selects the variables, we set the seed to a known constant value to remove the randomness so that if the program had to be re-run, the same results would be produced. Out of 50, 49, and 49 variables used in the original models for adults, adolescents, and children, respectively, the LASSO identified 28, 28, and 25 variables to be significant predictors of response for the three age groups, respectively, as indicated in Table 3-7.

### *The Chi-Square Automatic Interaction Detector (CHAID) for Cell Formation*

The next step was to apply the CHAID algorithm (Magidson, 2005) to the variables selected by the LASSO procedure. CHAID classifies the sampled individuals (i.e., the respondents and nonrespondents) into "cells" based on information available for all sample persons. The cells are formed in such a way that persons belonging to the same cell have similar propensities for being respondents. Using the variables selected by the LASSO as input, CHAID uses a weighted log-linear modeling (WLM) algorithm for the computation of chi-square statistics associated with each predictor, where the weight is the person base weight, $W_{hijk}^{(3)}$. An output of the CHAID procedure is a tree diagram that specifies the optimum number of final weighting cells, and their definitions based on the input predictor variables. The depth limit of the tree was set to 5, and the minimum subgroup size required to allow splitting and minimum terminal node size were set to 50 observations (both respondents and nonrespondents).

To create the CHAID tree for adults, gender (variable SEX) and an age-derived variable (specifically, whether the person was between the ages of 15-17 or 17+ (the derived variable H_AGETEENYEARS_C defined in Table 3-8), were forced into the model to make the initial splits. The reason for doing this was because males and females and adults 15-17 and adults 17+ received different questions; without forcing these variables into the model, the resulting tree would not have been created correctly. After forcing the two variables in the model, the tree was then allowed to grow freely. The CHAID algorithm selected 16, 11, and 10 variables for adults, adolescents, and children, respectively, that were used to create the weighting classes for nonresponse adjustment. Table 3-8 summarizes the variables that were included in the final CHAID models. The trees produced by CHAID are provided in Appendix C.

The final cells produced by CHAID were used to specify the nonresponse adjustment classes. However, cells that either had fewer than 30 respondents or had a weighted response rate of 50 percent or less, were collapsed with neighboring cells after reviewing the detailed CHAID trees. A total of 36 final weighting adjustment cells were created for adults, 21 cells for adolescents, and 16 cells for children. The final weighting cells created for nonresponse adjustment are documented in Appendix C.

Table 3-7 Variables in the original model, variables selected by LASSO, and variables selected by CHAID, and final adjustment cells

| Age Group | Variables in original model | Variables selected by the LASSO | Variables selected by CHAID | Number of nonresponse adjustment cells |
|---|---|---|---|---|
| Adults | 50 | 28 | 16 | 36 |
| Adolescents | 49 | 28 | 11 | 21 |
| Children | 49 | 25 | 10 | 16 |

Table 3-8 Variables selected by CHAID to produce classes for interview nonresponse adjustment

| Age group | Number | Variable name | Description |
|---|---|---|---|
| Adult | 1 | F_SPOUSEYN | calc - Does fname have a spouse or co-habitating partner who usually lives in the household or stayed here last night? (hidden) |
| | 2 | H_AGETEENYEARS_C | 1: 15-17; 2: Other; based on AGEYEARS (roster) |
| | 3 | H_AGEYEARS_C | Best AGEYEARS categorical |
| | 4 | H_ECON3 | Received some economic support on the past 3 months |
| | 5 | H_HAVERADTVREF_C | Household has radio, television, refrigerator |
| | 6 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 7 | H_ROOFWALFLR_C | Roof/Wall/Floor materials: Natural, metal/cement, asbestos, etc |
| | 8 | H_ROOMSLEEP_C | No. Rooms to sleep: 1, 2, 3, 4+ |
| | 9 | H_TOILETSHARENUM_C | |
| | 10 | H_TOILET_C | Toilet Shared, Not shared: Flush, Latrine, Bucket/Other |
| | 11 | H_WTRSRC | Water Source: Pipe, Tube, Well, Spring/Rain, truck/bottled, other |
| | 12 | M_SPOUSEYN | calc - Does mname have a spouse or co-habitating partner who usually lives in the household or stayed here last night? (hidden) |
| | 13 | SEX | calc - Is name Male or Female? (hidden) |
| | 14 | STRATA | Design strata |
| | 15 | SUPPORTSCHOL12 | calc - In the last 12 months, has your household received any support for kidname's schooling, such as allowance, free admission, books, or supplies, for which you did not have to pay? (hidden) |
| | 16 | URBAN_RURAL | Urban/Rural indicator: 1=Urban, 2=Rural |
| Adolescent | 1 | DADHHM | calc - Does kidname's natural father usually live in this household or was a guest last night? (hidden) |
| | 2 | H_ECON3 | Received some economic support on the past 3 months |
| | 3 | H_HAVERADTVREF_C | Household has radio, television, refrigerator |
| | 4 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 5 | H_MOMGUARD | Mother or female guardian in HH |
| | 6 | H_PARENTSICK_C | Categorical Parent Sick |
| | 7 | H_ROOMSLEEP_C | No. Rooms to sleep: 1, 2, 3, 4+ |
| | 8 | H_WATER_C | Water treated, Not treated/Water Source (given in H_WTRSRC variable) |
| | 9 | SEX | calc - Is name Male or Female? (hidden) |
| | 10 | SICKFLAGHH | calc - flag household with sick adult (hidden) |
| | 11 | STRATA | Design strata |
| Children | 1 | DADHHM | calc - Does kidname's natural father usually live in this household or was a guest last night? (hidden) |
| | 2 | DEATHS | calc - Now I would like to ask you more questions about your household. Has any usual resident of your household died since 2013? (hidden) |
| | 3 | H_ECON12 | Received some economic support on the past 12 months |

| Age group | Number | Variable name | Description |
|---|---|---|---|
| | 4 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 5 | H_MOMGUARD | Mother or female guardian in HH |
| | 6 | H_OWNSMLANIMAL_C | Household owns small animals |
| | 7 | H_OWNTRNSPRT_C | Household owns transportation |
| | 8 | H_ROOFWALFLR_C | Roof/Wall/Floor materials: Natural, metal/cement, asbestos, etc |
| | 9 | H_WATER_C | Water treated, Not treated/Water Source (given in H_WTRSRC variable) |
| | 10 | STRATA | Design strata |

### *Calculation of Nonresponse-Adjusted Person Weights*

The general approach for computing the nonresponse-adjusted person-level interview weights was as follows. Within each of the final adjustment cells, the full-sample weighted response rate, $R_m^{(int)}$, was computed as

$$R_m^{(int)} = \sum_{k=1}^{n_m^{resp}} W_{mk}^{(3)} \; / \; ( \sum_{i=1}^{n_m^{resp}} W_{mk}^{(3)} + \sum_{i=1}^{n_m^{nr}} W_{mk}^{(3)} ),$$

where m denotes the adjustment cell, $W_{mk}^{(3)}$ is the base weight for person k in cell m, $n_m^{resp}$ = the number of responding persons in cell m, and $n_m^{nr}$ = the number of eligible nonresponding persons in cell m.

The corresponding replicate-specific weighted response rates were similarly computed for jackknife replicate r = 1, 2, ..., 248 as

$$R_{(r)m}^{(int)} = \sum_{k=1}^{n_{(r)m}^{resp}} W_{(r)mk}^{(3)} \; / \; ( \sum_{i=1}^{n_{(r)m}^{resp}} W_{(r)mk}^{(3)} + \sum_{i=1}^{n_{(r)m}^{nr}} W_{(r)mk}^{(3)} ),$$

The interview nonresponse adjustment factor for cell m is $A_m^{(int)} = 1/R_m^{(int)}$ for the full sample, and $A_{(r)m}^{(int)} = 1/R_{(r)m}^{(int)}$ for jackknife replicate r = 1, 2, ..., 248.

The full-sample nonresponse-adjusted interview weight for responding person k in cell m was then computed as

$$W_{mk}^{(int)} = A_m^{(int)} W_{mk}^{(3)}$$

and the corresponding jackknife replicate weights for replicate r = 1, 2, ..., 248 were similarly computed as

$$W_{(r)mk}^{(int)} = A_{(r)m}^{(int)} W_{(r)mk}^{(3)}$$

Table 3-9 summarizes the number of weighting cells created for nonresponse adjustment, the overall weighted response rate, and the minimum and maximum adjustment for each of the three major age groups.

**Table 3-9 Characteristics of the weighting cells developed for interview nonresponse adjustment and weighted counts before and after adjustment**

| Age group | Number of Interview Respondents | Number of Adjustment Cells | Overall Weighted Response Rate | Adjustment Factor | | Weighted Count of Respondents | |
|---|---|---|---|---|---|---|---|
| | | | | Min. | Max. | Before Adjustment [1] | After Adjustment [2] |
| Adults 15 or older | 24,723 | 36 | 88.83 | 1.00 | 1.65 | 6,201,864 | 6,981,545 |
| Adolescents 10-14 | 2,342 | 21 | 78.99 | 1.00 | 2.67 | 1,150,595 | 1,456,635 |
| Children 0-9 | 6,091 | 16 | 92.91 | 1.00 | 3.69 | 3,023,892 | 3,254,490 |

[1] Weight is person base weight, $W_{mk}^{(3)}$.
[2] Weight is nonresponse-adjusted person weight, $W_{(r)mk}^{(int)}$.

### 3.4.3.3 Poststratification Adjustment

The final step in computing the individual interview weights was to adjust the nonresponse-adjusted interview weights to national population totals using a procedure called poststratification (Kalton and Kasprzyk, 1986). The primary goal of poststratification is to mitigate noncoverage biases that result when some persons in the study population do not have a chance to be sampled and interviewed. Undercoverage can occur:

- At the dwelling unit (DU) level if field operations fail to include all eligible dwelling units during the implementation of the listing procedures.
- At the household level if all households within multi-family dwelling units are not accounted for in sampling.
- At the person level where under- or overcoverage can occur if errors are made in the enumeration of household members.

To compensate for the types of coverage problems indicated above, the nonresponse-adjusted person weights were ratio-adjusted so that the resulting weighted sample counts match the population control totals indicated in Table 3-10. The population control totals given in this table are projected 2016 national population counts by gender and five-year age groups published by the Zimbabwe Statistical Office (ZIMSTAT). The post-stratified interview weights were computed as follows. Note that the poststratification adjustment was done only for the 0-59 year old age groups. Because of concerns about the stability of the pre-adjustment weighted counts for the 60-64 and 65+ year age groups, poststratification was not done for these age groups. In effect, the "poststratification adjustment" for these age groups is 1.00; i.e., the nonresponse-adjusted weights for persons in these age groups are used as the final weights for analysis.

Let $N_{ga}^{2016}$ denote the 2016 Zimbabwe population control total for gender *g* and (five-year) age group *a* as given in Table 3-10. The poststratification ratio adjustment factor for gender *g* and age group *a* was then computed for the 0-59 year age groups as:

$$T_{ga}^{2016} = N_{ga}^{2016} / \sum_{k=1}^{n_{ga}^{resp}} W_{gak}^{(int)}$$

where $W_{gak}^{(int)}$ is the nonresponse-adjusted interview weight for respondent *k* in gender group *g* and age group *a*.

The corresponding replicate-specific adjustment factors were computed in a similar way as:

$$T_{(r)ga}^{2016} = N_{ga}^{2016} / \sum_{k=1}^{n_{(r)ga}^{resp}} W_{(r)gak}^{(int)}$$

for the $r = 1, 2, \ldots, 248$ jackknife replicates.

The full-sample poststratified interview weight was then computed as:

$$W_{gak}^{(ps-int)} = T_{ga}^{2016} W_{gak}^{(int)}$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-int)} = T_{ga}^{2016} W_{(r)gak}^{(int)}$$

for $r = 1, 2, \ldots, 248$.

Weighted counts of the interview respondents before and after poststratification are summarized in Table 3-10.

Table 3-10 2016 Zimbabwe population projections (overall and by age and gender) and weighted counts before and after poststratification

| Age group | Male | | | Female | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] |
| 0-4 | 1,104,387 | 837,614 | 1.3185 | 1,128,036 | 826,777 | 1.3644 | 2,232,423 | 1,664,391 | 1.3413 |
| 5-9 | 933,376 | 794,028 | 1.1755 | 946,852 | 815,964 | 1.1604 | 1,880,228 | 1,609,992 | 1.1678 |
| 10-14 | 833,889 | 724,499 | 1.1510 | 841,379 | 717,498 | 1.1727 | 1,675,268 | 1,441,997 | 1.1618 |
| 15-19 | 824,397 | 635,488 | 1.2973 | 822,333 | 632,079 | 1.3010 | 1,646,730 | 1,267,566 | 1.2991 |
| 20-24 | 653,302 | 427,178 | 1.5293 | 687,020 | 536,295 | 1.2810 | 1,340,322 | 963,474 | 1.3911 |
| 25-29 | 521,360 | 325,545 | 1.6015 | 638,827 | 467,976 | 1.3651 | 1,160,187 | 793,521 | 1.4621 |
| 30-34 | 500,276 | 335,711 | 1.4902 | 583,120 | 465,457 | 1.2528 | 1,083,396 | 801,168 | 1.3523 |
| 35-39 | 418,493 | 295,655 | 1.4155 | 446,998 | 392,418 | 1.1391 | 865,491 | 688,073 | 1.2578 |
| 40-44 | 336,667 | 263,591 | 1.2772 | 344,564 | 314,089 | 1.0970 | 681,231 | 577,680 | 1.1793 |
| 45-49 | 238,251 | 184,858 | 1.2888 | 230,929 | 213,130 | 1.0835 | 469,180 | 397,988 | 1.1789 |
| 50-54 | 144,395 | 119,397 | 1.2094 | 182,266 | 194,971 | 0.9348 | 326,661 | 314,368 | 1.0391 |
| 55-59 | 128,507 | 119,252 | 1.0776 | 198,323 | 192,442 | 1.0306 | 326,830 | 311,694 | 1.0486 |
| 60-64 | 107,350 | 123,524 | 1.0000 [4] | 144,838 | 151,371 | 1.0000 [4] | 252,188 | 274,895 | 1.0000 [4] |
| 65+ | 227,015 | 254,359 | 1.0000 [4] | 312,433 | 331,505 | 1.0000 [4] | 539,448 | 585,864 | 1.0000 [4] |
| Total | 6,971,665 | 5,440,701 | — | 7,507,918 | 6,251,970 | — | 14,479,583 | 11,692,671 | — |

[1] Source: 2016 Zimbabwe population projections.

[2] Weighted count of interview respondents using nonresponse-adjusted interview weight, $W_{gak}^{(int)}$.

[3] Ratio of population control total to weighted count of interview respondents using nonresponse-adjusted interview weight, $W_{gak}^{(int)}$.

[4] Poststratification was not done for the 60-64 and 65+ age groups; hence, the adjustment factor is 1.00.

### 3.4.4 Person-Level Blood Test Weights

Not every interview respondent also provided a useable blood sample. Thus, a separate set of weights is required for analysis of the blood test results. Like the construction of the interview weights described previously, development of the final blood test weights involves adjustments for nonresponse and poststratification to 2016 population control totals.

#### 3.4.4.1 Initial Weights

The starting point for the construction of the blood test weights is the set of final full-sample nonresponse-adjusted interview weights and corresponding replicate weights described in Section 3.4.3.2. These weights are given by $W_{hijk}^{(int)}$ and $W_{(r)hijk}^{(int)}$ (for $r = 1, 2, \ldots, 248$), respectively, where $k$ denotes the interview respondent, $h$ denotes the province, $i$ denotes the PSU, and $j$ denotes the household. These weights have already been adjusted for interview nonresponse, and thus act as the "base" weights for developing nonresponse adjustments for the blood tests. Note that persons who provided a valid blood sample are considered to be interview respondents for the weighting purposes (e.g., see Tables 2-9A through 2-9C). Table 3-11 summarizes the counts of individuals by gender/age group and blood test response status, and the corresponding weighted counts using the person-level interview weights, $W_{hijk}^{(int)}$.

**Table 3-11 Distribution of sample persons completing the blood test by gender and age group and response status**

| Group | Age [1] | Blood Test Status [2] | Count | Weighted count [3] |
|---|---|---|---|---|
| Adult Males | 15+ | Respondent | 9,243 | 2,781,200 |
| | | Nonrespondent | 963 | 303,359 |
| Adult Females | 15+ | Respondent | 13,258 | 3,555,088 |
| | | Nonrespondent | 1,196 | 336,644 |
| Adolescent Males | 10-14 | Respondent | 1,113 | 683,405 |
| | | Nonrespondent | 63 | 41,094 |
| Adolescent Females | 10-14 | Respondent | 1,133 | 677,717 |
| | | Nonrespondent | 66 | 39,781 |
| Children | 0-9 | Respondent | 4,786 | 2,544,215 |
| | | Nonrespondent | 1,335 | 730,168 |

[1] Age reported in the interview, which may differ from the age reported on the roster.

[2] Status among the interview respondents. Persons completing the blood test are considered to be interview respondents regardless of whether a completed interview was obtained.

[3] Weighted by the person-level interview weight, $W_{hijk}^{(int)}$.

#### 3.4.4.2 Nonresponse Adjustment of Blood Test Weights

To compensate for blood test nonresponse, the person-level interview weights were adjusted within cells defined by variables available for both the responding and nonresponding individuals. These variables included data from the household roster and other information collected in the household questionnaire, and selected PSU characteristics such as region (province) and urban/rural status, and the individual interview. The age and sex variables used to make the nonresponse adjustments are those reported in the interview.

The LASSO procedure was used to identify a reduced set of predictor variables to be used in the CHAID algorithm. Out of the over 100 variables initially specified for adults and adolescents, and the 67 variables specified for children, the LASSO reduced the number of variables shown in Table 3-12. No variables were selected for the group of adolescent males.

**Table 3-12 Variables in the original model, variables selected by LASSO, and variables selected by CHAID, and final adjustment cells for blood test weights**

| Age/Sex Group | Variables in original model | Variables selected by the LASSO | Variables selected by CHAID | Number of nonresponse adjustment cells |
|---|---|---|---|---|
| Adult Male | 144 | 42 | 14 | 32 |
| Adult Female | 166 | 58 | 22 | 41 |
| Adolescent Male | 104 | 0 | 3 | 5 |
| Adolescent Female | 101 | 10 | 4 | 8 |
| Children | 67 | 43 | 32 | 56 |

Table 3-13 summarizes the variables that were included in the final CHAID models for the blood tests. As noted above, no variables were selected by the LASSO for the adolescent males. For this group, the variables used as input to the CHAID algorithm were BEST_AGE, STRATA, and URBAN_RURAL. The trees produced by CHAID are provided in Appendix C.

Table 3-13 Variables selected by CHAID to produce classes for blood test nonresponse adjustment

| Age group | Number | Variable name | Description |
|---|---|---|---|
| Adult Male | 1 | FEARTEST | Do you think people hesitate to take an HIV test because they are afraid of how other people will react if the test result is positive for HIV? |
| | 2 | HFHIVTSTOFFER | During any of your visits to the health facility in the last 12 months, did a doctor, clinical officer or nurse offer you an HIV test? |
| | 3 | H_COOKFUEL_C | Cooking Fuel: Elect., Gas, Parfin/Kerosene/coal/charcoal/wood, Other |
| | 4 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 5 | H_OWNTRNSPRT_C | Household owns transportation |
| | 6 | H_TOILET_C | Toilet Shared, Not shared: Flush, Latrine, Bucket/Other |
| | 7 | KNOWN_HIV_STATUS_R | Known HIV Status, derived variable based on the questionnaire |
| | 8 | MCPLANS | Are you planning to get circumcised? |
| | 9 | QXA1205 | Are all of the listed household members your wives/partners who live in the household? |
| | 10 | SCHLHI | What is the highest level of school you attended: primary, secondary, or higher? |
| | 11 | SERVICECLINIC | Would you prefer to receive sexual and reproductive health and HIV services together at the same clinic or separately at different clinics? |
| | 12 | STRATA | Design strata |
| | 13 | TALKBAD | Do people talk badly about people living with HIV or who are thought to be living with HIV? |
| | 14 | TBCUREHIV | Can TB be cured in people living with HIV? |
| Adult Female | 1 | AT_BESTAGE_C | BEST AGE (based on the interview age) |
| | 2 | AT_PREGNUM | Number of pregnancy, capped at 10 |
| | 3 | CERVCNTST | Have you ever been tested for cervical cancer? |
| | 4 | CNDMSEX | Do you believe women who carry condoms have sex with a lot of men? |
| | 5 | FIRSTSEXCNDM | The first time you had sex, was a condom used? |
| | 6 | HIVTSTEVER | Have you ever tested for HIV? |
| | 7 | H_COOKFUEL_C | Cooking Fuel: Elect., Gas, Parfin/Kerosene/coal/charcoal/wood, Other |
| | 8 | H_HAVERADTVREF_C | Household has radio, television, refrigerator |
| | 9 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 10 | H_OWNTRNSPRT_C | Household owns transportation |
| | 11 | H_ROOFWALFLR_C | Roof/Wall/Floor materials: Natural, metal/cement, asbestos, etc |
| | 12 | H_ROOMSLEEP_C | No. Rooms to sleep: 1, 2, 3, 4+ |
| | 13 | H_TOILET_C | Toilet Shared, Not shared: Flush, Latrine, Bucket/Other |
| | 14 | H_WTRSRC | Water Source: Pipe, Tube, Well, Spring/Rain, truck/bottled, other |
| | 15 | KNOWN_HIV_STATUS_R | Known HIV Status, derived variable based on the questionnaire |

| Age group | Number | Variable name | Description |
|---|---|---|---|
| | 16 | MCCNDMS | Do you agree or disagree with the following statement: Men who are circumcised do not need to use condoms to protect themselves from HIV |
| | 17 | PRGCARE | When you were pregnant with ${namedis}*, did you visit a health facility for antenatal care? |
| | 18 | RELIGION | What is your religion? |
| | 19 | SERVICECLINIC | Would you prefer to receive sexual and reproductive health and HIV services together at the same clinic or separately at different clinics? |
| | 20 | STDTRT | Did you get treatment for these problems? |
| | 21 | STRATA | Design strata |
| | 22 | SYPHTTK | When you were pregnant with ${namedis}*, were you tested for syphilis? |
| Adolescent Male | 1 | BEST_AGE | |
| | 2 | STRATA | Design strata |
| | 3 | URBAN_RURAL | Urban/Rural indicator: 1=Urban, 2=Rural |
| Adolescent Female | 1 | ADPLHIV | Would you play with someone who has HIV? |
| | 2 | H_OWNBIGANIMAL_C | Household owns big animals |
| | 3 | H_OWNSMLANIMAL_C | Household owns small animals |
| | 4 | STRATA | Design strata |
| Children | 1 | AT_PREGNUM | Number of pregnancy, capped at 10 |
| | 2 | AVOIDPREG | Are you or your partner currently doing something or using any method to delay or avoid getting pregnant? |
| | 3 | AWY12MOMS | In the last 12 months, have you been away from home for more than one month at a time? |
| | 4 | CH_KIDCRCMFUTR | Are you planning to have ${curchnm}** circumcised in the future? |
| | 5 | CH_KIDGENDER | Is ${curchnm}* a boy or girl? |
| | 6 | CH_KIDMISSCHL | During the last school week, did ${curchnm}* miss any school days for any reason? |
| | 7 | CH_KIDWEIGHIN12 | In the last 12 months, how often did a doctor, clinical officer or nurse weigh ${curchnm}*? |
| | 8 | CNDMSEX | Do you believe women who carry condoms have sex with a lot of men? |
| | 9 | CONDOMGET | If you wanted a condom, would it be easy for you to get one? |
| | 10 | DEATHS | calc - Now I would like to ask you more questions about your household. Has any usual resident of your household died since 2013? (hidden) |
| | 11 | FAMSHAME | Do you agree or disagree with the following statement: I would be ashamed if someone in my family had HIV. |
| | 12 | HFHIVTSTOFFER | During any of your visits to the health facility in the last 12 months, did a doctor, clinical officer or nurse offer you an HIV test? |
| | 13 | HUSOTWIF | Does your husband or partner have other wives or does he live with other women as if married? |
| | 14 | H_HAVERADTVREF_C | Household has radio, television, refrigerator |
| | 15 | H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| | 16 | H_OWNSMLANIMAL_C | Household owns small animals |

| Age group | Number | Variable name | Description |
|---|---|---|---|
| | 17 | H_OWNTRNSPRT_C | Household owns transportation |
| | 18 | H_ROOFWALFLR_C | Roof/Wall/Floor materials: Natural, metal/cement, asbestos, etc |
| | 19 | H_ROOMSLEEP_C | No. Rooms to sleep: 1, 2, 3, 4+ |
| | 20 | H_TOILET_C | Toilet Shared, Not shared: Flush, Latrine, Bucket/Other |
| | 21 | H_WATER_C | Water treated, Not treated/Water Source (given in H_WTRSRC variable) |
| | 22 | MCRISKR | Does male circumcision alone reduce the risk, or chance, of a man getting HIV completely, somewhat or not at all? |
| | 23 | MOMHHM | calc - Does kidname's natural mother usually live in this household or was a guest last night? (hidden) |
| | 24 | PRGCARE | When you were pregnant with ${namedis}*, did you visit a health facility for antenatal care? |
| | 25 | PROXY_GENDER | |
| | 26 | P_BESTAGE_C | BEST AGE (based on the interview age) |
| | 27 | RESPECT | Do people living with HIV, or thought to be living with HIV, lose the respect of other people? |
| | 28 | SERVICECLINIC | Would you prefer to receive sexual and reproductive health and HIV services together at the same clinic or separately at different clinics? |
| | 29 | SICKFLAGHH | calc - flag household with sick adult (hidden) |
| | 30 | STRATA | Design strata |
| | 31 | SYPHTTK | When you were pregnant with ${namedis}*, were you tested for syphilis? |
| | 32 | WORK12MO | Have you done any work in the last 12 months for which you received a paycheck, cash or goods as payment? |

### *Calculation of Nonresponse-Adjusted Blood Test Weights*

The general approach for computing the nonresponse-adjusted person-level blood test weights was as follows. Within each of the final adjustment cells, the full-sample weighted response rate, $R_m^{(BT)}$, was computed as

$$R_m^{(BT)} = \sum_{k=1}^{n_m^{BT}} W_{mk}^{(int)} \; / \; \left( \sum_{i=1}^{n_m^{BT}} W_{mk}^{(int)} + \sum_{i=1}^{n_m^{NBT}} W_{mk}^{(int)} \right),$$

where $m$ denotes the adjustment cell, $W_{mk}^{(int)}$ is the final interview weight for interview respondent $k$ in cell $m$, $n_m^{BT}$ = the number of interview respondents in cell $m$ who provided a useable blood sample, and $n_m^{NBT}$ = the number of interview respondents in cell $m$ who did not provide a useable blood sample.

The corresponding replicate-specific weighted response rates were similarly computed for jackknife replicate $r = 1, 2, ..., 248$ as

$$R_{(r)m}^{(BT)} = \sum_{k=1}^{n_{(r)m}^{BT}} W_{(r)mk}^{(int)} / \left( \sum_{i=1}^{BT} W_{(r)mk}^{(int)} + \sum_{i=1}^{n_{(r)m}^{NBT}} W_{(r)mk}^{(int)} \right),$$

The blood test nonresponse adjustment factor for cell $m$ is $A_m^{(BT)} = 1/R_m^{(BT)}$ for the full sample, and $A_{(r)m}^{(BT)} = 1/R_{(r)m}^{(BT)}$ for jackknife replicate $r = 1, 2, ..., 248$.

The full-sample nonresponse-adjusted interview weight for interview respondent $k$ in cell $m$ was then computed as

$$W_{mk}^{(BT)} = A_m^{(BT)} W_{mk}^{(int)}$$

and the corresponding jackknife replicate weights for replicate $r = 1, 2, ..., 248$ were similarly computed as

$$W_{(r)mk}^{(BT)} = A_{(r)m}^{(BT)} W_{(r)mk}^{(int)}$$

Table 3-14 summarizes the number of weighting cells created for nonresponse adjustment of the blood test weights, the overall weighted response rate, and the minimum and maximum adjustment for each of the five major gender/age groups.

Table 3-14 Characteristics of the weighting cells developed for blood test nonresponse adjustment and weighted counts before and after adjustment

| Group | Number of Blood Test Respondents | Number of Adjustment Cells | Overall Weighted Response Rate [1] | Adjustment Factor | | Weighted Count of Respondents | |
|---|---|---|---|---|---|---|---|
| | | | | Min. | Max. | Before Adjustment [2] | After Adjustment [3] |
| Adults 15+/Male | 9,243 | 32 | 90.17 | 1.00 | 1.65 | 2,781,200 | 3,084,559 |
| Adults 15+/Female | 13,258 | 41 | 91.35 | 1.00 | 2.55 | 3,555,088 | 3,891,732 |
| Adolescents 10-14/Male | 1,113 | 5 | 94.33 | 1.00 | 1.13 | 683,405 | 724,499 |
| Adolescents 10-14/Female | 1,133 | 8 | 94.46 | 1.00 | 1.16 | 677,717 | 717,498 |
| Children 0-9 | 4,786 | 56 | 77.70 | 1.00 | 2.42 | 2,544,215 | 3,274,383 |

[1] Among the interview respondents.
[2] Weight is person interview weight, $W_{mk}^{(int)}$.
[3] Weight is nonresponse-adjusted blood test weight, $W_{(r)mk}^{(BT)}$.

### 3.4.4.3 Poststratification Adjustment

Like the nonresponse-adjusted interview weights described previously, the nonresponse-adjusted blood test weights were poststratified to projected 2016 population counts within classes defined by gender and five-year age groups for persons 0-59 years old. Poststratification was not done for the 60-64 and 65+ year age groups. In effect, the "poststratification adjustment" for these age groups is 1.00; i.e., the nonresponse-adjusted blood test weights for persons in these age groups are used as the final weights for analysis.

Let $N_{ga}^{2016}$ denote the 2016 Zimbabwe population control total for gender *g* and (five-year) age group *a* as given in Table 3-15. The poststratification ratio adjustment factors used to adjust the blood test weights was computed for the 0-59 year age groups as:

$$T_{ga}^{2016} = N_{ga}^{2016} / \sum_{k=1}^{n_{ga}^{BT}} W_{gak}^{(BT)}$$

where $W_{gak}^{(BT)}$ is the nonresponse-adjusted blood test weight for blood test respondent *k* in gender group *g* and age group *a*.

The corresponding replicate-specific adjustment factors were computed in a similar way as:

$$T_{(r)ga}^{2016} = N_{ga}^{2016} / \sum_{k=1}^{n_{(r)ga}^{BT}} W_{(r)gak}^{(BT)}$$

for the $r = 1, 2, \ldots, 248$ jackknife replicates.

The full-sample poststratified blood test weight was then computed as:

$$W_{gak}^{(ps-BT)} = T_{ga}^{2016} W_{gak}^{(BT)}$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-BT)} = T_{ga}^{2016} W_{(r)gak}^{(BT)}$$

for $r = 1, 2, \ldots, 248$.

Weighted counts of the blood test respondents before and after poststratification are summarized in Table 3-15.

Table 3-15 2016 Zimbabwe population projections (overall and by age and gender) and weighted counts of blood test respondents before and after poststratification

| Age group | Male | | | Female | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] | Population control total [1] | Wtd. count before post-stratification [2] | Post-stratification adjustment factor [3] |
| 0-4 | 1,104,387 | 817,510 | 1.3509 | 1,128,036 | 819,704 | 1.3762 | 2,232,423 | 1,637,214 | 1.3636 |
| 5-9 | 933,376 | 819,914 | 1.1384 | 946,852 | 817,255 | 1.1586 | 1,880,228 | 1,637,169 | 1.1485 |
| 10-14 | 833,889 | 724,499 | 1.1510 | 841,379 | 717,498 | 1.1727 | 1,675,268 | 1,441,997 | 1.1618 |
| 15-19 | 824,397 | 650,400 | 1.2675 | 822,333 | 644,903 | 1.2751 | 1,646,730 | 1,295,304 | 1.2713 |
| 20-24 | 653,302 | 423,207 | 1.5437 | 687,020 | 536,340 | 1.2809 | 1,340,322 | 959,547 | 1.3968 |
| 25-29 | 521,360 | 329,233 | 1.5836 | 638,827 | 465,544 | 1.3722 | 1,160,187 | 794,777 | 1.4598 |
| 30-34 | 500,276 | 326,551 | 1.5320 | 583,120 | 467,157 | 1.2482 | 1,083,396 | 793,708 | 1.3650 |
| 35-39 | 418,493 | 291,594 | 1.4352 | 446,998 | 388,433 | 1.1508 | 865,491 | 680,027 | 1.2727 |
| 40-44 | 336,667 | 261,393 | 1.2880 | 344,564 | 313,488 | 1.0991 | 681,231 | 574,882 | 1.1850 |
| 45-49 | 238,251 | 185,085 | 1.2872 | 230,929 | 210,522 | 1.0969 | 469,180 | 395,607 | 1.1860 |
| 50-54 | 144,395 | 118,731 | 1.2162 | 182,266 | 197,946 | 0.9208 | 326,661 | 316,677 | 1.0315 |
| 55-59 | 128,507 | 118,594 | 1.0836 | 198,323 | 194,982 | 1.0171 | 326,830 | 313,575 | 1.0423 |
| 60-64 | 107,350 | 124,643 | 1.0000 [4] | 144,838 | 150,187 | 1.0000 [4] | 252,188 | 274,830 | 1.0000 [4] |
| 65+ | 227,015 | 255,128 | 1.0000 [4] | 312,433 | 322,229 | 1.0000 [4] | 539,448 | 577,357 | 1.0000 [4] |
| Total | 6,971,665 | 5,446,483 | — | 7,507,918 | 6,246,188 | — | 14,479,583 | 11,692,671 | — |

[1] Source: 2016 Zimbabwe population projections.

[2] Weighted count of blood test respondents using nonresponse-adjusted blood test weight, $W_{gak}^{(BT)}$.

[3] Ratio of population control total to weighted count of blood test respondents using nonresponse-adjusted blood test weight, $W_{gak}^{(int)}$.

[4] Poststratification was not done for the 60-64 and 65+ age groups; hence, the adjustment factor is 1.00.

# 4 Special Purpose Weights

In addition to the analytic weights described in Section 3, four sets of special purpose weights were created for analysis of specific sections of the individual questionnaire. The four sections of interest are (a) the violence module (VM), (b) the HIV knowledge (HIVK) module, (c) a module on the use of computer-assisted self interview (CASI), and (d) weight and height measurements for children. Special weights are required for analyses of these sections because the relevant modules were administered to different random subsamples of the interview respondents.

## 4.1 Weights for Analysis of the Violence Module

The violence module (VM) was administered to a random sample of women 15+ years of age. The module does not apply to men 15+ years of age nor to children 0-14 years of age.

### 4.1.1 Selection Criteria for the Violence Module

One eligible adult female aged 15+ years old was randomly selected per household to respond to the questions in the violence module. The criteria used to identify persons eligible for the violence module are given in Appendix D.

### 4.1.2 Definition of Response Status for the Violence Module

For adult females who were designated to receive the violence module, their violence respondent status is based on whether they answered key questions within the violence module. For weighting purposes, respondents are defined to be those women who (a) provided a VALID response to all four "how many times" questions, or (b) provided a VALID response to the VLNC question (see Appendix D). This definition results in an unweighted response rate of 94.9% (9,713/10,231). Table 4-1 summarizes the number of responses to the five key adult violence questions.

Table 4-1 Distribution of responses to five key variables in the violence module.

| TOUCHTIMES | CMPLSXTIMES | FRCSXTIMES | PRSSXTIMES | VLNC | Frequency |
|---|---|---|---|---|---|
| Missing | Missing | Missing | Missing | Missing | 514 |
| Missing | Missing | Missing | Missing | Invalid | 2 |
| Missing | Missing | Missing | Missing | Valid | 28 |
| Missing | Missing | Missing | Valid | Valid | 4 |
| Missing | Missing | Valid | Valid | Valid | 4 |
| Missing | Valid | Missing | Missing | Valid | 1 |
| Missing | Valid | Valid | Missing | Valid | 2 |
| Missing | Valid | Valid | Valid | Invalid | 1 |
| Missing | Valid | Valid | Valid | Valid | 17 |
| Valid | Missing | Missing | Missing | Valid | 5 |
| Valid | Missing | Missing | Valid | Valid | 5 |
| Valid | Missing | Valid | Valid | Valid | 4 |
| Valid | Valid | Missing | Missing | Valid | 1 |
| Valid | Valid | Missing | Valid | Valid | 9 |
| Valid | Valid | Valid | Missing | Invalid | 1 |
| Valid | Valid | Valid | Missing | Valid | 1 |
| Valid | Valid | Valid | Valid | Invalid | 12 |
| Valid | Valid | Valid | Valid | Valid | 9,620 |

### 4.1.3 Construction of Weights for the Violence Module

The following steps were implemented to construct the violence weights.

- Each eligible woman 15+ years of age who was selected for the violence module was assigned an appropriate base weight, $W_{jk}^{viol-bw}$, reflecting the probability of selection for the violence module, as follows:

$$W_{jk}^{viol-bw} = W_{jk}^{bw} N_j^F,$$

where $N_j^F$ = the number of eligible women 15+ in household *j* (based on roster) if there were four or less eligible women in the household or $N_j^F = 4$ it there were five or more eligible women in the household, and where $W_{jk}^{bw}$ is the corresponding base weight from the regular weighting process (see Section 3.4.3.1). The number of eligible women in the household used to compute the violence module weight was top-coded to a value of four as a way to prevent the creation of large person weights in households with a large number of eligible respondents. The small bias introduced by top coding is mitigated by the poststratification adjustment described below. The top-coded value was determined by examining the design effects and the bias and variance trade-offs of estimates of the total population using nonresponse-adjusted weights based on different top-coded values.

- Next, the response-status for persons selected for the violence module was assigned as described in Section 4.2. Note that respondents to the violence module also completed the regular interview.
- A CHAID analysis was then applied to the sample of persons selected for the violence module, separately by sex, using the same predictors identified for the regular interview weights (see Table 3-8).
- The final cells identified from the CHAID analysis were used to compute the nonresponse-adjusted weights for the violence module, $W_{jk}^{viol-nr} = A_{jk}^{nradj} W_{jk}^{viol-bw}$.
- The last step was to poststratify the $W_{jk}^{viol-nr}$s to appropriate population counts by detailed age groups for the population of 15+ year old females.

Table 4-2 lists the variables that were used to create the nonresponse-adjustment cells for creating the violence weights. Table 4-3 summarizes selected unweighted and weighted counts associated with the VM weighting process.

**Table 4-2 List of variables identified by CHAID**

| NAME | LABEL |
|---|---|
| H_AGETEENYEARS_C | 1: 15-17; 2: Other; based on AGEYEARS (roster) |
| H_AGEYEARS_C | Best AGEYEARS categorical |
| H_ECON3 | Received some economic support on the past 3 months |
| H_HAVERADTVREF_C | Household has radio, television, refrigerator |
| H_HH_SIZE_C | 1-9, where 9 includes all HHs with 9 or more rostered eligible people |
| H_MATWALL | RECODED MATEXWALLS |
| H_OWNBIGANIMAL_C | Household owns big animals |
| H_POWER_C | Power: Electricity, Solar energy, Battery, No Power |
| H_ROOFWALFLR_C | Roof/Wall/Floor materials: Natural, metal/cement, asbestos, etc |
| H_ROOMSLEEP_C | No. Rooms to sleep: 1, 2, 3, 4+ |
| H_TOILET_C | Toilet Shared, Not shared: Flush, Latrine, Bucket/Other |
| H_WTRSRC | Water Source: Pipe, Tube, Well, Spring/Rain, truck/bottled, other |
| STRATA | Design strata |

**Table 4-3 Selected statistics on the creation of the weights for the violence module**

| Age group | Number selected for violence module | Base - weighted count of persons selected for violence module | Number of respondents | Base - weighted count of respondents to violence module | Weighted count of respondents after nonresponse adjustment | Weighted count of respondents after post-stratification |
|---|---|---|---|---|---|---|
| Females 15-49 | 7,807 | 3,013,177 | 7,425 | 2,815,450 | 3,013,996 | 3,753,791 |
| Females 50+ | 2,424 | 851,218 | 2,288 | 792,683 | 850,399 | 863,465 |
| Total | 10,231 | 3,864,395 | 9,713 | 3,608,133 | 3,864,395 | 4,617,256 |

## 4.2 Weights for Analysis of the HIV Knowledge Module

The HIV Knowledge (HIVK) module was administered to a random sample of adults 15+ years of age. The adolescent version of HIV Knowledge module was administered to children 10-14 years of age. Since all adolescents were required to respond to this module, no separate HIVK weights were produced for adolescents. The module does not apply to children 0-9 years of age.

### 4.2.1 Selection Criteria for the HIV Knowledge Module

Each adult 15+ years of age had an independent probability of selection of 50% for the HIVK module, regardless of the number of other adults in the household. The criteria used to identify persons eligible for the HIVK module are given in Appendix E.

### 4.2.2 Definition of Response Status for the HIV Knowledge Module

For weighting purposes, respondents are those individuals selected for HIVK with a valid answer to the first HIVK question, ONEPARTNR ("Can the risk of HIV transmission be reduced by having sex with only one uninfected partner who has no other partners?"). The valid answers are "Yes = 1", "No = 2", and "Don't Know = 3". The answer "Refused = -9" is considered invalid, i.e., nonresponse. Of the 12,295 adults (15+) who were respondents to the individual interview and were selected for the HIVK module, 12,291 (99.97%) are HIVK "respondents" under the above definition. Table 4-4 summarizes the number of responses to key HIVK variables (see Appendix E for descriptions of the variables).

Table 4-4 Distribution of responses to key variables in the HIVK module

| Variable Name | Total (# cases =12,295) [1] | | Male (# cases = 5,086) [1] | | Female (# cases = 7,209) [1] | |
|---|---|---|---|---|---|---|
| | # with valid answer | Unwtd RR | # with valid answer | Unwtd RR | # with valid answer | Unwtd RR |
| ONEPARTNR | 12,291 | 100% | 5,084 | 100% | 7,207 | 100% |
| MOSQUITO | 12,290 | 100% | 5,084 | 100% | 7,206 | 100% |
| CONDOMS | 12,289 | 100% | 5,085 | 100% | 7,204 | 100% |
| SHAREFOOD | 12,291 | 100% | 5,085 | 100% | 7,206 | 100% |
| HEALTHYINF | 12,292 | 100% | 5,085 | 100% | 7,207 | 100% |
| BUYFOOD | 12,291 | 100% | 5,085 | 100% | 7,206 | 100% |
| KIDSSCHOOL | 12,292 | 100% | 5,085 | 100% | 7,207 | 100% |
| FEARTEST | 12,286 | 100% | 5,085 | 100% | 7,201 | 100% |
| TALKBAD | 12,290 | 100% | 5,083 | 100% | 7,207 | 100% |
| RESPECT | 12,292 | 100% | 5,084 | 100% | 7,208 | 100% |
| SALIVA | 12,289 | 100% | 5,085 | 100% | 7,204 | 100% |
| FAMSHAME | 12,285 | 100% | 5,082 | 100% | 7,203 | 100% |

{1} Counts are of individuals 15+ years of age who were selected for the HIVK module.

### 4.2.3 Construction of Weights for the HIV Knowledge Module

The following steps were implemented to construct the HIVK weights.

- Each eligible person 15+ years of age who was selected for the HIVK module was assigned a base weight, $W_{jk}^{HIVK(bw)}$, reflecting the probability of selection for the HIVK module, as follows:

$$W_{jk}^{HIVK(bw)} = 2\, W_{jk}^{(int)},$$

  where $W_{jk}^{(int)}$ is the corresponding nonresponse-adjusted interview weight from the regular weighting process (see Section 3.4.3.1).

- To reduce the variability of the weights which can lead to inflated sampling variances, an adjustment known as "weight trimming" was applied to the $W_{jk}^{HIVK(bw)}$s. The same trimming rules described in Sections 3.4.3.3 and 3.4.4.3 were applied. As shown in Table 4-5, the weight of one female respondent 15-49 years of age was trimmed.

- Because nonresponse to the HIVK module among those individuals completing the regular interview was trivial (0.03%), the final step was to poststratify the trimmed weights $W_{jk}^{HIVK(trim)}$s to appropriate population counts using procedures similar to those described in Section 3.4.3.4.

Table 4-5 summarizes selected unweighted and weighted counts associated with the HIVK weighting process.

**Table 4-5 Selected statistics on the creation of the weights for the HIV knowledge module**

| Sex/age group [1] | Number selected for HIVK module | Base - weighted count of persons selected for HIVK module | Number of HIVK respondents | Base - weighted count of HIVK respondents | Number of HIVK respondents trimmed | Weighted count of HIVK respondents | |
|---|---|---|---|---|---|---|---|
| | | | | | | after trimming | after post-stratification |
| Females 15-49 | 5,519 | 3,005,115 | 5,517 | 3,002,893 | 1 | 3,002,228 | 3,753,791 |
| Females 50+ | 1,690 | 883,962 | 1,690 | 883,962 | . | 883,962 | 863,465 |
| Males 15-49 | 4,020 | 2,463,566 | 4,019 | 2,463,157 | . | 2,463,157 | 3,492,746 |
| Males 50+ | 1,066 | 606,597 | 1,065 | 605,961 | . | 605,961 | 650,786 |
| Total | 12,295 | 6,959,240 | 12,291 | 6,955,973 | 1 | 6,955,309 | 8,760,787 |

[1] Sex and age are based on household roster data except for the post-stratified weighted counts in the last column of table. For the latter, sex and age are based on interview responses.

## 4.3 Weights for Analysis of the Computer Assisted Self Interview (CASI) Module

The Computer Assisted Self Interview (CASI) module was administered to a random sample of adults 15-49 years of age. The purpose of this module was to obtain information on how the use of this interviewing technique would affect data collection if offered in future surveys. The module does not apply to persons 50+ years of age nor to children 0-14 years of age.

### 4.3.1 Selection Criteria for the CASI Module

Among the over 15,000 households sampled for ZIMPHIA, 2,005 households were randomly selected to provide a male person 15-49 years of age to respond to questions in the CASI module. A separate (non-overlapping) random sample of 1,999 households was selected to provide a female person 15-49 years of age to respond to the CASI module. However, not all of the designatated households yielded persons eligible to receive the CASI module. For example, some households were not respondents to the survey, and some did not contain a person 15-49 years old of the designated gender. Within each of the male designated households, one male 15-49 years of age was randomly selected for the CASI module. Similarly, within each female designated household, one female 15-49 years of age was randomly selected for the CASI module. Note that households with no eligible persons of the specified gender were out-of-scope for the CASI module.

### 4.3.2 Definition of Response Status for the CASI Module

Persons 15-49 years old who were selected for the CASI module are identified in the PHIA data files by the variable CASI_FLAG, which takes on the value of 1 for the selected individuals (and 0 for the non-selected individuals). The selected individuals are those who completed the main PHIA interview and were randomly selected to answer the CASI module. Table 4-6 shows a cross tabulation of adults 15 to 49 years of age by (main) interview response status and CASI_FLAG. As indicated in the table, 2,055 eligible adults were selected for the CASI module. Of these, 1,843 were respondents to the individual interview. The data set for weighting the CASI module thus consists of these 1,843 cases.

**Table 4-6 Distribution of persons 15-49 years of age by interview response status and CASI selection status**

| Interview response status (INDIV_STATUS) | CASI selection flag (CASI_FLAG) | | | TOTAL |
|---|---|---|---|---|
| | Selected and consistent gender (1) | Selected but inconsistent gender (1) | Not selected (0) | |
| Respondent (Status 1) | 1,843 | 20 | 17,247 | 19,110 |
| Nonrespondent (Status 2) | 212 | 0 | 2,186 | 2,398 |
| TOTAL (adults 15-49) | 2,055 | 20 | 19,670 | 21,508 |

Table 4-7 summarizes the number of responses to key CASI questions. Among the 1,843 interview respondents who were selected for the CASI module, the CASI respondents are those with valid answers to the first CASI question, CSOLDLB ("How old were you at your last birthday?"). With this definition, 1,722 (93.4%) of the 1,843 persons selected for the CASI module are CASI "respondents." Table 4-8 summarizes the distribution of the persons selected for the CASI module by sex and response status. Appendix F provides additional details about the CASI response status variable.

Table 4-7 Distribution of responses to key questions in the CASI module

| Zimbabwe ADULT CASI variable analysis<br>15 <= CONFAGEY_RECODE <= 49 & INDIV_STATUS = 1 & CASI_FLAG = 1 & BEST_GENDER match CASI Flag | | | | | | |
|---|---|---|---|---|---|---|
| Variable Name | Total (# cases = 1,843) | | Male (# cases = 787) | | Female (# cases = 1,056) | |
| | # with valid answer | Unwtd RR | # with valid answer | Unwtd RR | # with valid answer | Unwtd RR |
| CSOLDLB | 1,722 | 93.43% | 732 | 93.01% | 990 | 93.75% |
| CSOLDLBDKS [1] | | | | | | |
| CSHSAPSH | 1,707 | 92.62% | 720 | 91.49% | 987 | 93.47% |
| CSWORKMO | 1,721 | 93.38% | 728 | 92.50% | 993 | 94.03% |
| CSMRLIVETOG | 1,722 | 93.43% | 730 | 92.76% | 992 | 93.94% |
| CSDRATYDAY | 1,666 | 90.40% | 698 | 88.69% | 968 | 91.67% |
| (CSDRATYDAY > 1) | 322 | | | | | |
| CSHMDRATYDAY | 292 | | 249 | | 43 | |
| CSHODROOCA | 306 | | 255 | | 51 | |
| CSHODSEX | 1,449 | 78.62% | 579 | 73.57% | 870 | 82.39% |
| CSHODSEXFT "never had sex" | 242 | | 135 | | 107 | |
| CSPPOSWP | 1,425 | | 566 | | 859 | |
| CSPPOSWPM [1] | - | | - | | - | |
| (CSPPOSWP > 1) | 1,110 | | 431 | | 679 | |
| CSLTRIMHPSW | 1,101 | | 426 | | 675 | |
| CSLTRIMHPSWU [1] | - | | - | | - | |
| CSXPHMSC | 978 | | 371 | | 607 | |
| CSXPHMSCTI [1] | - | | - | | - | |
| CSLTUSECU | 1,104 | | 428 | | 676 | |
| CSRELASUOWAY | 1,074 | | 415 | | 659 | |
| CSILUSSM | 1,101 | | 426 | | 675 | |
| CSILHUPMFS | 1,102 | | 426 | | 676 | |
| (CSILHUPMFS = Yes) | 22 | | 16 | | 6 | |
| CSILTHIMHPUP | 22 | | 16 | | 6 | |
| CSILTHIMHPTHS [1] | - | | - | | - | |
| CSXPSWCEOTI | 22 | | 16 | | 6 | |
| CSXPSWCEOTIME [1] | - | | - | | - | |
| CSBVISITHOTO | 1,722 | 93.43% | 729 | 92.63% | 993 | 94.03% |
| (CSBVISITHOTO = Yes) | 1,215 | | | | | |
| CSMAYHIVTESTM | 1,090 | | 383 | | 707 | |
| CSMAYHIVTESTY | 1,170 | | 417 | | 753 | |
| CSHIVRESULT | 1,201 | | 422 | | 779 | |
| (CSHIVRESULT = Positive) | 187 | | 58 | | 129 | |
| CSACTARV | 187 | | 58 | | 129 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| CSDOHOBYWIFE | 1,711 | 92.84% | 727 | 92.38% | 984 | 93.18% |
| CSDORELATIONWIVE | 1,705 | 92.51% | 723 | 91.87% | 982 | 92.99% |
| CSDPDINVIEW | 1,674 | 90.83% | 709 | 90.09% | 965 | 91.38% |
| CSAMOREPRIVATE | 1,669 | 90.56% | 704 | 89.45% | 965 | 91.38% |
| | | | | | | |

[1] Follow-up questions if the previous questions were left blank, asking why left blank ("-8 = "Don't know" and -9 "I want to skip to the next question" are not counted as valid responses).

**Table 4-8 Distribution of persons selected for the CASI module by sex and response status**

| CASI response status | Sex [1] | | Total |
|---|---|---|---|
| | 1 (Male) | 2 (Female) | |
| 1 (respondent) | 732 | 990 | 1,722 |
| 2 (nonrespondent) | 55 | 66 | 121 |
| Total | 787 | 1,056 | 1,843 |

[1] Sex and age are based on household roster data.

### 4.3.3 Construction of Weights for the CASI Module

The following steps were implemented to construct weights for analysis of individuals who were selected for and asked to complete a small number of survey items using a CASI instrument. Since the primary objective was to estimate the impact that the CASI would have on both participation in PHIA and the resulting quality of data provided by those who completed the CASI module, weights were computed for all selected individuals (both respondents and nonrespondents).

- First, we identified the set of individuals for whom a final (positive) trimmed nonresponse-adjusted person-level *interview* weight, $W_{jk}^{NR}$, had been computed for the first report (i.e., these are the regular nonresponse-adjusted interview weights described in Section 3.4.3.2) where *j* denotes the household and *k* denotes the individual within the household.
- Next, we identified the subset of interview respondents who were selected for the CASI module. These are cases for which CASI_FLAG = 1.
- We assigned a CASI "base" weight to the *k*th person who was sampled for the CASI module as

$$W_{jk}^{BW:CASI} = (15009/2000) * W_{jk}^{NR} N_j^{15-49},$$

where $N_j^{15-49}$ = the number of "gender-eligible" rostered individuals 15 - 49 years old in household *j*.

The factor (15009/2000) = 7.5045 reflects the fact that an expected 2,000 (randomly selected) households were designated for CASI interviews for men, and another 2,000 (non-overlapping) households were designated for CASI interviews for women. The factor inflates the interview weights to adjust for the subsampling of households for the CASI. In the male-designated households, $N_j^{15-49}$ = the number of rostered males 15 - 49 years of age, whereas in female-designated households, $N_j^{15-49}$ = the number of rostered females 15 - 49 years of age.

- Finally, we post-stratified the $W_{jk}^{BW:CASI}$s of the individuals who were selected for the CASI module by sex and age group to obtain the final CASI weight, $W_{jk}^{CASI}$.

Table 4-9 summarizes selected unweighted and weighted counts associated with the CASI weighting process.

**Table 4-9 Selected statistics on the creation of the weights for the CASI module**

| Sex/age group [1] | Number selected for CASI module | Base -weighted count of persons selected for CASI module | Weighted count of persons selected for CASI module after post-stratification |
|---|---|---|---|
| Adult male (15-49) | 787 | 2,314,059 | 3,492,746 |
| Adult female (15-49) | 1,056 | 2,821,407 | 3,753,791 |
| Total | 1,843 | 5,135,466 | 7,246,537 |

[1] Sex and age are based on household roster data except for the post-stratified weighted counts in the last column of table. For the latter, sex and age are based on interview responses.

## 4.4 Weights for Analysis of Children's Weight and Height Measurements

A subsample of children 0-60 months of age was selected to obtain weight and height measurements for a nutritional assessment.

### 4.4.1 Selection Criteria for the Weight and Height Measurements

All children 0-60 months of age who tested HIV positive and a random sample of approximately 5 percent of children 0-60 months of age who tested HIV negative were selected for the weight and height measurements.

### 4.4.2 Definition of Response Status for the Weight and Height Measurements

Table 4-10 summarizes the distribution of children 0-60 months old for whom a blood test weight had been computed by the standard PHIA weighting procedures described in Section 3.4.4 by (a) HIV testing status (HIVSTATUS, HIVSTATUSC), (b) weight/height measurement selection status (CWH_FLAG), and (c) the presence or absence of reported height (CWHHEIGHT) and weight (CWHWEIGHT). The number of cases to be weighted are shown in the last column of the table, and are those for which CWH_FLAG = 1 and for which the weight and height measurements are not both missing. Additional details about the creation of the response status variable is given in Appendix G.

**Table 4-10 Distribution of children 0-60 months old with a blood test weight by HIV test result and selection status**

| HIVSTATUS[1] (1 = pos.; 2 = neg.) | HIVSTATUSC[2] (1 = pos.; 2 = neg.) | CWH_FLAG 1=selected; 0 = not. sel. | CWHHEIGHT | CWHWEIGHT | Cases [3] with a blood test weight | Cases to weighted for weight and height analysis |
|---|---|---|---|---|---|---|
| . | 1 | 1 | NON-MISS | NON-MISS | 33 | 33 |
| . | 2 | 0 | MISS | MISS | 575 | 0 |
| . | 2 | 1 | NON-MISS | NON-MISS | 34 | 34 |
| 1 | . | 1 | NON-MISS | NON-MISS | 15 | 15 |
| 2 | . | 0 | MISS | MISS | 1,737 | 0 |
| 2 | . | 1 | NON-MISS | NON-MISS | 82 | 82 |
| TOTAL | — | — | — | — | 2,476 | 164 |

[1] HIVSTATUS is the HIV result variable for children who are older than 18 months.

[2] HIVSTATUSC is the HIV result variable for children 18 months or younger.

[3] Children with a confirmed age of 0-60 months for whom a blood test was previously computed (see Section 3.4.4)

### 4.4.3 Construction of Weights for the Weight and Height Measurements

The basic steps for creating the analytic weights required for analysis of the weight and height measurements were as follows:

- A "base" weight, $W_i^{WH:base}$, was assigned to those cases with CWH_FLAG = 1 as follows:

$$W_i^{WH:base} = K W_i^{BT}$$

where $W_i^{BT}$ is the final blood test weight for child *i* (see Section 3.4.4) and

*K* = 1 if the child tested HIV positive;

*K* = 20 if the child tested HIV negative, was selected for weight and height measurements, and the reported weight and height measurements were not both missing.

From Table 4-10, it can be seen that 164 cases in Zimbabwe were included in the weighting process. Note that since all the sampled children provided weight and height measurements, a separate nonresponse adjustment was not done.

- Next, the base weights, $W_i^{BT}$, were poststratified so that the final weighted counts match the corresponding full-sample weighted counts by gender.

  Specifically, let $W_{gi}^{WH:PS}$ denote the final weight for child *i* of gender *g*. Then $W_{gi}^{WH:PS}$ was computed as:

  $$W_{gi}^{WH:PS} = W_{gi}^{WH:base} \ (A_g \ / \ B_g \ )$$

  where

  $W_{gi}^{WH:base}$ = the base weight for child *i* of gender *g* as computed above,

  $A_g$ = $\sum_{j=1}^{n_g} W_{gj}^{BT}$

  $W_{gj}^{BT}$ = the previously-computed full-sample blood test weight for child *j* of gender *g*

  $n_g$ = the number of children of gender *g* in the *full* sample for which $W_{gj}^{BT} > 0$.

  $B_g$ = $\sum_{j=1}^{n_g^{WH}} W_{gj}^{WH:base}$

  $n_g^{WH}$ = the number of children of gender *g* who were selected for and provided weight/height measurements

- The above steps were repeated for each of the jackknife replicates to provide the corresponding jackknife weights for variance estimation.

Table 4-11 summarizes selected unweighted and weighted counts associated with the weighting process.

Table 4-11 Selected statistics on the creation of the weights for children's weight and height measurements

| Sex/age group [1] | Number providing weight and height measurements (respondents) | Base-weighted count of respondents | Final (post-stratified) weighted count of respondents |
|---|---|---|---|
| Females 0-60 mos. | 91 | 1,234,499 | 1,147,947 |
| Males 0-60 mos. | 73 | 895,375 | 1,125,466 |
| Total | 164[2] | 2,129,873 | 2,273,414 |

[1] Sex and age are based on household roster data except for the post-stratified weighted counts in the last column of table. For the latter, sex and age are based on interview responses.

[2] Represents an unweighted response rate of 164/164 = 1.000 (see Table 4-10).

# References R


Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics. http://www.springer.com/us/book/9780387848570

Johnston, G. and Rodriguez, R (2015). Introducing the HPGENSELECT Procedure: Model Selection for Generalized Linear Models and More. Paper SAS1742-2015. https://support.sas.com/resources/papers/proceedings15/SAS1742-2015.pdf

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology 12*, 1-16.

Kish, L. (1965). *Survey Sampling*. New York, NY: John Wiley & Sons.

Magidson, J. (2005) SI-CHAID Users Guide. Statistical Innovations. https://www.statisticalinnovations.com/wp-content/uploads/SICHAIDusersguide.pdf

Valliant, R., Dever, J., & Kreuter, F. (2013). Practical Tools for Designing and Weighting Survey Samples. New York, NY: Springer.

# Appendix A

# Definition of Eligibility for Dwelling Unit/Household Sampling

# Definition of Eligibility for Dwelling Unit/Household Sampling

The listing process for the ZIMPHIA was done manually. The sampling frames of listed dwelling units/households were entered in 10 separate Excel files, one for each province. Each Excel file included separate worksheets ("tabs") for each of the sampled PSUs within the province.

Each worksheet contained four columns:

- Household number (generally, a sequence number for the eligible households in the PSU),
- HH To Interview (generally left blank),
- Residence (assigned as "Y" for the eligible cases), and
- Observations (notes entered by the lister indicating vacant, not at home, refuse to respond, etc.).

The Zimbabwe team assigned a household number to those dwelling units/households that were eligible for sampling purposes. These corresponded to cases where Residence = Y and included vacant or non-occupied units that could potentially be occupied at the time of interview. Table A-1 summarizes the distribution of the dwelling units/households in the listing data files by eligibility status and province.

**Table A-1 Summary of Excel files used to create the sampling frame**

| Province | Number of PSUs | Number of Households Not Eligible for Sampling | Number of Households Eligible for Sampling |
|---|---|---|---|
| BULAWAYO | 43 | 2 | 4,137 |
| HARARE | 57 | 0 | 6,129 |
| MANICALAND | 54 | 28 | 5,872 |
| MASH CENTRAL | 56 | 1 | 6,457 |
| MASH EAST | 50 | 2 | 5,738 |
| MASH WEST | 52 | 0 | 6,157 |
| MASVINGO | 53 | 1 | 5,596 |
| MATABELELAND NORTH | 44 | 0 | 4,795 |
| MATEBELELAND SOUTH | 40 | 1 | 4,429 |
| MIDLANDS | 51 | 1 | 5,712 |
| *Total* | *500* | *36* | *55,022* |

# Appendix B

# Definition of Household, Interview, and Blood Test Response Status

# Definition of Household, Interview, and Blood Test Response Status

## B.1 Survey Status for Household: HH_STATUS

Table B-1 Household response status codes (HH_STATUS)

| Value | Meaning | Comments |
|---|---|---|
| 1 | Responding household | All households with Roster records |
| 2 | Nonresponding in-scope household | Household with a record, no roster data, and judged in-scope for the survey based on the RESULTNDT or RESULTNDTOTHR variables |
| 3 | Household not in scope for the survey | Households with a record, no roster data, and judged not in-scope for the survey based on the RESULTNDT or RESULTNDTOTHR variables |
| 4 | Household with no roster data, but unable to determine whether the household was in scope for the survey | In the weighting process the base weights for these cases is distributed among the other household records. |

## SAS Code for HH_STATUS

```
 attrib HH_eligible length=3 label="Household Eligibility flag – will be used to create
HH_STATUS";
   if     STARTINT=1 and TAPGOOD=1 and RESULTNDT=" " then HH_eligible = 1;
/* Complete */
   else if STARTINT=1                          then HH_eligible = 2;          /* Partial complete */
   else if STARTINT=2 and RESULTNDT in ('3','5')     then HH_eligible = 3;          /* Eligible
NR */
   else if STARTINT=2 and RESULTNDT in ('6','7')     then HH_eligible = 4;          /* Known
Ineligible */
   else if STARTINT=2 and RESULTNDT in ('8','10')    then HH_eligible = 5;          /*
Unknown Ineligible */

   attrib HH_STATUS length=3 label="HH disposition code";
   if     HH_eligible = 1     then HH_STATUS= 1;                    /* Eligible Respondent */
   else if HH_eligible in(2,3) then HH_STATUS= 2;                    /* Eligible
NonRespondent */
   else if HH_eligible = 4     then HH_STATUS= 3;                    /* Ineligible */
```

```
   else if HH_eligible = 5      then HH_STATUS= 4;                          /* Unknown eligibility
Status */

   if HH_ELIGIBLE = 2 and ROSTERCOUNT > 0 then HH_STATUS= 1 ;                /*
Eligible Respondent */

   if HH_ELIGIBLE = 5 and UPCODE_STAT_HH in (2,3,4) then HH_STATUS =
UPCODE_STAT_HH;
```

**Notes regarding this code:**

The statement "if HH_ELIGIBLE = 2 and ROSTERCOUNT > 0 then HH_STATUS= 1" resets HH_STATUS to 1 = Eligible Respondent for "partly complete" households that have roster records. (The variable ROSTERCOUNT is created earlier in the program; it counts the number of individual records on the file phiazim_cff_roster_20161201 for each value of EA_HHID_FIXED.)

The statement "if HH_ELIGIBLE = 5 and UPCODE_STAT_HH in (2,3,4) then HH_STATUS = UPCODE_STAT_HH;" moves cases from HH_Status 4 = Unknown Eligibility Status to one of the other status codes that apply to household records with no response. (The variable UPCODE_STAT_HH is created based on the text in RESULTNDTOTHR. The DM team, the ICAP team and the statistical team all contributed to evaluating the text comments and assigning codes based on the text.)

## B.2 Survey Status for Individual Interview: INDIV_STATUS

Table B-2 Individual response status codes (INDIV_STATUS)

| Value | Meaning | Comments |
|---|---|---|
| 1 | Responding, in-scope individual | Individual from in-scope household; for children, must also be in household with ChildFlag turned on, has questionnaire data and/or biomarker data |
| 2 | Nonresponding in-scope individual | Individual from in-scope household; for children, must also be in household with ChildFlag turned on, no questionnaire data or biomarker data |
| 4 | Individual with a record, no data, but unable to determine whether the individual was in scope for the survey | reflects ambiguous "reason for no data" (4 cases) |
| 7 | Rostered in error | based on "reason for no data" (15 cases) |
| 8 | Not Sampled | Child in household with Child Flag not turned on |
| 9 | De Jure Ineligible | Slept here last night? = NO |

## SAS Code for INDIV_STATUS

"Indiv_elig_check = first check of roster information to determine eligibility of rostered person"

```
indiv_elig_check = 0;
  if in_roster = 1 and (livehere = 1 or sleephere = 1) then do;
      if ageyears >= 15 then indiv_elig_check = 1;
      else if ageyears =< 14 and child_smpflg_r = 1 then do;
          if (momfemname ^= . or dadmalename ^= .) then indiv_elig_check = 1;
      end;
```

NOTE on this piece. Cases rostered deFacto or DeJure and roster age15+ have preliminary status "Eligible"; for DeFacto and DeJure cases roster ages 0 to 14, must have child flag "yes" and a linked adult to be preliminary "eligible"

"indiv_nonelig_reason = reason for indiv_elig_check = 0"

```
if INDIV_ELIG_CHECK = 0 then do;
   if AGEYEARS >= 15 then do;
            if SLEEPHERE = 2 and
```

```
            LIVEHERE  = 1 then INDIV_NONELIG_REASON=1;
        else
         if SLEEPHERE = 2 and
            LIVEHERE  = 2 then INDIV_NONELIG_REASON=2;
end;
else if AGEYEARS < 15 then do;
            if SLEEPHERE = 2 and
                LIVEHERE  = 1 then INDIV_NONELIG_REASON=3;
            else
                if SLEEPHERE = 2 and
                    LIVEHERE  = 2 then INDIV_NONELIG_REASON=4;
            else
                    if child_smpflg_r = 2 and
                    SLEEPHERE =1 then INDIV_NONELIG_REASON=5;
            else
                    if child_smpflg_r   =1  and
                    SLEEPHERE   =1  and
                    MOMFEMNAME  = .  and
                    DADMALENAME = .    then INDIV_NONELIG_REASON=6;
 end;
```

NOTE on this piece: Cases given preliminary status "not eligible" are given a code as to why:

| INDIV_NONELIG_REASON | Value label |
|---|---|
| 1 | Adults (>=15) usually live here but didn't sleep here |
| 2 | Adults (>=15) neither live here nor slept here |
| 3 | Children (<15) usually live here but didn't sleep here |
| 4 | Children (<15) neither live here nor slept here |
| 5 | Children (<15) slept here but with child flag off |
| 6 | Children (<15) slept here, with child flag on, but had no linked guardians |

Cases in category 6 will be returned to "Eligible Nonrespondent" status later.

Create INDIV_AGEGROUP:

```
    IF CONFAGEY_RECODE ^= . THEN DO;
        BEST_AGE = CONFAGEY_RECODE;
        BEST_AGE_FLAG = 1;
```

```
END;
ELSE DO;
    BEST_AGE = AGEYEARS;
    BEST_AGE_FLAG = 2;
END;
IF GENDR ^= . THEN DO;
   BEST_GENDER = GENDR;
   BEST_GENDER_FLAG = 1;
END;
ELSE DO;
   BEST_GENDER = SEX;
   BEST_GENDER_FLAG = 2;
END;
if 0 <= BEST_AGE <= 9 then indiv_agegroup = 1;
else
    if 10 <= BEST_AGE <= 14 then indiv_agegroup = 2;
   else
       if BEST_AGE >= 15 then indiv_agegroup = 3;
```

NOTE: Section above creates INDIV_AGEGROUP based on CONFAGEY_RECODE when available, otherwise AGEYEARS.

```
indiv_qxstatus = "Completion of questionnaire";

indiv_qxstatus = 0;

if (INDIV_AGEGROUP = 1 and
   (CH_KIDAGEY       => 0 or
    CH_KIDGENDER     => 0 or
    CH_KIDENROLL     => 0 or
    CH_KIDHIVTESTEVR => 0 or
    CH_KIDWEIGHIN12 => 0 or
    CH_KIDVISTTBCLIN => 0 or
    CH_KIDDIAGTB     => 0)) or
```

```
   (indiv_agegroup = 2 and icnsnt = 1 and indfinrslt in (1, 2) and adattck in (1,2)) then
indiv_qxstatus = 1;

else
  if (indiv_agegroup = 3 and icnsnt = 1 and indfinrslt in (1, 2)) then do;
    two_flag = 0;
    do i = 1 to 12;
      if miles(i) = 2 then two_flag = 1;
    end;
   if two_flag = 0 then indiv_qxstatus = 1;

end;
```

NOTE: INDIV_QXSTATUS analyzes the relevant interview variables for each INDIV_AGEGROUP. Value INDIV_QXSTATUS = 1 indicates enough interview data to consider the interview completed. For ages 0 -9 the determination is based on the Module 3A variables from the linked adult.

```
label indiv_status   = "Individual Response Status"

indiv_status = 0;
if sleephere = 2 then indiv_status = 9;
else
  if indiv_nonelig_reason = 5 then indiv_status = 8;
  else
    if indiv_nonelig_reason = 6 then indiv_status = 2;
    else
      if in_indiv = . and indiv_elig_check = 1 then indiv_status = 2;
      else
        if hiv1statusfinalsurvey in ("Negative", "Positive") then indiv_status = 1;
        else
          if indiv_qxstatus = 1 then indiv_status = 1;
          else
            indiv_status = 2;
run;
```

NOTE: Base definition of INDIV_STATUS. IN_INDIV = . indicates a rostered case with no individual cff record.

```
 if indiv_status not in (8,9) and upcode_stat not in (.,3) and indiv_agegroup in (2,3) then
indiv_status = upcode_stat;
 else
  if indiv_status not in (8,9) and upcode_stat in (7,9) and indiv_agegroup = 1 then indiv_status =
upcode_stat;
```

NOTE:UPCODE_STAT is the recode of INDFINRSLT_DISP (the text) when IND0040 = 10 "Reason for no data: OTHER (specify)" It is used to reassign INDIV_STATUS for cases where this occurred,"

```
  If EA_HHID_LN_FIXED in ("101071007011904", "426095003000503", "804281005006202",
"921315014002506") then INDIV_STATUS = 8;
```

NOTE: Hardcode the INDIV_STATUS of four records where all roster items were missing to "not sampled".

## B.3 BTEST Survey Status for Individual Blood Test Data

**Table B-3 Blood test response status codes (BTEST)**

| Value | Meaning | Comments |
|---|---|---|
| 1 | Has blood test | Responding individuals with hiv1statusFinalSurvey with values 'Positive' or 'Negative' |
| 2 | Does not have blood test | All other responding individuals |

## SAS Code for BTEST

```
 ATTRIB BTEST LABEL="Was blood test done: 1=YES, 2=NO";
   IF HIV1statusfinalsurvey In (1,3) THEN BTEST=1;
   ELSE                                BTEST=2;
```

NOTE: HIV1statusfinalsurvey is changed to numeric when read in:

VALUE 1 = '1 - Negative'

2 = '2 - Unknown'

3 = '3 - Positive'

# Appendix C

# CHAID Trees and Definition of Final Nonresponse-Adjustment Weighting Cells

# CHAID Trees and Definition of Final Nonresponse-Adjustment Weighting Cells

## C.1 Final CHAID Trees

The final CHAID trees used to construct the weighting cells for nonresponse adjustment are documented in PDF files in the zipped file Appendix_C.zip. There are a total of eight PDF files corresponding to the three groups for which the CHAID analysis was conducted for adjustment of the interview weights (Section 3.4.3.2) and the five groups for which the CHAID analysis was conducted for adjustment of the blood test weights (Section 3.4.4.2). The names of the eight PDF files containing the CHAID trees are listed below. Each tree indicates diagrammatically how the final weighting cells were created by successively partitioning the sample into heterogeneous subsets with respect to response propensity. The final cells (prior to collapsing, if done to control variation in weights) are indicated by the number underneath the box defining the cell.

### Individual Interview

AD_INDIV_STATUS.pdf (Persons 15 years or older)

TN_INDIV_STATUS.pdf (Adolescents 10-14 years)

CH_INDIV_STATUS.pdf (Children 0-9 years)

### Blood Test

AM_BTEST.pdf (Males 15 years or older)

AF_BTEST.pdf (Females 15 years or older)

TM_BTEST.pdf (Males 10-14 years)

TF_BTEST.pdf (Females 10-14 years)

C_BTEST.pdf (Children 0-9 years)

## C.2 Final Nonresponse-Adjustment Weighting Cells

The final nonresponse-adjustment weighting cells are documented in Excel files in the zipped file Appendix_C.zip. There are eight Excel files corresponding to the groups for which the nonresponse adjustments were made. The names of the Excel files are listed below. Each row of the Excel file corresponds to a weighting cell, and shows the variables and the corresponding values used to define the weighting cell, the numbers of responding and nonresponding cases in the cell, the weighted counts of the responding and nonresponding cases, the weighted response rate, and the nonresponse weight adjustment factor (which is defined to be the reciprocal of the weighted response rate). Cells that were collapsed to control the variation in weights are highlighted.

### Individual Interview

Zim_AD_INDIV.xlsx (Persons 15 years or older)

Zim_TN_INDIV.xlsx (Adolescents 10-14 years)

Zim_CH_INDIV.xlsx (Children 0-9 years)

### Blood Test

Zim_AM_BT.xlsx (Males 15 years or older)

Zim_AF_BT.xlsx (Females 15 years or older)

Zim_TM_BT.xlsx (Males 10-14 years)

Zim_TF_BT.xlsx (Females 10-14 years)

Zim_CH_BT.xlsx (Children 0-9 years)

# Appendix D

# Violence Module Variables, Eligibility Criteria, and Program Code

# Violence Module Variables, Eligibility Criteria, and Program Code

## D.1 Variables in the Violence Module

| Variable | Question Text |
|---|---|
| vlnc | Has anyone ever done any of these things to you: - Punched, kicked, whipped, or beat you with an object - Slapped you, threw something at you that could hurt you, pushed you or shoved you - Choked, smothered, tried to drown you, or burned you intentionally - Used or threatened you with a knife, gun or other weapon? |
| vlncfrstage | How old were you the first time one of these things happened to you? |
| vlncfrstagedk | Please provide the reason this previous question was left blank: How old were you the first time one of these things happened to you? |
| vlnc12motimes | In the past 12 months, how many times did someone: - Punched, kicked, whipped, or beat you with an object - Slapped you, threw something at you that could hurt you, pushed you or shoved you<br>- Choked, smothered, tried to drown you, or burned you intentionally<br>- Used or threatened you with a knife, gun or other weapon? |
| vlnc12moptnr | In the past 12 months, did a partner do any of these things to you? |
| seekhelp | Thinking about all these experiences that we just discussed, whether someone has done the following: - Punched, kicked, whipped, or beat you with an object - Slapped you, threw something at you that could hurt you, pushed you or shoved you - Choked, smothered, tried to drown you, or burned you intentionally - Used or threatened you with a knife, gun or other weapon Did you try to seek professional help or services for any of these incidents from any of the following? |
| seekhelpwhynot | What was the main reason that you did not try to seek professional help or services? |
| touchtimes | How many times has anyone ever touched you in a sexual way without your permission, but did not try and force you to have sex? |
| touchtimesdk | Please provide the reason this previous question was left blank: How many times has anyone ever touched you in a sexual way without your permission, but did not try and force you to have sex? |
| touchage | How old were you the first time this happened? |
| touchagedk | Please provide the reason this previous question was left blank: How old were you the first time this happened? |
| touchrelat | The first time this happened, what was this person's relationship to you? If it was more than one person, what was the relationship with the person you knew the best? |
| cmplsxtimes | How many times in your life has anyone tried to make you have sex against your will but did not succeed? This includes someone using harassment, threats, tricks, or physical force. |
| cmplsxtimesdk | Please provide the reason this previous question was left blank: How many times in your life has anyone tried to make you have sex against your will but did not succeed? This includes someone using harassment, threats, tricks, or physical force. |

| Variable | Question Text |
|---|---|
| cmplsxage | How old were you the first time someone tried to make you have sex against your will but did not succeed? |
| cmplsxagedk | Please provide the reason this previous question was left blank: How old were you the first time someone tried to make you have sex against your will but did not succeed? |
| frcsxtimes | How many times in your life have you been physically forced to have sex? |
| frcsxtimesdk | Please provide the reason this previous question was left blank: How many times in your life have you been physically forced to have sex? |
| frcsxage | How old were you the first time someone physically forced you to have sex? |
| frcsxagedk | Please provide the reason this previous question was left blank: How old were you the first time someone physically forced you to have sex? |
| frcsxrelat | What was this person's relationship to you? If it was more than one person, what was the relationship with the person you knew the best? |
| frcsx12mo | In the past 12 months, did someone physically force you to have sex? |
| frcsx12mopt | In the past 12 months, did a partner physically force you to have sex? |
| prssxtimes | How many times in your life has someone pressured you to have sex through harassment, threats and tricks and did succeed? |
| prssxtimesdk | Please provide the reason this previous question was left blank: How many times in your life has someone pressured you to have sex through harassment, threats and tricks and did succeed? |
| prssxage | How old were you the first time someone pressured you to have sex and did succeed? |
| prssxagedk | Please provide the reason this previous question was left blank: How old were you the first time someone pressured you to have sex and did succeed? |
| prssxrelat | What was this person's relationship to you? If it was more than one person, what was your relationship with the person you knew the best? |
| prssx12mo | In the past 12 months, did someone pressure you to have sex and did succeed? |
| prssx12mopt | In the past 12 months, did a partner pressure you to have sex and did succeed? |
| uwntsxhelp | After any of these unwanted sexual experiences, did you try to seek professional help or services from any of the following? |
| unwntsxnohlp | What was the main reason that you did not try to seek professional help or services? |

## D.2 Eligibility Criteria for the Violence Module

The variable VM_STATUS was created to identify individuals eligible to receive the violence module and was assigned to every rostered record, with values as shown in the table below. Codes 1 through 9 were assigned only to cases flagged to receive the violence module.

| VM_STATUS | Description |
|---|---|
| 0 | Not selected for Violence Module |
| 1 | Violence Module Respondent |
| 2 | In-scope for Violence Module, Non-Respondent |
| 3 | Out of scope for Violence Module, changed to male in Interview |
| 4 | Out of scope for Violence Module, changed age out of age range for Violence Module in Interview |
| 5 | No data, unknown whether eligible for survey |
| 6 | Collected in Another Tablet |
| 7 | Rostered in Error |
| 8 | Not Sampled (adults over the age limit of participation for the country and children in households with child flag = NO) |
| 9 | Extraneous Cases – De Jure Ineligible |

## D.3 Code to Define Violence Module Status (VM_STATUS)

```
DATA HH_QX;
        LENGTH EA_HHID_VIOL $15;
        LENGTH VIOLFLAG_X $2;
        SET w11.HH_QX(KEEP=EA_HHID_FIXED CHILDFLAG VIOLFLAG);

        VIOLFLAG_X = PUT(VIOLFLAG,Z2.0);

        IF VIOLFLAG ^= . THEN DO;
                EA_HHID_VIOL = EA_HHID_FIXED||VIOLFLAG_X;
        END;
RUN;

DATA ROSTER;
        SET W11.ROSTER;
```

```
        IF AGEYEARS < 15 THEN ROSTER_VIOL_AGECAT = 1;  /* Roster age less than 15
        */
        ELSE IF AGEYEARS > 14 THEN ROSTER_VIOL_AGECAT = 2;  /* Roster age 15+
        */

        LABEL ROSTER_VIOL_AGECAT = "Violence weighting age categories from Roster
        Age";
RUN;

PROC SORT DATA=HH_QX; BY EA_HHID_FIXED; RUN;
PROC SORT DATA=ROSTER; BY EA_HHID_FIXED; RUN;

DATA NEW_ROSTER;
        MERGE ROSTER (IN=AA) HH_QX (IN=BB);
        BY EA_HHID_FIXED;

        LABEL VM_FLAG = "Adult Female age 15 and older Selected for Violence Module"

        VM_FLAG = 0;

        IF AA AND BB then do;
                IF ROSTER_VIOL_AGECAT = 2 THEN DO;
                        IF EA_HHID_LN_FIXED=EA_HHID_VIOL THEN VM_FLAG = 1;
                END;
        END;

        ELSE IF AA THEN OUTPUT;

RUN;

DATA INDIV;
        SET w30.W30_indiv_qx_reduced;
        IF (TOUCHTIMES >= 0 AND CMPLSXTIMES >= 0 AND FRCSXTIMES >= 0 AND
        PRSSXTIMES >= 0) OR compress(VLNC) in ('1','2')

        THEN VM_QXSTATUS = 1;
        ELSEVM_QXSTATUS = 0;
RUN;

PROC SORT DATA=NEW_ROSTER; BY EA_HHID_LN_FIXED; RUN;
PROC SORT DATA=INDIV; BY EA_HHID_LN_FIXED; RUN;

DATA INDIV w31.W31_viol;
        MERGE INDIV(IN=A)
        NEW_ROSTER(KEEP=EA_HHID_LN_FIXED VM_FLAG
ROSTER_VIOL_AGECAT);
        BY EA_HHID_LN_FIXED;
```

```
        IF A;

        Label INDIV_VIOL_AGEGROUP = "Violence age group from Best Age";

        INDIV_VIOL_AGEGROUP = 0;
        IF INDIV_AGEGROUP = 3 THEN INDIV_VIOL_AGEGROUP = 2;/* Adult (15 - 64)
        */
        ELSE IF INDIV_AGEGROUP in(1,2) THEN INDIV_VIOL_AGEGROUP = 1;/*
        Child/Adolescent (0-14) */

        IF VM_FLAG = 0 THEN VM_STATUS = 0; /* Not selected for Violence Module */
        ELSE IF INDIV_STATUS = 4 THEN VM_STATUS = 5; /* Unknown Eligibility for
        Questionnaire*/
        ELSE IF INDIV_STATUS NOT IN (1, 2)   THEN VM_STATUS = INDIV_STATUS;
        /* others */
        ELSE IF BEST_GENDER ^= '2' THEN VM_STATUS = 3; /* Out of scope for Violence
        Module, changed to male in Interview */
        ELSE IF INDIV_VIOL_AGEGROUP IN (1,3) THEN VM_STATUS = 4; /* Out of
        scope for Violence Module, changed age out of 15 - 64 in Interview */
        ELSE IF VM_QXSTATUS = 1 THEN VM_STATUS = 1; /* Violence Module
        Respondent */
        ELSE VM_STATUS = 2; /* In-scope for Violence Module, Non-Respondent */
RUN;
```

# Appendix E

# HIV Knowledge Module Variables, Eligibility Criteria, and Program Code

# HIV Knowledge Module Variables, Eligibility Criteria, and Program Code

## E.1 List of HIVK Knowledge Variables

| NAME | LABEL |
|---|---|
| ONEPARTNR | Can the risk of HIV transmission be reduced by having sex with only one uninfected partner who has no other partners? |
| MOSQUITO | Can a person get HIV from mosquito bites? |
| CONDOMS | Can a person reduce their risk of getting HIV by using a condom every time they have sex? |
| SHAREFOOD | Can a person get HIV by sharing food with someone who has HIV? |
| HEALTHYINF | Can a healthy-looking person have HIV? |
| BUYFOOD | Would you buy fresh vegetables from a shop keeper or vendor if you knew the person had HIV? |
| KIDSSCHOOL | Do you think children living with HIV should be allowed to attend school with children who do not have HIV? |
| FEARTEST | Do you think people hesitate to take an HIV test because they are afraid of how other people will react if the test result is positive for HIV? |
| TALKBAD | Do people talk badly about people living with HIV or who are thought to be living with HIV? |
| RESPECT | Do people living with HIV, or thought to be living with HIV, lose the respect of other people? |
| SALIVA | Do you fear that you could get HIV if you come into contact with the saliva of a person living with HIV? |
| FAMSHAME | Do you agree or disagree with the following statement: I would be ashamed if someone in my family had HIV. |

## E.2 Eligibility Criteria for HIVK Module

The variable HIVK_STATUS was created to identify individuals eligible to receive the HIV knowledge module and was assigned to every rostered record, with values as shown in the table below. Codes 1 through 9 were assigned only to cases flagged to receive the HIV knowledge module.

| HIVK_STATUS | Description |
|---|---|
| 0 | Not selected for HIVK Module |
| 1 | HIVK Module Respondent |
| 2 | HIVK Module Eligible Non-Respondent |
| 4 | No data, unknown whether eligible for survey |
| 7 | Rostered in Error |
| 8 | Not Sampled (children in households with child flag = NO) |
| 9 | Extraneous Cases – De Jure Ineligible |

## E.3 Program Code for HIVK Response Status

```
data eligibles (keep = ea_hhid_ln_fixed hivk_status onepartnr);
set w30.w30_indiv_qx_reduced;
    where confagey_RECODE >= 15   and
            indiv_hivkflag= "1" and
             indiv_status = 1;

if onepartnr in ("1","2","3") then HIVK_STATUS = 1;
else
  if onepartnr in ("-9"," ") then HIVK_STATUS = 2;
run;

proc sort data = eligibles (drop = onepartnr);
  by ea_hhid_ln_fixed;
run;

proc sort data = w30.w30_indiv_qx_reduced  out = w30_indiv_qx_reduced;
  by ea_hhid_ln_fixed;
run;
```

```
data W32_HIVK;
merge eligibles(in=a) w30_indiv_qx_reduced (in=b);
  by ea_hhid_ln_fixed;
  if b;
  if b and not a then HIVK_STATUS=0;
run;

data w32.w32_hivk;
set w32_hivk;
if indiv_status => 3 then hivk_status = indiv_status;
run;
```

# Appendix F

# CASI Module Variables, Eligibility Criteria and Program Code

# CASI Module Variables, Eligibility Criteria and Program Code

## F.1 List of CASI Variables

| Variable | Label |
|---|---|
| CASTT1 | Are you currently in Zimbabwe? |
| CASTT2 | What is the day after Wednesday? |
| CASTT3 | What are the first three letters of Zimbabwe? |
| CASTT4 | Please type in the number 18. |
| CSOLDLB | How old were you at your last birthday? |
| CSOLDLBDKS | Please provide the reason this previous question was left blank: How old were you at your last birthday? |
| CSHSAPSH | What is the highest level of school you attended: primary, secondary, or higher? |
| CSWORKMO | Have you done any work in the last 12 months for which you received a paycheck, cash, or goods as payment? |
| CSMRLIVETOG | Have you ever been married or lived together with a [${prtgnd_disp}***] as if married? |
| CSDRATYDAY | How often do you have a drink containing alcohol? |
| CSHMDRATYDAY | How many drinks containing alcohol do you have on a typical day? |
| CSHODROOCA | How often do you have six or more drinks on one occasion? |
| CSHODSEX | How old were you when you had sex for the very first time? |
| CSHODSEXFT | Please provide the reason this previous question was left blank: How old were you when you had sex for the very first time? |
| CSPPOSWP | People often have sex with different partners over their lifetime. In total, with how many different people have you had sex in the last 12 months? |
| CSPPOSWPM | Please provide the reason this previous question was left blank: People often have sex with different partners over their lifetime. In total, with how many different people have you had sex in the last 12 months? |
| CSLTRIMHPSW | In the last three months, how many partners have you had sex with? |
| CSLTRIMHPSWU | Please provide the reason this previous question was left blank: In the last three months, how many partners have you had sex with? |
| CSXPHMSC | Of these [${csltrimhpsw}*] partners, with how many did you have sex without a condom, even if it was only one time? |
| CSXPHMSCTI | Please provide the reason this previous question was left blank: Of these [${csltrimhpsw}*] partners, with how many did you have sex without a condom, even if it was only one time? |
| CSIYSP | What are the initials of your last sex partner? |
| CSLTUSECU | The last time you had sex with [${csiysp}*] was a condom used? |
| CSRELASUOWAY | Did you enter into a sexual relationship with [${csiysp}*] because [${csiysp}*] provided you with or you expected that [${csiysp}*] would provide you with material support in other ways? |
| CSILUSSM | In the last 12 months, have you sold sex for money? |
| CSILHUPMFS | In the last 12 months, have you paid money for sex? |
| CSILTHIMHPUP | In the last 3 months, how many people did you pay to have sex with? |
| CSILTHIMHPTHS | Please provide the reason this previous question was left blank: In the last 3 months, how many people did you pay to have sex with? |
| CSXPSWCEOTI | Of these [${csilthimhpup}*] partners, with how many did you have sex without a condom, even if it was only one time? |

| Variable | Label |
|---|---|
| CSXPSWCEOTIME | Please provide the reason this previous question was left blank: Of these [${csilthimhpup}*] partners, with how many did you have sex without a condom, even if it was only one time? |
| CSBVISITHOTO | Before we visited your house today, have you ever tested for HIV? |
| CSMAYHIVTESTM | MONTH: |
| CSMAYHIVTESTY | YEAR: |
| CSHIVRESULT | What was the result of that HIV test? |
| CSACTARV | Are you currently taking ARVs, that is, antiretroviral medications? |
| CSDOHOBYWIFE | Do you believe it is right for a man to hit or beat his wife if she refuses to have sex with him? |
| CSDORELATIONWIVE | Do you believe married men need to have sex with women they are not married to, even if they have good relationships with their wives? |
| CSDPDINVIEW | Do you prefer to do such an interview by yourself on the computer or do you prefer our staff to read the questions to you? |
| CSAMOREPRIVATE | Are you more likely to give private information to a computer or to a person? |

## F.2 Eligibility Criteria for CASI Module

The variable CASI_STATUS was created to identify interview respondents selected to receive the CASI module and was assigned to every rostered record, with values as shown in the table below. Codes 1 through 3 were assigned only to interviewed cases flagged to receive the CASI module.

| CASI_STATUS | Description |
|---|---|
| 0 | Not selected for CASI Module |
| 1 | CASI Module Respondent |
| 2 | CASI Module Eligible NonRespondent |
| 3 | Gender did not match CASIFLAG gender |
| 4 | Unknown eligibility Status |
| 6 | Collected in Another Tablet |
| 7 | Rostered in Error |
| 8 | Not Sampled |
| 9 | Extraneous Cases – De Jure Ineligible |

## F.3 Program Code for CASI Response Status

```
data newhhqx (rename=(DMFLAG=DMFLAG_hhqx));
    set casi.pzim_ffcorr_hhqx_casi_20181115 (keep = EA_HHID_FIXED CASIMFLAG
CASIMMAX CASIMCHOICE CASIFFLAG CASIFMAX CASIFCHOICE DMFLAG);
    where dmflag in (" ","Corrected") and
       ((casimmax not in (" ","0") or not missing (casimchoice)) or
       (casifmax not in (" ","0") or not missing (casifchoice)));
```

```
If CASIMMAX not in ("", "0") or not missing (CASIMCHOICE) then CASI_FMFLAG=1;  /*
Select Male */
Else
  if CASIFMAX not in ("", "0") or not missing (CASIFCHOICE) then CASI_FMFLAG=2;   /*
Select Female */
If not missing (CASIMCHOICE) then CASI_LN = CASIMCHOICE;
Else
  if not missing (CASIFCHOICE) then CASI_LN = CASIFCHOICE;
run;

Data NewRoster(rename=(DMFLAG=DMFLAG_ROSTER));
set casi.pzim_ffcorr_Roster_casi_20181115 (keep = EA_HHID_FIXED EA_HHID_LN_FIXED
CASIRAND CASIMRAND CASIFRAND DMFLAG);
  where DMFLAG in (" ", "Corrected");
run;

proc sort data = NewRoster nodupkey;
  by EA_HHID_FIXED EA_HHID_LN_FIXED;
run;

proc sort data = Newhhqx;
  by ea_hhid_fixed;
run;

data NewRoster1;
merge NewRoster (in=aa) NewHHqx;
  by ea_hhid_fixed;
if aa then do;
  last_two_ea = substr(ea_hhid_ln_fixed,14,2) * 1;
  if last_two_ea = CASI_LN then CASI_FLAG = 1;
  else
    CASI_FLAG = 0;
  output;
end;
run;
```

Merge in data file from weighting process (to pick up BEST_AGE BEST_GENDER and INDIV_STATUS)

```
If 15 <= BEST_AGE <=49 and
   INDIV_STATUS =1     and
   CASI_FLAG= 1        Then Do;
  If 0 <= CSOLDLB <= 2015 then CASI_STATUS=1;
  Else
    CASI_STATUS=2;
End;
Else
  if INDIV_STATUS => 3 then CASI_STATUS=INDIV_STATUS;
  Else
    CASI_STATUS = 0;
If CASI_STATUS in (1,2) then do;
  If CASI_FMFLAG ^= BEST_GENDER Then CASI_STATUS =3;
  else
    IF CASI_FMFLAG = . then CASI_STATUS = 0;
end;
run;
```

# Appendix G

# Eligibility Criteria and Program Code for Weight and Height Measurements

# Eligibility Criteria and Program Code for Weight and Height Measurements

## G.1 Eligibility Criteria for Weight and Height Measurements

The variable CWH_STATUS was created to identify children eligible to receive weight and height measurements and was assigned to children 0-60 months old who had a blood test weight, with values as shown in the table below.

| CWH_STATUS | Description |
|---|---|
| 1 | Provided W/H measurements |
| 2 | Did not provide W/H measurements |
| . | Not selected for W/H measurements |

## G.2 Program Code for Response Status for Weight and Height Measurements

```
DATA CWH;
   SET W100.Blood_delivery;

   IF CONFAGEM not in (' ' 'AGE NOT RECORDED')
               THEN CONFAGEM_r = CONFAGEM+0;

              CWH_FLAG = 0;
              IF CWHDATE > 0 OR HIVSTATUS = 1 OR HIVSTATUSC = 1
                 THEN CWH_FLAG = 1;

   IF 0<= CONFAGEM_r <=60 AND BTWT0 > 0;

   RUN;

DATA FRM;
   SET CWH (RENAME=(CWHHEIGHT=CWHHEIGHT_A
CWHWEIGHT=CWHWEIGHT_A ));
```

```
    CWHHEIGHT=INPUT(CWHHEIGHT_A,8.2);
    CWHWEIGHT=INPUT(CWHWEIGHT_A,8.2);

    IF      CWH_FLAG=1 and CWHHEIGHT ^= . AND CWHWEIGHT ^= . THEN
CWH_RESP = 1;
    ELSE IF CWH_FLAG=1 and (CWHHEIGHT = . OR CWHWEIGHT = .)  THEN
CWH_RESP = 2;
    ELSE        CWH_RESP = .;

    RUN;
```

# Appendix H

# Child module weight creation and eligibility criteria

## H.1 Purpose of the child module weights

As described in Section 2.4.5, a subset of all sampled households was randomly selected for additional child data collection. In these selected households, children were eligible for blood testing, and additional interview questions were asked either of the child (for adolescents) or the parent/guardian. In other households this additional data was not collected. The blood test and interview weights (btwt and intwt, respectively) on the child biomarker and individual datasets allow for analysis of the variables only collected in the households selected for additional child data collection.

Although the information available for children in the selected households is more detailed, questions included in the child module of the adult interview were administered to parents and guardians of all children in the household. The household roster also contains information about all children in the household. If an analysis aims to use these data, the sample population is different: specifically, this sample includes all rostered children who would have been eligible to participate, irrespective of whether their household was flagged for child data collection. In these situations, a separate set of weights is needed. These are referred to hereafter as child module weights.

## H.2 Child module weight creation process

Three main steps were carried out to create the child module weights:

1. Create a list of all children aged 0-14 rostered in any responding household who were de facto eligible (i.e., slept in the household the night before) and had a responding parent or guardian, and link each child to their parent or guardian using the line number of the responding adult in the household (parentguardqx variable in the child interview dataset).
2. Assign each child an initial weight equal to the linked adult's non-response adjusted (but not post-stratified) interview weight (trmpnr1w0 from the intermediary weights file). We refer to this weight as the child module base weight, chmodbw0.
3. Post-stratify the resulting set of weights to ensure that the total populations by five-year age group and gender sum to the control totals used for the blood test and interview weights. We refer to the resulting weight as the child module final weight, chmodfw0.

In step one, individuals in the child dataset were included as possible guardians, because there can be cases where someone under 15 years of age responded as the parent or guardian of another child in the household. Records for children who would not have been eligible for the survey were excluded.

Potentially eligible children have indstatus = 1 or 8 (see section E.5 below for full details on the eligibility criteria).

In step two, if the adult did not respond or was deemed ineligible for some reason (for example, if they did not stay in the household the previous night), their interview weight was set to zero, so their associated children will also have a child module weight of zero.

The post-stratification in step three used an adjustment factor that was computed for each cell defined by gender and five-year age group of the rostered children. This adjustment factor is equal to the control total in each cell divided by the sum of the chmodbw0 weights of the children in the cell. Each child's chmodbw0 weight in the given cell was multiplied by the corresponding adjustment factor to obtain the final weight, chmodfw0.

Steps two and three were repeated for each replicate weight set (trmpnr1w001-trmpnr1wXXX) to create the associated jackknife replicate weights for the child module. First, the child module replicate base weights were computed as chmodbw001 = trmpnr1w001, chmodbw002 = trmpnr1w002, ..., chmodbwXXX = trmpnr1wXXX. Each set of jackknife replicate weights was then used to compute the corresponding replicate-specific post-stratification adjustment factors and final post-stratified replicate weights, chmodfw001, chmodfw002, ..., chmodfwXXX.

## H.3 Variables available for all children and when to use these weights

The child module weights should only be used when the analysis variables are collected for all rostered children (i.e., eligibility for data collection is not restricted to whether the household was flagged for child data collection). In general, this includes variables from the roster, such as age and gender, as well as questions from the adult questionnaire's children module that have been attached to the child records. These variables can be identified by filtering the variable category in the child interview dataset codebook to "Adult questionnaire - Module 3A: children" (note that the module number may vary by country). Most of these variables have the prefix "ch_" in their variable names to assist with identification. Additional information about the mother is available for linked children in the variables prefixed "mom". Questions from the "Household questionnaire – Child" category are also available for all children because these are completed by the head of household.

Variables which are asked in the adolescent interview or related to blood testing are not available for children in non-selected households, so the child module weights should not be used for these.

## H.4 Further non-response adjustments

The child module weights are general-purpose weights which are a reasonable approximation of the weights that would be obtained through a more complex non-response adjustment procedure like that used for the main child interview weights. A major assumption is that the non-response pattern for children is captured fully by the non-response adjustments carried out for the linked adults. It is possible that these non-response adjustments do not fully account for some specific characteristics of the child. For example, older children may tend to have more missing data than younger ones, and missing parent/guardian links may occur at different rates for different ages or other groups of children. To more fully compensate for these patterns a precise definition of response status for children would have to be developed based on the questions answered, and non-response adjustments applied to relevant response cells based on child-level characteristics. For highly detailed or specialized analysis we recommend that the non-response patterns be checked for the particular groups of interest for the analysis to determine whether any further adjustments may be needed.

## H.5 Child module weight eligibility criteria

The following table shows all combinations of values for variables defining eligibility for child module weights. Children who were unable to be linked to an adult (linked adult indstatus = .) or whose linked adult was not an eligible respondent (linked adult indstatus = 2, 7, 8, or 9) are ineligible. Among children who had an eligible, responding, linked adult, those with indstatus = 2, 6, 7, 9 were also ineligible (2 = non-responding sampled child, 6 & 7 = were duplicated or erroneous child records, 9 = de jure ineligible).

Only those children in rows 1 and 5 below in the table, with indstatus = 1 or 8, linked adult indstatus = 1, and sleephere = 1, are assigned child module weights.

Table H-1 Variables determining child module weight eligibility criteria

| Linked adult's indstatus | Child's indstatus | Child's sleephere | Explanation of child eligibility status |
|---|---|---|---|
| 1 | 1 | 1 | Eligible: Sampled child with responding adult. These children have valid individual weights (intwt) |
| 1 | 2 | 1 | Ineligible: Sampled child with linked adult, but considered non-respondent (e.g., parent refused consent or did not provide sufficient data) |
| 1 | 6 | 1 | Ineligible: the child record was collected in another tablet |
| 1 | 7 | 1 | Ineligible: the child was rostered in error |
| 1 | 8 | 1 | Eligible: Child with linked, responding adult, in a household not sampled for child blood testing. These children do not have individual weights (intwt) but are eligible for child module weights (chmodfwt) |
| 1 | 9 | 1 | Ineligible: Non-de facto child. The adult was an eligible respondent, but the child had ind0040 = 3 (not available), 6 (incapacitated), or did not sleep in HH the night before. |
| 2,7,8,9 | 1 | 1 | Ineligible: Ineligible or non-responding linked adult |
| 2,7,8,9 | 2 | 1 | Ineligible: Ineligible or non-responding linked adult |
| 2,7,8,9 | 6 | 1 | Ineligible: Ineligible or non-responding linked adult |
| 2,7,8,9 | 7 | 1 | Ineligible: Ineligible or non-responding linked adult |
| 2,7,8,9 | 8 | 1 | Ineligible: Ineligible or non-responding linked adult |
| 2,7,8,9 | 9 | 1 | Ineligible: Ineligible or non-responding linked adult |
| . | 2 | 1 | Ineligible: Not able to be linked to an adult |
| . | 8 | 1 | Ineligible: Not able to be linked to an adult |
| . | 9 | 2 | Ineligible: Not able to be linked to an adult |