SHIMS2
SWAZILAND HIV INCIDENCE MEASUREMENT SURVEY

SAMPLING AND WEIGHTING TECHNICAL REPORT
SHIMS2 2016-2017

APRIL 2019

ICAP GLOBAL. HEALTH. ACTION.
Columbia University Mailman School of Public Health

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

CDC	US Centers for Disease Control and Prevention
CHAID	Chi-square Automatic Interaction Detector
CI	Confidence Interval
CSO	Central Statistics Office
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
SHIMS	Swaziland HIV Incidence Measurement Survey
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

Introduction 1

The 2016 Swaziland HIV Incidence Measurement Survey (SHIMS 2) is a Population-based HIV Impact Assessment (PHIA) designed to assess the prevalence of key human immunodeficiency virus (HIV)-related health indicators. Data collection for SHIMS 2 was conducted between September 2016 and March 2017 with over 15,000 individuals in approximately 5,200 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 SHIMS 2 is a stratified multistage probability sample design, with strata defined by region and urban/rural status, 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 "primary sampling units" or PSUs) were stratified by urban/rural status within each of the four regions of the country, and then within each stratum were selected with probabilities proportionate to the number of households in the PSU based on the 2007 census. The allocation of the sample PSUs to the eight sampling strata was made in a manner designed to achieve specified precision levels for a national estimate of HIV incidence rate, and four regional 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 stratum to the extent feasible.

Within the sampled households, all eligible adults 15 years of age and 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 sample design employed for SHIMS 2 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, if necessary.
- 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 individuals within households, compensate for nonresponse to the blood test, and adjust for potential undercoverage through poststratification.

Technical details of the weighting procedures employed in SHIMS 2 are provided in Section 3.

Sample Design 2

2.1 Population of Inference

The population of inference for SHIMS 2 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 data collection; however, as discussed later in Section 2.5, only members of the *de facto* population are included in the study population. Table 2-1 summarizes projections of the 2017 Swaziland *de jure* population by gender and age group.

Table 2-1 Summary of 2017 population projections for Swaziland by gender and age group

Age group	Gender		Total
	Male	Female	
14 years or younger	212,548	204,691	417,239
15 to 49 years	280,129	331,589	611,718
50 and older years	52,132	64,881	117,013
Total	544,809	601,161	1,145,970

Source: 2017 population projections provided by the Swaziland Central Statistics Office (CSO).

2.2 Precision Specifications and Assumptions

The following specifications and assumptions were used to develop the sample design for SHIMS 2.

Specifications

- The relative standard error (RSE) of the national estimate of annual HIV incidence among persons aged 15-49 should be 20% or less.
- 95% confidence bounds around the estimated viral load suppression (VLS) rate among HIV+ persons aged 15-49 for each of the four regions of the country should be ± 10% or less.

Assumptions

- An overall HIV prevalence rate of 0.2770 (27.7%) that varies by region and urban/rural status (e.g., see Table 2-2). Source: 2011 Swaziland HIV Incidence Measurement Survey (SHIMS 1) conducted from December 2010 to June 2011 (see SHIMS, First Findings Report, November 2012).

- An annual HIV incidence rate for adults aged 15-49 of $P_a = 0.0189$ (1.89%). Source: UNAIDS estimate for 2014.

- A mean duration of recent infections (MDRI) of 130 days, yielding an annualization factor of 365/130= 2.8077. Hence, the estimated incidence rate for MDRI = 130 days is $P_m = 0.0189/2.8077 = 0.0067$ (0.67%).

- A viral load suppression (VLS) rate among HIV+ adults aged 15-49 in each region *h* of $P_{vh} = 50\%$. This is a conservative assumption because it will overstate the actual variance of the VLS rate.

- An average of 20 occupied sampled households per sampled cluster (PSU).

- An intra-cluster correlation (ICC) of $\rho = 0.05$ for prevalence and VLS rates. An intraclass of $\rho = 0.02$[1] was used for estimates of incidence. The ICC provides an average measure of the homogeneity of responses within the first-stage sampling units.

- An occupancy rate of 89.4% for sampled dwellings. Note that this is not included in the calculation of the overall survey response rate, but does determine the initial numbers of dwelling units to be sampled. Source: 2011 Swaziland HIV Incidence Measurement Survey (SHIMS 1).

- An overall household response rate of 94.4% among the occupied dwelling units. Source: 2011 Swaziland HIV Incidence Measurement Survey (SHIMS 1).

- The average number of persons aged 15 to 49 in a household is 2.33. Source: 2011 Swaziland HIV Incidence Measurement Survey (SHIMS 1) and 2016 national population projections produced by the Swaziland Central Statistics Office (CSO).

- The percentage of persons in households who are age 0-14 is 37.9%. Source: 2016 national population projections for Swaziland.

- The percentage of persons in households who are age 18-49 is 42.6%. Source: 2016 national population projections for Swaziland.

[1] This estimate was derived from tabulations of the 2011 SHIMS data files.

- The percentage of persons in households who are 50 years or older is 12.4%. Source: 2016 national population projections for Swaziland.
- Among the individuals 15 years of age and older, an overall biomarker response rate of 77.1% reflecting sample loss due to any of the following reasons: nonparticipation (refusal) of some sample households, nonresponse to the individual interview, refusal to provide a blood sample, or providing a non-analyzable blood sample. Source: 2006-07 Swaziland Demographic and Health Survey
- Among the children in the eligible responding households, an overall biomarker response rate of 72.1% for persons 0-14 years of age. This value is the comparable response rate for adults minus 5%.

Based on the assumptions listed above, a sample of 287 clusters (EAs) was determined to be the minimum needed to meet the specified precision goals. The allocation of the sample to the eight urban/rural subregions (strata) is shown in Table 2-2. The expected numbers of households included in the study and the corresponding projected numbers of respondents by age group are also summarized in Table 2-2. The actual numbers of respondents obtained in SHIMS 2 are presented in Sections 2.4 and 2.5, and differ from the counts in Table 2-2 because of differences between the response rates and other assumptions used to develop the sample design and the actual values achieved during data collection. Further details about the sampling of households are given in Section 2.4.

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

Stratum [1]	Est. HIV prevalence rate [2]	Sample clusters (EAs)	Target no. dwelling units to be sampled [3]	Exp. no. households [4]	Projected number of respondents [5]		
					15 to 49	50 or older	0 to 14 [6]
Hhohho - Urban	0.2751	19	425	380	613	155	226
Manzini - Urban	0.2803	22	492	440	732	169	248
Shiselweni - Urban	0.2680	12	268	240	386	113	140
Lubombo - Urban	0.2677	14	313	280	478	106	168
Hhohho - Rural	0.2707	63	1,409	1,260	2,153	523	689
Manzini - Rural	0.2948	72	1,610	1,440	2,368	517	735
Shiselweni - Rural	0.2651	41	917	820	1,451	431	589
Lubombo - Rural	0.2871	44	984	880	1,498	346	566
TOTAL	0.2770	287	6,417	5,740	9,680	2,362	3,361

[1] Sampling strata defined by crossing the four regions of the country by urban/rural status

[2] Source: 2011 SHIMS and Spectrum 2014

[3] These are the original targets specified under the design

[4] Assumes occupancy rate of 0.894

[5] Entries are projected counts based on the assumptions used to develop the sample design

[6] 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 SHIMS 2 are defined to be the Enumeration Areas (EAs) created for the 2007 Swaziland Census of Population and Housing. The 2007 sampling frame consisted of 2,064 EAs containing an estimated 212,000 households and 1,005,000 persons, with an average number of households and persons per EA of 103 and 487, respectively.
Before the EA sample was selected, 43 EAs with less than 25 households were excluded from the selection. The excluded EAs accounted for less than 0.48% of all households in the sampling frame.

2.3.2 Selection of the PSU Sample

A stratified sample of 287 EAs was selected from the final EA sampling frame in accordance with the sample allocation given in Table 2-2. The eight strata specified for sampling were defined by crossing the four regions Hhohho, Manzini, Shiselweni and Lubombo by urban/rural status. The

EA samples were selected systematically and with probabilities proportionate to a measure of size (MOS) within each stratum. The MOS used for sampling was equal to the number of households in the EA based on the 2007 Census of Population and Housing.

The first step of the sampling process was to divide the sampling frame of EAs into the eight strata indicated in Table 2-2. Next, within each stratum, the EAs were sorted by the unique 5-digit EA identification code. 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 geography are included in the sample. To select the sample from a particular stratum, 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 stratum 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 within a particular stratum is proportional to the MOS of the EA in the stratum.

2.4 Selection of Households

The selection of households for SHIMS 2 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 child data collection.

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." Local leaders and knowledgeable community members were consulted to assist in the listing process. For each of the 287 EAs selected for the study, listers were provided with maps from which to delineate the boundaries of the EA, and to record the general locations of the dwelling units/households found by the listers in the field. Information about the listed dwelling units/households was entered into computer tablets. The information recorded in the tablets included the address or description of the listed dwelling unit/household, the name of the head of household, the type of structure (house, apartment, compound, etc.), occupancy status, and GPS coordinates. Vacant structures were listed along with households in occupied dwelling units. Over 51,500 dwelling units/households were listed for SHIMS 2.

2.4.3 Determination of Eligibility for Sampling

As indicated above, all known households at the time of listing, plus vacant dwelling units that could potentially be occupied at the time of interview, were initially entered into the computer tablets as separate records. However, not all of these records were eligible for subsequent sampling purposes. Those records marked with the notation "discard" were data entry errors and were eliminated from sampling consideration. To establish eligibility for the remaining records, three key variables collected during listing were used: (1) the structure type, (2) whether the listed structure was vacant or under construction, and (3) whether anyone was living in the structure at the time of listing. Based on the values of these three variables, those records meeting the criteria specified in Appendix A were eligible for household sampling. Table 2-3 summarizes the number of records entered into the tablets, the number of discarded listings, the numbers of unoccupied and occupied dwelling units eligible for sampling, and the total number of dwelling units/households (records) eligible for sampling.

Table 2-3 Distribution of records in listing file by type of record and eligibility status, by stratum

Stratum	Number of dwelling units/house-holds in listing file	Number of discarded listings	Number of unoccupied dwelling units [1]	Number of unoccupied dwelling units eligible for sampling [2]	Number of occupied dwelling units/house-holds [3]	Number of occupied dwelling units/house-holds eligible for sampling	Total number of dwelling units/house-holds eligible for sampling
Hhohho Rural	9,775	0	700	693	9,075	9,069	9,762
Hhohho Urban	3,384	1	138	135	3,245	3,245	3,380
Lubombo Rural	6,032	0	573	561	5,459	5,457	6,018
Lubombo Urban	2,618	0	516	514	2,102	2,098	2,612
Manzini Rural	14,585	0	648	624	13,937	13,936	14,560
Manzini Urban	7,404	0	175	174	7,229	7,229	7,403
Shiselweni Rural	4,519	1	315	302	4,203	4,203	4,505
Shiselweni Urban	3,261	0	75	75	3,186	3,186	3,261
TOTAL	51,578	2	3,140	3,078	48,436	48,423	51,501

[1] Records coded as vacant, under construction, or with no residents at time of listing (see Appendix A).

[2] Subset of the unoccupied dwelling units that could potentially serve as residential quarters (see Appendix A).

[3] All records not coded as vacant, under construction, or with no residents at time of listing (see Appendix A).

2.4.4 Selection of Dwelling Units

In order to achieve an equal probability sample of dwelling units within each of the four regions of the country, the sampling rates required to select dwelling units within an EA will depend on the difference between the size measure used in sampling (i.e., the number of households in the EA based on the 2007 census) and the actual number of dwelling units/households found at the time of listing in mid-2016. Thus, application of these within-EA sampling rates can yield more than the desired 20 households in EAs that have experienced growth in population since the 2007 census, and fewer than 20 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 region h = 1, 2, ..., 4, was computed as:

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

where

T_h = target sample size for region *h*
= $T_h^U + T_h^R$;
T_h^U = target sample size for the urban part of region *h* given in Table 2-2 ;
T_h^R = target sample size for the rural part of region *h* given in Table 2-2 ;
m_h = number of sample EAs in region *h*
= $m_h^U + m_h^R$
m_h^U = the number of sample EAs in the urban part of region *h* ;
m_h^R = the number of sample EAs in the rural part of region *h* ;
N_{hi} = number of eligible dwelling units in PSU *i* in region *h* based on listing counts;
P_{hi} = probability of selecting PSU *i* in region *h* .

The total *targeted* number of listings to be selected across all four regions is $\sum_{h=1}^{4} T_h$ = 6,417 (see Table 2-2). To obtain an equal probability sample within region *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 region *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 had experienced considerable growth, the maximum number of dwelling units to be selected in any EA was capped at 40, and the minimum number of be selected was set to 15, except for one EA that contained only one eligible dwelling unit. The difference between the number of dwelling units that would have been selected using the sampling rates, f_{hi}^{within} , and the corresponding capped number was then re-distributed to the other EAs in the same region so as to maintain the desired total sample size for the region. In other words, the within-EA sampling rates, f_{hi}^{within}, defined above were adjusted to reflect the capping and the redistribution of the sample within the region as:

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

where the adjustment factors, A_{hi}, were determined such that $15 \leq N_{hi} A_{hi} f_{hi}^{within} \leq 40$ 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 unit/household records in each EA were sorted by lister, village/neighborhood name, structure number, apartment number if applicable, and finally by household number assigned at the time of listing. Dwelling units/households within the EA were then selected systematically from the ordered list of records at the rates, $f_{hi}^{adj(w)}$, specified above. In addition, a random half sample of the selected dwelling units/households was designated (flagged) for child data collection.

2.4.5 Results of Second-Stage Sampling

Table 2-4 summarizes the number 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 (region). As indicated in the previous section, there was one EA containing only one eligible dwelling unit for which no minimum sample size was set. Though given an appropriate chance of selection, no dwelling units were selected from this EA. Thus, only 286 of the 287 sampled EAs contained any dwelling units selected for data collection.

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 (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 2-4 Number of sampled dwelling units/households and expected unequal weighting design effects by region

Region	No. sample PSUs (clusters)	Number of sampled dwelling units/house-holds	Number of dwelling units/house-holds flagged for child data collection	Minimum PSU sample size	Maximum PSU sample size	UEW DEFF for sample after capping
Hhohho	82	1,833	916	15	40	1.01
Lubombo	57 [1]	1,297	648	15	36	1.00
Manzini	94	2,102	1,051	15	43	1.07
Shiselweni	53	1,185	593	15	40	1.03
Total	286	6,417	3,208	15	43	1.05 [2]

[1] No dwelling units were selected from one EA in this region due the small number of eligible dwellings.

[2] Reflects total variation in weights within and across strata.

Table 2-5 summarizes the number of dwelling units selected for SHIMS 2 by final household response status. Of the 6,417 sampled dwelling units 310 (4.8%) were determined during data collection to be vacant/unoccupied, 51 (0.8%) for which eligibility for the survey (i.e., occupancy status) could not be established, 871 (13.6%) were determined to be eligible for the study (i.e., contained eligible household members) but did not complete the household roster, and 5,185 (80.8%) completed the household roster. The overall unweighted household response rate was 84.9%.

Table 2-5 Distribution of dwelling unit sample by region and response status

Region	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 nonresp-onding households	Unweighted response rate [3]
Hhohho	1,833	89	20	1,474	250	0.846
Lubombo	1,297	72	11	1,064	150	0.869
Manzini	2,103	108	12	1,703	280	0.854
Shiselweni	1,184	41	8	944	191	0.826
Total	6,417	310	51	5,185	871	0.849

[1] Vacant or unoccupied dwelling units, households with no persons eligible for SHIMS 2.

[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.

2.5 Selection of Individuals

The selection of individuals for SHIMS 2 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; and (3) selecting for the study 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 2-6 summarizes the number of households completing the roster and the corresponding number of rostered individuals by stratum and resident status.

Table 2-6 Number of households completing rosters and number of persons by resident status

Region	Number of households completing rosters	Usual resident but did not sleep here	Usual resident and slept here	Nonresident but slept here	Total
Hhohho	1,474	124	5,689	174	5,987
Lubombo	1,064	114	4,171	143	4,428
Manzini	1,703	106	6,064	201	6,371
Shiselweni	944	141	3,904	166	4,211
Total	5,185	485	19,828	684	20,997

2.5.2 Selecting Individuals for Data Collection

All of the individuals listed in the household rosters who were 15 years of age and older and were either usual residents of the household or slept in the household were eligible for data collection. However, children 0-14 years of age were eligible for data collection only if the household in which

they resided had been randomly designated for child data collection (see Section 2.4.5). Table 2-7 summarizes the number of individuals eligible for data collection by region, age group, and resident status.

Although data collection was attempted for all of the 13,209 adults and 4,076 children indicated in Table 2-7, 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 12,867 adults and 4,006 children who slept in the household during the night prior to the interview.

Table 2-7 Number of individuals eligible for data collection

Region	Adults 15 and older [1]				Children 0-14 [1]			
	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
Hhohho	82	3,722	117	3,921	22	1,084	36	1,142
Lubombo	81	2,480	87	2,648	19	866	25	910
Manzini	81	3,898	136	4,115	11	1,128	42	1,181
Shiselweni	98	2,308	119	2,525	18	804	21	843
Total	342	12,408	459	13,209	70	3,882	124	4,076

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

2.5.3 Distribution of Person Samples

Tables 2-8A through 2-8C summarize the number of individuals selected for data collection and the corresponding numbers completing the interview and blood test, for adults 15 years old and older, 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 SHIMS 2 study population are shown under the *de facto* heading in these tables. Note that for children 0-9 years in Table 2-8C, 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 2-8A Distribution of completed interviews and blood tests for adults 15 years and older

Region	*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]
Hhohho	3,839	3,527	3,276	82	43	41
Lubombo	2,567	2,333	2,255	81	60	58
Manzini	4,034	3,655	3,402	81	48	47
Shiselweni	2,427	2,149	1,991	98	63	57
Total	12,867	11,664	10,924	342	214	203

Note: This table includes adults 15 years and older based on age recorded in roster. In a small number of cases, the actual age at interview may be different. See Section 3.4.3.

[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 10,924 de facto cases completing the blood test, 5 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 2-8B Distribution of completed interviews and blood tests for adolescents 10-14 years

Region	*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]
Hhohho	378	346	336	5	4	4
Lubombo	291	261	259	9	7	7
Manzini	375	343	336	1	1	1
Shiselweni	277	250	244	7	5	4
Total	1,321	1,200	1,175	22	17	16

Note: This table includes adolescents 10-14 years based on age recorded in roster. In a small number of cases, the actual age at interview may be different. See Section 3.4.3.

[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 1,175 de facto cases completing the blood test, there were no cases who did not complete the interview. See Appendix B for more information about the response status categories defined for the blood tests.

Table 2-8C Distribution of completed interviews and blood tests for children 0-9 years

Region	*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]
Hhohho	742	727	625	17	14	12
Lubombo	600	582	514	10	8	3
Manzini	795	764	655	10	9	5
Shiselweni	548	516	413	11	9	5
Total	2,685	2,589	2,207	48	40	25

Note: This table includes children 0-9 years based on age recorded in roster. In a small number of cases, the actual age at interview may be different. See Section 3.4.3.

[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 2,207 de facto cases completing the blood test, 44 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.

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, p. 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 SHIMS 2, 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 SHIMS 2 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, which are computed 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 strata had no household data collected, an EA nonresponse adjustment is made for the remaining "responding" EAs in this stratum.

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 1.0% 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 2017 population projections produced by the Swaziland Central Statistics Office (CSO).

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 SHIMS 2 survey data that are used to construct the sampling weights are contained in the following data files.

- **psw_ffcorr_hh_qx_STAT_20170505**: A household (HH) file that contains the majority of household data collected in the HH questionnaire.
- **psw_ffcorr_deaths_STAT_20170505**: A household (HH) file that contains data collected in the HH questionnaire regarding any deaths that have occurred in the household since 2013.
- **psw_ffcorr_roster_STAT_20170505**: A file that contains the roster of household members collected in the HH questionnaire with a record for each rostered person.
- **psw_ffcorr_indiv_STAT_20170505**: 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.
- **swabiomarker20170505**: 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:

- Checking IDs, merging household survey files with sampling files, and accounting for records found in one file and not the other. This type of check for the EAs occurs as part of the household (HH) selection process.
- Check counts of sampled and responding HHs against what was expected, overall and by stratum.
- 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 (c) 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 SHIMS 2: 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) used in the SHIMS 2 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, the PSUs were selected systematically from a list of PSUs that had been ordered by region, urban/rural status within region, and EA code within each urban/rural status. To take account of the precision benefits of implicit stratification as fully as possible, the 286 sampled PSUs from which at least one dwelling unit was sampled were paired off in the systematic order in which they were selected within each region and urban/rural stratum, treating each pair as a variance-estimation stratum. When there was an odd number of sampled PSUs in a stratum, one of the variance-estimation strata was defined to contain three sampled PSUs.

For SHIMS 2, a total of 141 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 k = 1, 2, ..., 141 variance-estimation strata, resulting in a total of 141 jackknife replicates. 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		No. PSUs	No. variance strata consisting of pairs	No. variance strata consisting of triplets	Number of jackknife replicates
Hhohho	Rural	63	30	1	31
	Urban	19	8	1	9
Lubombo	Rural	44	22	0	22
	Urban	13	5	1	6
Manzini	Rural	72	36	0	36
	Urban	22	11	0	11
Shiselweni	Rural	41	19	1	20
	Urban	12	6	0	6
Total		286	137	4	141

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 SHIMS 2 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 141 variance-estimation strata and the variable **VarUnit** indicates the primary sampling unit (PSU) or cluster within the variance-estimation stratum. This pair of variables, which are provided in the SHIMS 2 data sets, 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 region 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 each of 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 stratum *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).

As indicated in Table 3-1, 141 jackknife replicates were formed from the 286 sampled PSUs. For variance estimation, replicate-specific PSU weights, $W_{(r)hi}^{(1)}$, r = 1, 2, ..., 141 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 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, ..., 141, the corresponding replicate-specific PSU base weight was computed as

$$\begin{aligned} W_{(r)hi}^{(1)} &= a\,W_{hi}^{(1)} && \text{if } h = r \text{ and PSU } i \text{ in variance-estimation stratum } h \text{ is included in replicate } r \\ &= 0 && \text{if } h = r \text{ and PSU } i \text{ in variance-estimation stratum } h \text{ is not included in replicate } r \\ &= W_{hi}^{(1)} && \text{if } h \neq r \end{aligned}$$

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

Table 3-2 Number of PSUs and weighted sums by region

Region	Number of sampled PSUs (EAs)	PSUs weighted by PSU base weights [1]	Number of PSUs with responding households
Hhohho	82	577	82
Lubombo	57	452	57
Manzini	94	589	94
Shiselweni	53	452	53
Total	286	2,070	286

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

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 (described in Section 3.4.1) and the reciprocal of the within-PSU household selection probability; i.e., the household base weight for sampled dwelling unit/household *j* in PSU *i* in stratum *h* was computed as:

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

where

$W_{hi}^{(1)}$ = the final weight for PSU *i* in stratum *h*

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

The corresponding weights for jackknife replicate $r = 1, 2, \ldots, 141$, were computed as:

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

where $W_{(r)hi}^{(1)}$ is the PSU weight for PSU *I* in stratum *h* in replicate *r* described in Section 3.4.1.

Next, the sampled dwelling units/households were assigned to one of the four response status groups specified in Table 3-3 based on criteria indicated in Appendix B. In Table 3-4, we show the corresponding weighted sums by response status and stratum 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 stratum.

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

Household response status group [1]	Description	Number of dwelling units/households
1	Eligible respondent	5,185
2	Eligible nonrespondent	871
3	Ineligible/out-of-scope	310
4	Unknown eligibility status	51

[1] See Appendix B for definitions.

Table 3-4 Weighted Sums of Household Base Weights by Response Status

Region	Household Response Status Code [1]				
	Status code 1: Eligible Respondents	Status code 2: Eligible Nonresp-ondents	Status code 3: Not Eligible (Vacant, Destroyed, not a DU, etc.)	Status code 4: Could not determine eligibility	Weighted Count of Households [2]
Hhohho	66,430	11,891	4,008	913	83,242
Lubombo	47,126	6,619	3,134	486	57,364
Manzini	88,483	15,535	5,591	623	110,231
Shiselweni	38,754	7,918	1,684	326	48,683
Total	240,792	41,963	14,417	2,348	299,520

[1] See Appendix B for definitions.

[2] 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 is to increase the weights of responding households so that they represent the nonresponding households in the same PSU. Because such nonresponse can 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 SHIMS 2), 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 SHIMS 2, a total of 4 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 combined with the nearest PSU on the sorted list of sample PSUs until the adjustment factor was below 2.00 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 3.64 (for a PSU with a response rate of 27.5%). After the grouping, the highest adjustment factor was reduced to 1.49.

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 stratum *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 stratum *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 stratum *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 stratum *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 stratum *h*

The first-phase household nonresponse adjustment factor for PSU weighting cell *i* in stratum *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 stratum *h*, and where the sum in the numerator extends over the entire sample of dwelling units/households in PSU weighting cell *i* in stratum *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 stratum *h* was then computed as:

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

The corresponding replicate weights for replicate $r = 1, 2, \ldots, 141$ 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 $r = 1, 2, \ldots, 141$.

The effect of this adjustment is to distribute the total weight of the undetermined-eligibility cases (i.e., the estimated 2,348 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

Region	Household Response Status				
	Status code 1: Eligible responding households	Status code 2: Eligible nonresponding households	Status code 3: Ineligible dwellings	Total dwelling units/house-holds	Total eligible households
Hhohho	67,187	12,026	4,029	83,242	79,213
Lubombo	47,506	6,711	3,147	57,364	54,217
Manzini	89,003	15,598	5,631	110,231	104,601
Shiselweni	39,028	7,964	1,691	48,683	46,992
Total	242,724	42,298	14,498	299,520	285,022

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 stratum *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 stratum *h*, and where the sum in the number extends over the sample of responding and nonresponding households in PSU weighting cell *i* in stratum *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 stratum *h* was then computed as:

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

The corresponding replicate weights for replicate $r = 1, 2, \ldots, 141$ 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 SHIMS 2 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 SHIMS 2 are categorized into three age groups for which different data elements are collected: (1) adults 15 or older, 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 are in the *de facto* population and 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 15 or older), 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 stratum *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 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, $W_{(r)hijk}^{(3)}$, $r = 1, 2, \ldots, 141$, 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 or older	Eligible Respondent	11,664	630,379
		Eligible Nonrespondent	1,186	65,412
Adolescents	10-14	Eligible Respondent	1,200	125,369
		Eligible Nonrespondent	120	12,949
Children	0-9	Eligible Respondent	2,589	275,550
		Eligible Nonrespondent	95	10,728

[1] Based on age reported at time of rostering.

[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 stratum 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). 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 54, 53, and 53 variables used in the original models for adults, adolescents, and children, respectively, the LASSO identified 25, 5, and 10 variables to be significant predictors of response for the three age groups, respectively, as shown 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 14, 4, and 3 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. A total of 35 final weighting adjustment cells were created for adults, 6 cells for adolescents, and 7 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	54	25	14	35
Adolescents	53	5	4	6
Children	53	10	3	7

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

Age group	Number	Variable name	Description
Adult	1	F_SPOUSEYN	Does [fname] have a spouse or co-habitating partner who usually lives in the household or stayed here last night?
	2	H_AGETEENYEARS	1: 15-17; 2: Otherwise; based on AGEYEARS (roster)
	3	H_AGEYEARS	Age (categorical), based on roster age. Matches postsratification cells
	4	H_ECON3	Received economic support in the past 3 months: 1 - Yes; 0 - No
	5	H_HH_SIZE_C	1-9, where 9 includes all HHs with 9 or more people
	6	H_MATRF	Roof material: 1 - no roof, thatch, mat, wood planks, cardboard; 2 - corrugated iron; 9 - other
	7	H_MATWALL	Wall material: 1 - cane, mud; 2 - cement, stone; 3 - bricks; 4 - cement blocks; 9 - other
	8	H_ROOMSLEEP	Number of rooms used for sleeping: 1, 2, 3, 4+
	9	HHELITER	Is household respondent literate?
	10	HHRSCHLHL	What was the highest level of school completed by name
	11	MOSNETS	Does your household have any mosquito nets that can be used while sleeping?
	12	NOEAT4WKFREQ	How often did this happen in the past 4 weeks?
	13	SEX	Roster gender
	14	STRATA	Numeric code for EA sampling stratum
Adolescent	1	HHR12MOAWAY	In the last 12 months, was name away from home for more than one month at a time?
	2	H_TLETTYP	Toilet type: 1 - flush; 2 - latrine; 3 - otherwise
	3	MOSNETS	Does your household have any mosquito nets that can be used while sleeping?
	4	SEX	Roster gender
Children	1	H_HAVE_ELEC_DEVICE	Household has: 1 - electricity and all fridge/tele/radio; 2 - electricity and either fridge/tele/radio; 3 - no electricity but either fridge/tele/radio; 9 - otherwise
	2	NOFOOD4WKYN	In the past 4 weeks, was there ever no food to eat of any kind in your household because of lack of resources to get food?
	3	STRATA	Numeric code for EA sampling stratum

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

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

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, ..., 141 as

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

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, ..., 141.

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, ..., 141 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

Group	Age	Number of interview respondents	Number of adjustment cells	Overall weighted response rate	Minimum adjustment factor	Maximum adjustment factor	Weighted count of respondents before adjustment [1]	Weighted count of respondents after adjustment [2]
Adults	15 or older	11,664	35	90.60	1.00	1.49	630,379	695,791
Adolescents	10-14	1,200	6	90.64	1.04	1.29	125,369	138,318
Children	0-9	2,589	7	96.25	1.00	1.07	275,550	286,278

[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 2017 national population projections by gender and five-year age groups published by the Swaziland Central Statistics Office (CSO). The post-stratified interview weights were computed as follows.

Let N_{ga}^{2017} denote the 2017 Swaziland 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 as:

$$T_{ga}^{2017} = N_{ga}^{2017} / \sum_{k=1}^{n_{ga}^{resp}} W_{gak}^{(int)},$$

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:

$$T_{(r)ga}^{2017} = N_{ga}^{2017} / \sum_{k=1}^{n_{(r)ga}^{resp}} W_{(r)gak}^{(int)}$$

for the $r = 1, 2, \ldots, 141$ jackknife replicates.

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

$$W_{gak}^{(ps-int)} = T_{ga}^{2017} W_{gak}^{(int)},$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-int)} = T_{ga}^{2017} W_{(r)gak}^{(int)}$$

for $r = 1, 2, \ldots, 141$.

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

Table 3-10 2017 Swaziland 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	77,477	71,523	1.0832	74,648	71,717	1.0409	152,125	143,240	1.0620
5-9	71,216	71,957	0.9897	67,920	72,526	0.9365	139,136	144,483	0.9630
10-14	63,855	66,766	0.9564	62,123	66,900	0.9286	125,978	133,666	0.9425
15-19	63,188	65,725	0.9614	64,186	62,957	1.0195	127,374	128,682	0.9898
20-24	59,293	44,420	1.3348	62,471	56,393	1.1078	121,764	100,813	1.2078
25-29	50,075	40,354	1.2409	56,271	51,116	1.1008	106,346	91,470	1.1626
30-34	38,866	35,665	1.0897	49,324	47,445	1.0396	88,190	83,111	1.0611
35-39	29,675	29,458	1.0074	41,613	36,221	1.1489	71,288	65,679	1.0854
40-44	21,745	20,627	1.0542	32,190	26,149	1.2310	53,935	46,776	1.1530
45-49	17,287	15,393	1.1231	25,534	23,652	1.0796	42,821	39,045	1.0967
50-54	13,683	12,676	1.0795	19,139	21,680	0.8828	32,822	34,356	0.9554
55-59	10,988	11,137	0.9866	14,616	17,223	0.8486	25,604	28,361	0.9028
60+	27,461	30,001	0.9153	31,126	50,705	0.6139	58,587	80,706	0.7259
Total	544,809	515,702	1.0564	601,161	604,684	0.9942	1,145,970	1,120,388	1.0228

[1] Source: 2017 Swaziland 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, $W_{gak}^{(int)}$.

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 2017 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 $r = 1, 2, \ldots, 141$), respectively, where k denotes the interview respondent, h denotes the stratum, 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 weighting purposes, even though there were cases that did not complete the interview (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 age group and response status

Group	Age [1]	Gender	Blood Test Status [2]	Count	Weighted count [3]
Adults	15 and older	Male	Respondent	4,506	279,669
			Nonrespondent	398	25,788
		Female	Respondent	6,428	372,730
			Nonrespondent	341	20,810
Adolescents	10-14	Male	Respondent	576	65,587
			Nonrespondent	10	1,179
		Female	Respondent	578	64,875
			Nonrespondent	15	2,025
Children	0-9	Male and female	Respondent	2,218	244,068
			Nonrespondent	383	43,655

[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, selected PSU characteristics such as stratum and urban/rural status, and data from 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. Table 3-12 shows the number of variables used in the models for adults,

adolescents, and children, respectively, and the number of variables identified by the LASSO to be significant predictors of response for the three age groups, respectively. Because LASSO did not select any variables for adolescent females, all variables that were used for the initial LASSO model for adolescent females were added manually to CHAID files for adolescent females. In addition, three variables that were typically selected as significant predictors of response in previous countries, namely, a categorized age based on interview (BEST_AGE), a stratification variable (STRATA) and urban/rural definition (URBAN_RURAL) were added manually for both adolescent males and females. Gender of the responding parent or guardian (variable PROXY_GENDER) was forced into the CHAID tree model for children to make the initial splits, and then the tree was allowed to grow freely. The reason for doing this was because children's parents/guardians completing the interview received different questions based on their gender (see section 3.4.3.2). A complete list of these variables is given in Table 3-13.

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

Group	Age	Gender	Variables in original model	Variables selected by the LASSO	Variables selected by CHAID	Number of nonresponse adjustment cells
Adults	15 and older	Male	128	32	17	29
		Female	137	20	15	38
Adolescents	10-14	Male	108	14	4	5
		Female	105	0	5[1]	7
Children	0-9	Male and female	72	28	13	23

[1] Selected from the 105 variables in the original model.

Table 3-13 summarizes the variables that were included in the final CHAID models. 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	CONFEMAN	Individual's confirmed emancipation status
	2	H_COOKFUEL	Cooking Fuel: 1 - Electricity, 2 - Gas, paraffin; 3 - Coal; 4 - wood, 9 - other'
	3	H_ECON3	Received economic support in the past 3 months: 1 - Yes; 0 - No
	4	H_HAVE_ELEC_DEVICE	Household has: 1 - electricity and all fridge/tele/radio; 2 - electricity and either fridge/tele/radio; 3 - no electricity but either fridge/tele/radio; 9 - otherwise
	5	H_OWNANML	Household owns animals: 1 - cow(s) (and some other animal); 2 - only poultry; 3 - no cow(s) but other animals, including poultry; 9 - doesn't own animals
	6	HEALTHC	Who usually makes decisions about health care for yourself: you, your (spouse/partner), you and your (spouse/partner) together, or someone else?
	7	HHRSCHLHL	What was the highest level of school completed by name
	8	HIVONLY	Only persons who think they might be infected with HIV should get an HIV test.
	9	KNOWN_HIV_STATUS_R	Categorical known HIV status
	10	LOSERESPECT	People who are suspected of having HIV/AIDS lose respect in the community
	11	MCSTATUS	Some men are uncomfortable talking about circumcision but it is important for us to have this information. Some men are circumcised and others are not. Are you circumcised?
	12	Q1060LEV	What is the highest grade/form that you have completed?
	13	SCHLCUR	Are you enrolled in school?
	14	STRATA	Numeric code for EA sampling stratum
	15	WORK30DAYS	Have you done any work in the last thirty days for which you received cash or goods as payment?
	16	WORKOWN	Other (Home/Work) Phone belongs to
	17	WTRSFR	Do you do anything to the water to make it safer to drink?
Adult Female	1	H_COOKFUEL	Cooking Fuel: 1 - Electricity, 2 - Gas, paraffin; 3 - Coal; 4 - wood, 9 - other'
	2	H_ECON3	Received economic support in the past 3 months: 1 - Yes; 0 - No
	3	H_HAVE_ELEC_DEVICE	Household has: 1 - electricity and all fridge/tele/radio; 2 - electricity and either fridge/tele/radio; 3 - no electricity but either fridge/tele/radio; 9 - otherwise
	4	H_OWNTRNSPRT	Household owns transportation: 1 - owns car and no other transport; 2 - owns car and one or more other transport; 3 - does not own car but has other transport; 4 - does not own transport
	5	H_TLETTYP	Toilet type: 1 - flush; 2 - latrine; 3 - otherwise

Age group	Number	Variable name	Description
	6	H_WTRSRC	Water Source: 1 - pipe, 2 - well; 3 - spring/rainwater; 4 - tank; 9 - other
	7	KNOWN_HIV_STATUS_R	Categorical known HIV status
	8	MO12AWAY	In the last 12 months, have you been away from home for more than one month at a time?
	9	NOFOOD4WKYN	In the past 4 weeks, was there ever no food to eat of any kind inf your household because of lack of resources to get food? (hidden)
	10	PREGNANT	Are you or your partner pregnant now?
	11	Q1060LEV	What is the highest grade/form that you have completed?
	12	SCHLHI	What is the highest level of school you attended: primary, secondary, or higher?
	13	STRATA	Numeric code for EA sampling stratum
	14	WORK30DAYS	Have you done any work in the last thirty days for which you received cash or goods as payment?
	15	WORKOWN	Other (Home/Work) Phone belongs to
Adolescent Male	1	ADMEDLL	Are there medicines that people with HIV or AIDS can take to help them live longer?
	2	ADMOSQUITO	Can a person get HIV from a mosquito bite?
	3	H_OWNANML	Household owns animals: 1 - cow(s) (and some other animal); 2 - only poultry; 3 - no cow(s) but other animals, including poultry; 9 - doesn't own animals
	4	H_OWNTRNSPRT	Household owns transportation: 1 - owns car and no other transport; 2 - owns car and one or more other transport; 3 - does not own car but has other transport; 4 - does not own transport
Adolescent Female	1	CH_KIDHIGHLVL	What is the highest level of school ${curchnm}* has attended: primary, secondary or high school?
	2	H_COOKFUEL	Cooking Fuel: 1 - Electricity, 2 - Gas, paraffin; 3 - Coal; 4 - wood, 9 - other'
	3	H_ROOFWALFLR	Roof/wall/floor material: 1 - iron/cane,mud/cement, 2 - iron/cement,stone/cement, 3 - iron/bricks/ceramic tiles, 4 - iron/bricks/cement, 5 - iron/cement blocks/ceramic tiles, 6 - iron/cement blocks/cement, 9 - other
	4	H_SUPPORT	In the last 12 months household received any emotional/material/medical/schooling/social support for kidname: 1 - received any support; 2 - did not receive any support
	5	STRATA	Numeric code for EA sampling stratum
Children	1	AT_BESTAGE_C	Categorical age based on interview age (BEST AGE)
	2	CH_KIDENROLL	Is ${curchnm}* enrolled in school?
	3	CHEAGE	Does the parent/guardian for ${ind_nm}*** confirm that ${ind_nm}*** is ${inageconfp_disp}* years?
	4	H_DADGUARD	Does kid's natural father or male guardian usually live in this HH: 1 - Yes, 2 - No
	5	H_MOMGUARD	Does kid's natural mother or female guardian usually live in this HH: 1 - Yes, 2 - No
	6	H_TLETTYP	Toilet type: 1 - flush; 2 - latrine; 3 - otherwise

Age group	Number	Variable name	Description
	7	HHELITER	Is household respondent literate?
	8	HHRSCHLHL	What was the highest level of school completed by name
	9	HIVONLY	Only persons who think they might be infected with HIV should get an HIV test.
	10	HIVTOPG	During any of your clinic visits for the pregnancy when you were pregnant with ${namedis}*, were you offered an HIV test?
	11	HUSLIVEW	Is your husband or partner living with you now or is he staying elsewhere?
	12	PROXY_GENDER	Gender of responding parent / guardian
	13	STRATA	Numeric code for EA sampling stratum

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

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

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, ..., 141 as

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

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, ..., 141.

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 $r = 1, 2, ..., 141$ 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 three major age groups.

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

Group	Age	Gender	Number of		Overall weighted response rate [1]	adjustment factor		Weighted count of respondents	
			blood test respondents	adjustment cells		Min	Max	before adjustment [2]	after adjustment [3]
Adults	15 and older	Male	4,506	29	91.56	1.00	1.8047	279,669	305,458
		Female	6,428	38	94.71	1.00	1.5228	372,730	393,540
Adolescents	10-14	Male	576	5	98.23	1.00	1.1002	65,587	66,766
		Female	578	7	96.97	1.00	1.2252	64,875	66,900
Children	0-9	Male and female	2,218	23	84.83	1.00	1.9970	244,068	287,723

[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 2017 population projections within classes defined by gender and five-year age group.

Let N_{ga}^{2017} denote the 2017 Swaziland population control total for gender *g* and (five-year) age group *a* as given in Table 3-15. The poststratification ratio adjustment factor used to adjust the blood test weights for gender *g* and age group *a* was computed as:

$$T_{ga}^{2017} = N_{ga}^{2017} / \sum_{k=1}^{n_{ga}^{BT}} W_{gak}^{(BT)},$$

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:

$$T_{(r)ga}^{2017} = N_{ga}^{2017} / \sum_{k=1}^{n_{(r)ga}^{BT}} W_{(r)gak}^{(BT)}$$

for the *r* = 1, 2, …, 141 jackknife replicates.

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

$$W_{gak}^{(ps-BT)} = T_{ga}^{2017} W_{gak}^{(BT)},$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-BT)} = T_{ga}^{2017} W_{(r)gak}^{(BT)}$$

for *r* = 1, 2, …, 141.

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

Table 3-15 2017 Swaziland population projections (overall and by age and gender) and weighted counts of blood test respondents before and after poststratification

	Male			Female			Overall		
Age group	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	77,477	71,177	1.0885	74,648	70,609	1.0572	152,125	141,786	1.0729
5-4	71,216	72,652	0.9802	67,920	73,285	0.9268	139,136	145,937	0.9534
10-14	63,855	66,766	0.9564	62,123	66,900	0.9286	125,978	133,666	0.9425
15-19	63,188	66,302	0.9530	64,186	63,878	1.0048	127,374	130,181	0.9784
20-24	59,293	44,782	1.3240	62,471	56,464	1.1064	121,764	101,246	1.2027
25-29	50,075	40,078	1.2494	56,271	51,441	1.0939	106,346	91,519	1.1620
30-34	38,866	34,749	1.1185	49,324	46,429	1.0624	88,190	81,178	1.0864
35-39	29,675	29,145	1.0182	41,613	36,069	1.1537	71,288	65,215	1.0931
40-44	21,745	20,534	1.0590	32,190	26,318	1.2231	53,935	46,852	1.1512
45-49	17,287	15,752	1.0974	25,534	23,454	1.0887	42,821	39,206	1.0922
50-54	13,683	12,356	1.1074	19,139	21,522	0.8893	32,822	33,878	0.9688
55-59	10,988	11,268	0.9752	14,616	17,330	0.8434	25,604	28,598	0.8953
60+	27,461	30,491	0.9006	31,126	50,633	0.6147	58,587	81,125	0.7222
Total	544,809	516,052	1.0557	601,161	604,332	0.9948	1,145,970	1,120,387	1.0228

[1] Source: 2017 Swaziland 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, $W_{gak}^{(int)}$.

Weights for Analysis of the Violence Module 4

In addition to the analytic weights described in Section 3, a set of special purpose weights was created for analysis of questions in the violence module (VM) of the individual interview. Such weights are required because the relevant questions were administered to random subsamples of individuals who responded to the main interview. The violence module (VM) was administered to women 15 years of age or older, and to adolescents 10-14 years of age.

4.1 Selection Criteria for the Violence Module

- Among adult females aged 15 or older only one was designated in each household to receive the violence module. This designation was implemented after the household roster was completed. The number of adult females eligible for data collection was then determined for each household, and a random number generated by the computer tablet was used to designate one adult female to receive the violence module. Each eligible female within a household had an equal chance of selection. The criteria used to identify persons eligible for the violence module are given in Appendix D.
- All adolescents 10-14 years of age in households designated for child data collection (See Section 2.5.3) were selected to answer portions of the VM.

4.2 Definition of Response Status for the Violence Module

For adult females who were designated to receive the violence module, respondents are those who provided a valid response to the question "Have you ever been touched by someone in a sexual way without your permission?" This question corresponds to the variable TOUCHEVER given in Appendix D. This definition results in an unweighted response rate of 94% (3,888/4,156). Table 4-1 summarizes the number of responses to the adult violence variables.

For adolescents 10-14 years of age who were selected to answer the violence questions, respondents are those who provided a valid answer to the question "Has anyone ever done any of these things to you: - Punched, kicked, whipped, or beat you with an object such as a stick; - Choked smothered, tried to drown you, or burned you intentionally; - Used or threatened you with a knife, gun or other

weapon?" This corresponds to the variable ADATTCK given in Appendix D. This definition results in an unweighted response rate of 89% (1155/1300). Table 4-2 summarizes the number of responses to the adolescent violence variables.

Table 4-1 Distribution of responses to key variables in the violence module for women 15 years or older

Variable	# with valid answer	# ans Yes or >0	yes or >0	# cases Indiv status = 1 or 2	Unwtd RR among Indiv status = 1 or 2
TOUCHEVER	3888	177	yes	4156	94%
TOUCHTIMES	164	146	>0	4156	4%
TOUCHTIMESDK [1]				4156	0%
TOUCHAGE	136			4156	3%
TOUCHAGEDK [1]				4156	0%
FRCSXTIMES	3439	114	> 0	4156	83%
FRCSXTIMESDK [1]				4156	0%
FRCSXAGE	105			4156	3%
FRCSXAGEDK [1]				4156	0%
FRCSX12MO	114	6	yes	4156	3%
FRCSX12MOPT	6			4156	0%
UWNTSXHELP	626	436	Type of help	4156	15%
UNWNTSXNOHLP	165			4156	4%
VLNC12MOTIMES	2424	139	> 0	4156	58%
VLNC12MOPTNR	137			4156	3%
SEEKHELP	133	85	Type of help	4156	3%
SEEKHELPWHYNOT	46			4156	1%

[1] TOUCHTIMESDK asks why TOUCHTIMES was left blank; the only answers are "don't know" or "refused" which we have not considered valid answers for this table. Other DK variables are similar.

Table 4-2 Distribution of responses to key variables in the violence module for adolescents 10-14 years old

Variable	Ages 10 to 14				Ages 13 to 14			
	Resp Rate	N (10 - 14)	# resp	# yes or >0	Resp Rate	N (13 - 14)	# resp	# yes or >0
ADATTCK	89%	1300	1155	102	87%	502	438	33
ADATTKREL	8%	1300	102		7%	502	33	
ADSXLTCH (10 - 12 only)								
Following for (13 - 14 only)								
ADTCHWOPM					86%	502	434	4
ADFTREL					1%	502	3	
ADPRSXSCC					87%	502	439	3
ADFRCSCC					88%	502	442	2
ADFRCREL					0%	502	2	
ADSXTELL					0%	502	2	2
ADPRSV					0%	502	2	2
ADRSPRSVC					0%	502	0	
ADSDSC					0%	502	2	2

4.3 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 W_{jk}^{bw} is the corresponding base weight from the regular weighting process (see Section 3.4.3.1) and N_j^F is the number of eligible women 15+ years of age in household *j* (based on roster) if there were four or less eligible women 15+ years of age in the household or $N_j^F = 4$ it there were five or more eligible women 15+ years of age in the household. The number of eligible women 15+ years of age in the household used to compute the violence module initial 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 in step 5. 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. For adolescents 10-14 years old, $W_{jk}^{viol-bw}$ was set to the corresponding base weight, W_{jk}^{bw}, for the sampled child (see Section 3.4.3.1).

- 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 performed on the sample of persons selected for the violence module, using the same predictors identified for the regular interview weights (see Table 3-8).
- The final cells derived 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 2017 population projections by detailed age groups.

Table 4-2A and 4-2B 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 4-2A List of variables identified by CHAID for adult females 15+ years of age

Variable	Description
H_AGEYEARS	Age (categorical), based on roster age. Matches postsratification cells
H_HAVE_ELEC_DEVICE	Household has: 1 - electricity and all fridge/tele/radio; 2 - electricity and either fridge/tele/radio; 3 - no electricity but either fridge/tele/radio; 9 - otherwise
H_HH_SIZE_C	1-9, where 9 includes all HHs with 9 or more people
H_MATWALL	Wall material: 1 - cane, mud; 2 - cement, stone; 3 - bricks; 4 - cement blocks; 9 - other
H_OWNTRNSPRT	Household owns transportation: 1 - owns car and no other transport; 2 - owns car and one or more other transport; 3 - does not own car but has other transport; 4 - does not own transport
H_WTRSRC	Water Source: 1 - pipe, 2 - well; 3 - spring/rainwater; 4 - tank; 9 - other
STRATA	Numeric code for EA sampling stratum

Table 4-2B List of variables identified by CHAID for adolescents 10-14 years of age

Variable	Description
H_TLETTYP	Toilet type: 1 - flush; 2 - latrine; 3 - otherwise

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	3,011	282,452	2,810	260,736	276,372	331,589
Females 50+	1,145	107,075	1,078	100,446	106,545	64,881
Adolescents 10-14	1,300	134,709	1,155	118,971	134,662	125,978
Total	5,456	524,237	5,043	480,153	517,579	522,448

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

SHIMS (2012), Swaziland HIV Incidence Measurement Survey, First Findings Report, November 2012: https://www.k4health.org/sites/default/files/SHIMS_Report.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 was implemented by trained field staff using computer tablets. The aim in establishing eligibility was to make sure that all potentially-eligible dwelling units (e.g., including vacants or buildings under construction) are given appropriate chances of selection for the study. Based on three variables recorded for each listing in the computer tablets (the structure type, whether the structure was vacant or under construction, and whether the structure was occupied or not), an eligibility flag (ELIG_FLAG) was assigned to each combination of values of the three variable as either being eligible for the study (ELIG_FLAG = Y) or not (ELIG_FLAG = N).

Table A-1 shows all possible combinations of the three relevant variables used to define eligibility status and the corresponding counts of records in the Master Listing File. Table A-2 contains a detailed description of the three variables.

Of the 51,578 original listing records, two were coded as "discard" for various reasons, and not eligible for sample. This left a total of 51,576 listing records in the sample of 287 PSUs. Of these, 75 were classified as ineligible for sampling based on the structure type, vacancy status, and residential status.

Thus, 51,501 listing records in the Master Listing File were eligible for sampling. Examples of the eligible listing records are:

The 43,089 listings coded as 1, 1, 1 (that is, a single House/compound of HH; Not Vacant and not under construction; Occupied as a residence) are eligible for sampling (ELIG_FLAG=Y).

The 207 listings with codes 1, 1, 2 (that is, a single House/compound of HH; Not Vacant and not under construction; Not being occupied as a residence) are also be eligible for sampling since they appear to be dwelling units (DUs) with no one living there at the time of listing (but could potentially have occupants at the time of interview).

The 259 listings with codes 1, 2, 1 (a single House/compound of HH; Vacant; Occupied as a residence), though apparently contradictory appear to be potential HHs, and were made eligible for sampling.

The 1,427 listings with codes 1, 2, 2 are "vacants" with no one currently living there, but could have residents at the time of interview, and so they were considered eligible for sampling.

The 70 cases with codes 1, 3, 1 are currently under construction, but appear to have people living there and were also considered eligible for sampling.

The 744 listings with codes 1, 3, 2 are under construction and have no one currently living there. Although it is not possible to ascertain whether these dwelling units will have been completed at the time of interview, they are considered eligible to avoid undercoverage as CSO staff mentioned a construction boom.

The one listing with codes 1, 1, Missing (that is, a single House/compound of HH; Not Vacant and not under construction; Unknown occupancy status) was considered eligible because GPS data are available.

The one listing with codes 1, Missing, 1 (that is, a single House/compound of HH; Unknown Vacancy/Construction status; Occupied as a residence) was considered eligible because GPS data are available.

Table A-1 Definition of eligibility and number of records by eligibility status

Eligibility (ELIG_FLAG)	Structure Type (STOBS_D)	Structure vacant or under construction? STVAC_D	Anyone living in the structure? RESYN_D	Number of Records
Records coded as "Discarded" in the master listing file				2
Other ineligible listings in eligible EAs				75
N				13
N			2	1
N	3	1	2	4
N	3	2	2	13
N	3	3	2	10
N	4	1	2	2
N	4	2	2	1
N	5	3	2	3
N	6	1	2	6
N	6	2	2	8
N	7	2	2	4
N	8	2	2	3
N	8	3	1	1
N	8	3	2	6
TOTAL ELIGIBLE LISTINGS				51,501
Y	1		1	1
Y	1	1		1
Y	1	1	1	43,089
Y	1	1	2	207
Y	1	2	1	259
Y	1	2	2	1,427
Y	1	3	1	70
Y	1	3	2	744
Y	3	1	1	38
Y	4	1	1	4
Y	5	1	1	455
Y	5	1	2	4
Y	5	2	1	5
Y	5	2	2	20
Y	6	1	1	131
Y	7	1	1	91
Y	8	1	1	75
Y	9	1	1	4,538
Y	9	1	2	11
Y	9	2	1	21
Y	9	2	2	292
Y	9	3	1	1
Y	9	3	2	17
TOTAL NUMBER OF RECORDS IN THE MASTER LISTING FILE				51,578

Table A-2 Definition of variables used to define eligibility status

Structure Type (STOBS_D)	
1	Single House/compound of hh
2	Apartment bldg./gated comm.
3	Church/mosque/temple
4	Community center
5	School/University
6	Shop/business ctr/commerce bldg.
7	Clinic/hospital/dr.office
8	Other
Structure vacant or under construction? (STVAC_D)	
1	Not Vacant and not under construction
2	Vacant
3	under construction
Anyone living in the structure? (RESYN_D)	
1	Yes
2	No

APPENDIX B

Definition of Household, Interview, and Blood Test Response Status

Definition of Household, Interview, and Blood Test Response Status

The response status variables required for weighting as previously described in Section 3.4.2.1 (household weights), Section 3.4.3.1 (interview weights), and Section 3.4.4.1 (blood test weights) were created using the SAS program code given below. In general, a response code of 1 is assigned to respondents, 2 to (eligible) nonrespondents, 3 to ineligible/out-of-scope cases, and 4 to cases for which eligibility is unknown.

B.1 SAS Code for HH_STATUS

```
attrib HH_eligible length=3 label="Household Eligibility flag – will be used to create
HH_STATUS_0";

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_0 length=3 label="Intermediate HH disposition code";

if HH_eligible = 1     then HH_STATUS_0=1; /*Eligible Respondent*/
   else if HH_eligible in(2,3) then HH_STATUS_0=2;/*Eligible NonRespondent*/
   else if HH_eligible = 4 then HH_STATUS_0=3; /* Ineligible */
   else if HH_eligible = 5 then HH_STATUS_0=4; /*Unknown eligibility Status*/

if HH_ELIGIBLE = 2 and ROSTERCOUNT > 0 then HH_STATUS_0 = 1;    /* Eligible
Respondent */

if HH_ELIGIBLE = 5 and UPCODE_STAT_HH in (2,3,4) then HH_STATUS_0 =
UPCODE_STAT_HH;
```

```
if EA_HHID_FIXED in ('080807110081012034', '050504079212011036') then do;
        HH_STATUS_0 = 4;
              HH_eligible = 5;
              end;

IF UPCODE_RSLTNDT='96' THEN HH_STATUS_0=4;
IF HH_STATUS_0=. THEN HH_STATUS_0=4;
    HH_STATUS=HH_STATUS_0;
```

Notes:
The variable ROSTERCOUNT is created earlier in the program; it counts the number of non-empty individual records on the roster file for each value of EA_HHID_FIXED. Households with no questionnaire record but with at least one valid roster record are eligible respondent households.

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. It is used to assign nonresponding households where RESULTNDT = 10 "other, specify" into the three categories of households with no response.

B.2 SAS Code for INDIV_STATUS

```
label CHILD_SMPYN = "CHILD IS SAMPLED Y/N"

if 0 <= AGEYEARS <= 14 and
  CHILD_SMPFLG   = "1"  then CHILD_SMPYN = 1;
else
 CHILD_SMPYN = 0;

label INDIV_AGEGROUP = "INDIVIDUAL AGE GROUP BASED ON BEST_AGE"

    IF CONFAGEY_RECODE ^= . THEN DO;
        BEST_AGE = CONFAGEY_RECODE;
    END;
    ELSE DO;
        BEST_AGE = AGEYEARS;
    END;

    IF GENDR ^= . THEN DO;
       BEST_GENDER = GENDR;
    END;
```

```
    ELSE DO;
       BEST_GENDER = SEX;
    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;

label     INDIV_STATUS = "Individual Response Status"
          INDIV_QXSTATUS = "Completion of questionaire";

if (INDIV_AGEGROUP=1 and (CH_KIDAGEY => 0 and CH_KIDGENDER > "0") and
          (CH_KIDENROLL > "0" or CH_KIDHIVTESTEVR > "0" or CH_KIDVISTTBCLIN
          > "0")) or
    (INDIV_AGEGROUP=2 and INDFINRSLT in ("1","2") and ADOLTSEND ^= .) or
    (INDIV_AGEGROUP=3 and INDFINRSLT in ("1","2") and MILESTONEA1 ="1" and
          MILESTONEA2 ^= "2" and MILESTONEA3 ^= "2" and MILESTONEA4 ^= "2" and
          MILESTONEA5 ^= "2" and MILESTONEA7 ^="2" and MILESTONEA8 ^= "2" and
          MILESTONEA9 ^= "2" and MILESTONEA10 ^= "2" and MILESTONEA11 ^= "2"
          and MILESTONEA12 ^= "2") Then INDIV_QXSTATUS = 1;
 else
   INDIV_QXSTATUS = 0;

IF SLEEPHERE=2 then INDIV_STATUS =9;
ELSE
 if 0<= AGEYEARS <=14 and CHILD_SMPYN = 0 then INDIV_STATUS = 8;
 ELSE
  if upcase(HIV1STATUSFINALSURVEY) in ("NEGATIVE", "POSITIVE") OR
(INDIV_QXSTATUS =1) then INDIV_STATUS = 1;
   ELSE
    INDIV_STATUS = 2;

IF UPCODE_STAT ^= . THEN DO;
      If UPCODE_STAT in (6, 7) then INDIV_STATUS = UPCODE_STAT;
      Else
       if INDIV_AGEGROUP =1  and
          INDIV_STATUS  ^= 9 and
          UPCODE_STAT    = 9 then INDIV_STATUS = UPCODE_STAT;
       Else
        if INDIV_AGEGROUP in (2, 3) and
          INDIV_STATUS not in (5, 8, 9) then INDIV_STATUS = UPCODE_STAT;
end;

IF BEST_AGE = . and
  BEST_GENDER = " " THEN INDIV_STATUS = 7;
```

B.3 SAS Code for BT_STATUS

```
ATTRIB BT_STATUS

LABEL="Blood test disposition code:
  1=YES (valid lab results),
  2=NO (no valid lab results or didn't do BT)";

   IF HIV1statusfinalsurvey IN('Positive','Negative') THEN BT_STATUS=1;
   ELSE BT_STATUS=2;
```

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 subsets that varied 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)

TN_INDIV_STATUS.pdf (Adolescents 10-14 years)

CH_INDIV_STATUS.pdf (Children 0-9 years)

Blood Test

AM_BTEST.pdf (Males 15+ years)

AF_BTEST.pdf (Females 15+ years)

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

Swa_AD_INDIV.xlsx (Persons 15+ years)

Swa_TN_INDIV.xlsx (Adolescents 10-14 years)

Swa_CH_INDIV.xlsx (Children 0-9 years)

Blood Test

Swa_AM_BT.xlsx (Males 15+ years)

Swa_AF_BT.xlsx (Females 15+ years)

Swa_TM_BT.xlsx (Males 10-14 years)

Swa_TF_BT.xlsx (Females 10-14 years)

Swa_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 List of the questions in the adult questionnaire's violence module (Swaziland)

Variable	Question Text
touchever	Have you ever been touched by someone in a sexual way without your permission?
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 the previous question was left blank
touchage	How old were you the first time this happened?
touchagedk	Please provide the reason the previous question was left blank
frcsxtimes	How many times in your life have you been physically forced to have sex?
frcsxtimesdk	Please provide the reason the previous question was left blank
frcsxage	How old were you the first time someone physically forced you to have sex?
frcsxagedk	Please provide the reason the previous question was left blank
frcsx12mo	In the past 12 months, did someone physically force you to have sex?
frcsx12mopt	In the past 12 months, who physically forced you to have sex?
frcsx12mowhooth	Specify other (PII, not in data file)
uwntsxhelp	After any of these unwanted sexual experiences, did you try to seek professional help or services from any of the following? (data file variables UWNTSXHELP_A, _B, _C, _D, _E, _X, _Y, _Z
nosekoth	Specify other (PII, not in data file)
unwntsxnohlp	What was the main reason that you did not try to seek professional help or services?
unwntsxnohlpoth	Specify other (PII, not in data file)
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 - Choked, smothered, tried to drown you, or burned you intentionally - Used or threatened you with a knife, gun or other weapon?
vlnc12moptnr	In the last 12 months, who did any of these things to you?
vlnc12moith	Specify other (PII, not in data file)
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?
sekhlpsp	Specify other (PII, not in data file)
seekhelpwhynot	What was the main reason that you did not try to seek help or services?
seekhelpwhynototh	Specify other (PII, not in data file)

D-2 List of the questions in the adolescent questionnaire's violence module (Swaziland)

Variable	Question Text
Adattck (all 10 - 14)	Has anyone ever done any of these things to you:
	- Punched, kicked, whipped, or beat you with an object such as a stick
	- Choked smothered, tried to drown you, or burned you intentionally
	- Used or threatened you with a knife, gun or other weapon?
Adattkrel (follow-up to Adattck)	The first time this happened, what was your relationship to the person who did this? If it was more than one person, what was your relationship with the person you knew the best?
adattkreloth	Specify other (PII, not on data set)
Adsxltch (asked only of 10 - 12)	Has anyone ever touched your ${viol_area}* when you did not want them to? This could be pinching, grabbing, or touching you on or around your ${viol_area}*.
Adtchwopm (asked only of 13 - 14)	Has anyone ever touched you in a sexual way without your permission, but did not try and force you to have sex?
Adftrel (asked only of 13 - 14), follow-up to Adtchwopm	The first time this happened, what was your relationship to the person who did this? If it was more than one person, what was your relationship with the person you knew the best?
adftrelsp	Specify other(PII, not on data set)
Adprsxscc (asked only of 13 - 14)	Has anyone ever pressured you to have sex, through harassment, threats or tricks and did succeed?
Adfrcscc (asked only of 13 - 14)	Has anyone ever physically forced you to have sex and did succeed?
Adfrcrel (asked only of 13 - 14) Follow-up	The first time you were pressured or forced to have sex, what was your relationship to the person who did this?
adfrcrelsp	Specify other(PII, not on data set)
Adprsv (asked only of 13 - 14) Follow -up	After any of these unwanted sexual experiences, did you try to seek help or services from any of the following?
adprsvsp	Specify other(PII, not on data set)
Adrsprsvc (asked only of 13 - 14) Follow-up	What was the main reason that you did not try to seek professional help or services?
adrsprsvspy	Specify other(PII, not on data set)
Adsxtell (asked only of 13 - 14) Follow-up	After any of these unwanted sexual experiences, did you tell anyone about it?
Adsdsc (asked only of 13 - 14) Follow-up	Which of the following describes who you told about any of these unwanted sexual experience? (
adsdscspy	Specify other(PII, not on data set)

D.3 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.4 Code to Define Violence Module Status (VM_STATUS)

```
DATA HH_QX;
        SET INIT.HH_QX;
RUN;

PROC SORT DATA = HH_QX;
        BY EA_HHID_FIXED;
RUN;

PROC SORT DATA = INIT.ROSTER OUT = 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 Violence Module Selection Flag"

        IF AA AND BB;

        IF (AA AND BB) AND (LINENUM=VIOLFLAG) THEN VM_FLAG = 1;
        ELSE VM_FLAG = 0;
RUN;

PROC SORT DATA = INDIV.W30_INDIV_QX_REDUCED OUT = INDIV;
        BY EA_HHID_FIXED;
RUN;

DATA INDIV;
        MERGE INDIV (IN=AA) NEW_ROSTER;
        BY EA_HHID_FIXED;
        IF AA THEN OUTPUT;

        LABEL VM_STATUS  = "Adult Violence Module Weighting Disposition Code"

        IF TOUCHEVER IN ('1','2') THEN VM_QXSTATUS = 1;
        ELSE VM_QXSTATUS = 0;

        IF VM_FLAG = 0 THEN VM_STATUS = 0;
        ELSE IF INDIV_STATUS NOT IN (1,2) THEN VM_STATUS = INDIV_STATUS;
        ELSE IF CONFGEND = '1' THEN VM_STATUS = 3;
        ELSE IF INDIV_AGEGROUP IN (1,2) THEN VM_STATUS = 4;
        ELSE IF VM_QXSTATUS =1 THEN VM_STATUS = 1;
        ELSE VM_STATUS = 2;
RUN;
```