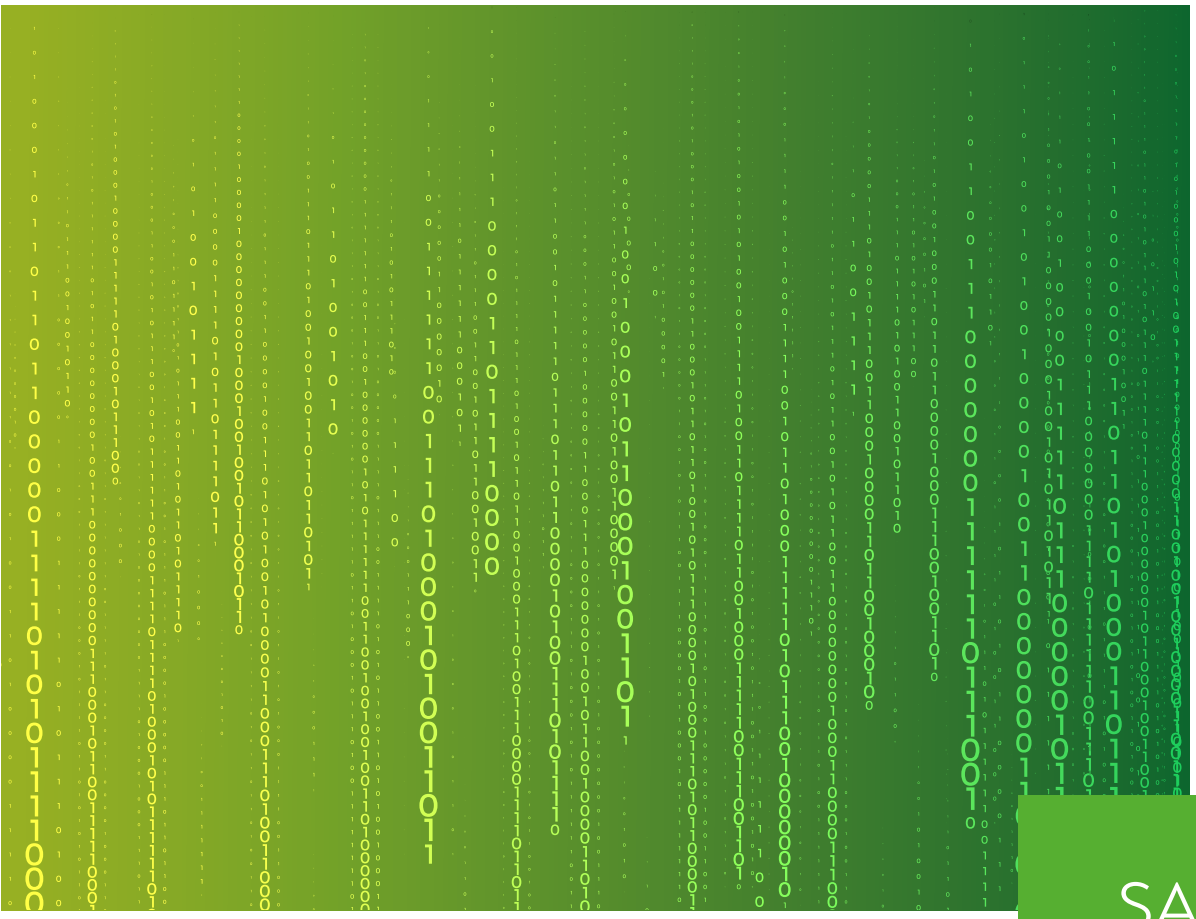
SHIMS3 2021

Eswatini Population-based HIV Impact Assessment



SAMPLING AND WEIGHTING TECHNICAL REPORT

SIYINQABA

Eswatini Population-based HIV Impact Assessment 2021

SHIMS3 2021

This project is supported by the US President's Emergency Plan for AIDS Relief (PEPFAR) through CDC under the terms of cooperative agreement U2GGH002173. The findings and conclusions are those of the authors and do not necessarily represent the official position of the funding agencies.

Table of Contents

Contents Continued

Acronyms

CDC	Centers for Disease Control and Prevention
CHAID	Chi-square Automatic Interaction Detector
CI	Confidence Interval
CV	Coefficient of Variation
DEFF	Design Effect
DU	Dwelling Unit
EA	Enumeration Area
HH	Household
HIV	Human Immunodeficiency Virus
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
PSU	Primary Sampling Unit
RSE	Relative Standard Error
SAS	Statistical Analysis System
SHIMS3	Swaziland HIV Incidence Measurement Survey
CSO	Central Statistics Office
UEW	Unequal Weighting
VLS	Viral Load Suppression
WLM	Weighted Log-linear Modeling

1. Introduction

The 2021 Swaziland HIV Incidence Measurement Survey (SHIMS3 2021) is a cross-sectional sample survey designed to assess the prevalence of key human immunodeficiency virus (HIV)-related health indicators among individuals 15 years or older. Data collection for SHIMS3 2021 was conducted between May 2021 and November 2021 with a temporary pause in data collection from June 2021 to August 2021 and again in October 2021 due to civil unrest within the country. The survey included approximately 12,000 interviewed individuals and approximately 11,200 individuals with valid blood tests in approximately 5,400 randomly-selected 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 SHIMS3 2021 is a stratified multistage probability sample design, with eight strata defined as the urban and rural portions of each of the four regions within the country, first-stage sampling units defined by enumeration areas (EAs) within strata, second-stage sampling units defined by households within EAs, and finally age-eligible persons within households. Within each sampling stratum, the first-stage sampling units (also referred to as "primary sampling units" or PSUs) were selected with probabilities proportionate to updated numbers of households in the PSU derived from the 2017 Eswatini Population and Housing Census. The allocation of the sample PSUs to the eight strata was made in a manner designed to achieve specified precision levels for: (a) national estimate of HIV incidence among persons 15 to 49 years old; (b) national and regional estimates of viral load suppression (VLS) rates among HIV-positive persons 15 to 49 years old; and (c) national and regional estimates of viral load suppression (VLS) rates among HIV-positive females 15 to 24 years old.

The second-stage sampling units were selected from lists of households and unoccupied dwelling units (referred to throughout as dwelling units/households) compiled by trained staff for each of the sampled PSUs. Upon completion of the listing process, random samples of specified numbers of dwelling units/households were selected from each PSU.

Within the responding households, all eligible persons 15 years of age and older who were present in the household on the night prior to the household interview were included in the study sample for SHIMS3 2021 data collection.

Details of the sample design employed for SHIMS3 2021 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 across 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 include

- 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, and to compensate for PSU nonresponse;
- 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;
- Post-stratification of the person-level interview weights to calibrate the weighted counts of persons completing the interview so that they match external population counts to adjust for potential undercoverage;
- 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

- Post-stratification of the person-level blood test weights to calibrate the weighed counts of persons so that they match external population counts to adjust for potential undercoverage.

Technical details of the weighting procedures employed for SHIMS3 2021 are provided in Section 3.

2. Sample Design

2.1 Population of Inference

The population of inference for SHIMS3 2021 is comprised of the *de facto* population of individuals 15 years of age and older. The *de facto* population is comprised of all individuals who were present in households (i.e., "slept in the household") on the night prior to the date of interview. 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. Individuals belonging to either the *de facto* or *de jure* populations were included on the rosters compiled for sampling purposes; however, only members of the *de facto* population were eligible for data collection. Table 2-1 summarizes estimates (projections) of the 2021 Eswatini population by gender and age group.

Table 2-1 2021 population estimates for Eswatini by gender and age group

Age group	Gender		Total
	Male	Female	
15 to 49 years	292,747	302,422	595,169
50 years or older	62,378	88,025	150,403
Total	355,125	390,447	745,572

Source: 2021 Population Projections from CSO website - https://eswatini.unfpa.org.

2.2 Precision Specifications and Assumptions

The following specifications and assumptions were used to develop the sample design for SHIMS3 2021.

2.2.1 Specifications

- Relative standard error (RSE) of the national estimate of HIV incidence among adults 15 to 49 years old should be 30% or less
- 95% confidence interval (CI) bounds around the estimated VLS rate among HIV positive adults aged 15 to 49 years for each of the four regions should be +/-0.10 or less
- 95% CI bounds around the national estimate of VLS rate among all HIV positive adults aged 15 to 49 years should be of +/-0.03 or less

- 95% CI bounds around the national estimate of VLS rate among all HIV positive females aged 15 to 24 years should be +/-0.06 or less

2.2.2 Statistical Assumptions

- National HIV prevalence rate of 0.272 (27.2%) for adults 15-49 years old that varies by region and urban/rural status (see Table 2-2);
- National HIV prevalence rate of 0.139 (13.9%) for women aged 15 to 24 years old that varies by region and urban/rural status (see Table 2-2);
- Annual national incidence rate for adults aged 15-49 of $p_a = 0.0128$ (1.28%);
- Stratum-level (region by urban/rural status) incidence rates of p_{ah}, $h = 1, 2, \ldots, 8$, which are obtained by adjusting the national incidence rate using the stratum-level (zonal) prevalence rates as follows:

$$p_{ah} = (p_h/p)\, p_a ,$$

  where p_h and p are the HIV prevalence rates for stratum h and the country, respectively, and p_a is the annual national incidence rate;
- Mean duration of recent infection (MDRI) of 130 days, yielding an annualization rate of 365/130= 2.8077;
- Estimated incidence rate for MDRI = 130 days of $p_m = 0.0128/2.8077 = 0.0046$ (0.46%), and the corresponding stratum-level (region by urban/rural status) estimates obtained by $p_{mh} = p_{ah}/2.8077$;
- Viral load suppression rate among HIV positive adults aged 15-49 of $p_{VLS} = 0.50$ (50%) in each region, which results in a conservative estimate of the underlying population variance associated with VLS rate;
- Intracluster correlation (ICC) of 0.033 for VLS and 0.039 for prevalence;
- ICC of 0.000 for incidence (Source: analyses of prior PHIA surveys);
- Overall sex-age distributions;
- Regional population distribution obtained from Swaziland Population Projections 2007-2030, Swaziland Central Statistics Office (CSO); and
- Estimates from SHIMS2 2016-2017 were used to inform the statistical assumptions.

2.2.3 Operational Assumptions

- Varying numbers of dwelling units/households to be sampled per PSU, resulting in an average of 35 sampled dwelling units/households per PSU;
- Occupancy rate of 96.2% for sampled dwelling units/households;
- Household response rate of 84.0% among occupied households;
- Average household size of 3.86 *(de facto)* persons per household;
- Overall percentage of *de facto* persons 15-49 years of age per household of 50.4%; and an overall percentage of *de facto* persons 50+ years of age of 12.9%;
- Within the responding households, a person-level interview response rate of 90.0%;
- Among persons completing the interview, a blood test response rate of 93.3%. Thus, among the persons selected for SHIMS3 2021, the assumed overall response rate for the blood tests is 84.0% (90.0% * 93.3%); and
- Estimates from SHIMS2 2016-2017 were used to inform the operational assumptions.

Based on the specifications and assumptions listed above, a sample of 200 EAs (clusters) was determined to be the minimum needed to meet the specified precision goals. The allocation of the sample to the eight strata of Eswatini 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 this table. The actual numbers of respondents achieved 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 those 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/households and projected sample sizes (expected number of respondents) by stratum

Stratum code	Stratum name	HIV prevalence rate[1]		Total Number of sample clusters	Target Number of DUs/HHs to be sampled	Number of participating HHs[2]	Projected Number of respondents[3]	
		Adults 15-49	Females 15-24				Adults 15-49	Adults 50+
1	Hhohho Rural	0.251	0.133	39	1,365	1,103	1,803	461
2	Hhohho Urban	0.266	0.141	18	630	509	832	213
3	Lubombo Rural	0.275	0.144	35	1,225	990	1,618	414
4	Lubombo Urban	0.387	0.203	7	245	198	324	83
5	Manzini Rural	0.262	0.130	38	1,330	1,075	1,757	449
6	Manzini Urban	0.295	0.146	28	980	792	1,295	331
7	Shiselweni Rural	0.265	0.134	31	1,085	877	1,433	366
8	Shiselweni Urban	0.288	0.146	4	140	113	185	47
All	Eswatini	0.272	0.139	200	7,000	5,657	9,248	2,363

DU = dwelling unit; HH= household

[1] Source: SHIMS2 2016-2017.

[2] Assumes occupancy rate of 96.2% and household response rate of 84.0%.

[3] Projected numbers of individuals providing valid blood draw based on assumptions used to develop the sample design.

2.3 Selection of the Primary Sampling Units (PSUs)

2.3.1 Definition of PSUs

In SHIMS3 2021 the first-stage sampling units, PSUs, were Central Statistics Office (CSO) enumeration areas (EAs). The term PSU is the more general statistical term. The first-stage SHIMS3 2021 sample was selected from a sampling frame of EAs that originally had been created for the 2017 Eswatini Population Census, and subsequently updated by CSO sometime prior to August 2019. The EAs in the updated sampling frame were generally the same as those created for the 2017 Population Census, except that some EAs were subdivided into separate EAs. The updated sampling frame consisted of slightly over 2,260 EAs containing an estimated 264,856 households as of 2019.

2.3.2 Selection of the PSU Sample

A stratified sample of 200 EAs was selected from the updated EA sampling frame in accordance with the sample allocation given in Table 2-2. The following procedure was used to select the EAs

for SHIMS3 2021. Within each stratum, the EAs in the updated sampling frame were sorted in the same way they had been sorted in the SHIMS2 2016-2017 frame to the extent feasible; i.e., by inkundla within region, ward within inkundla and finally by EA within ward[1]. The sorting of EAs prior to sample selection induces an implicit geographic substratification within each stratum.

Next, a systematic sample of the EAs was selected from each stratum. The EAs were selected with probabilities proportionate to a measure of size (MOS) equal to the estimated number of households in the EA in 2017. To select the sample from a given stratum, the cumulative MOS was determined for each EA in the ordered list of EAs, and the sample selections were designated using a random start and a sampling interval equal to the total MOS of the EAs in the stratum divided by the number of EAs to be selected. The resulting sample has the property that the probability of selecting an EA within a stratum is proportional to the MOS of the EA.

Of the 200 EAs selected using this method, 38 had been selected previously for the SHIMS2 2016-2017. Following recommendations by CSO, none of the 38 overlapping EAs was replaced by another EA.

2.3.3 Out-of-Scope PSUs

Out-of-scope PSUs are defined to be those EAs with no households (e.g., EAs that are no longer occupied due to flooding or other natural disasters, or where all residents have been permanently relocated). These are also sometimes referred to as "empty" PSUs. There were no out-of-scope PSUs in SHIMS3 2021 sample.

2.3.4 Non-responding PSUs and Substitution

A sampled PSU that contains eligible households is considered nonresponding if it cannot be entered (e.g., roads/bridges or other means of entry are temporarily closed, access points are flooded, the area contains army barracks or government facilities for which entry is prohibited), is subject to military conflict or other dangerous conditions, or if permission to visit sampled areas is

[1] A 9-digit EA code defined by CSO with region (1 digit), inkundla (2 digits), ward (2 digits) and EA number (4 digits) was used.

not received when such approval is needed. There were no nonresponding PSUs in the SHIMS3 2021 sample.

2.3.5 Summary of the PSU Sample

As indicated in the previous sections, 200 PSUs (EAs) were selected for SHIMS3 2021. There were no out-of-scope (ineligible) or nonresponding PSUs. Table 2-3 summarizes the distribution of the sampled PSUs by stratum and sampling status of the PSU.

Table 2-3 Distribution of sample PSUs by stratum and PSU sampling status

Stratum code	Stratum name	Sampled PSUs	Nonresponding PSUs excluded from 2nd stage DU/HH selection	Ineligible PSUs excluded from 2nd stage DU/HH selection	Number of in-scope PSUs included in study
1	Hhohho Rural	39	0	0	39
2	Hhohho Urban	18	0	0	18
3	Lubombo Rural	35	0	0	35
4	Lubombo Urban	7	0	0	7
5	Manzini Rural	38	0	0	38
6	Manzini Urban	28	0	0	28
7	Shiselweni Rural	31	0	0	31
8	Shiselweni Urban	4	0	0	4
All	Eswatini	200	0	0	200

DU = dwelling unit; HH= household

2.4 Selection of Households

The selection of dwelling units/households for SHIMS3 2021 involved the following steps: (1) listing all potentially eligible dwelling units/households within the sampled EAs, (2) assigning eligibility codes to the listed dwelling unit/household records based on characteristics of the listed units, and (3) selecting the sample of dwelling units/households from those records determined to be eligible for selection.

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 EA. For the purpose of PHIA, the sampling units are households. When vacancy of a "dwelling unit" cannot be determined, the "dwelling unit" is included on the household sampling frame. Therefore, the sampling frame and the sample of second-stage sampling units is a mixture of households and dwelling units.

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. Listers were provided with maps from which to delineate the boundaries of the EA, and to record the 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 (where available), the type of structure (house, apartment, compound, etc.), occupancy status, and GPS coordinates. Vacant structures were listed along with occupied households. Slightly over 39,600 records eligible for sampling were listed for SHIMS3 2021.

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 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 the listing file. 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 second-stage sampling. Table 2-4 summarizes the total number of records entered into

the tablets, the numbers of unoccupied dwelling units, households eligible for sampling, and the total number of dwelling units/households (records) eligible for sampling.

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

Stratum code	Stratum name	Number of records (DUs/HHs) in listing file[1]	Number of unoccupied DUs/HHs[2]	Number of unoccupied DUs/HHs eligible for sampling[3]	Number of occupied DUs/HHs eligible for sampling[4]	Total number of DUs/HHs eligible for sampling
1	Hhohho Rural	6,255	252	251	6,001	6,252
2	Hhohho Urban	2,866	71	70	2,795	2,865
3	Lubombo Rural	5,436	254	252	5,177	5,429
4	Lubombo Urban	1,815	41	41	1,774	1,815
5	Manzini Rural	6,755	352	352	6,403	6,755
6	Manzini Urban	10,437	316	316	10,121	10,437
7	Shiselweni Rural	5,134	209	208	4,925	5,133
8	Shiselweni Urban	933	21	21	911	932
All	Eswatini	39,631	1,516	1,511	38,107	39,618

, DU = dwelling unit; HH= household

[1] See Appendix A for additional details.

[2] Records coded as vacant, under construction, or with no residents at time of listing.

[3] Subset of the unoccupied DUs that could potentially become residential units by the time of data collection.

[4] All records not coded as vacant, under construction, or with no residents at the time of listing.

2.4.4 Selection of Dwelling Units/Households

In order to achieve equal-probability samples of dwelling units/households within each of the eight sampling strata, the sampling rates required to select dwelling units/households within a PSU depend on the difference between the size measure used in sampling (i.e., the estimated number of households in the PSU based on the most recent census projections) and the actual number of dwelling units/households found at the time of listing which took place between October and November 2019. Thus, application of the within-PSU sampling rates based on the size measure used in sampling can yield more than the desired number of dwelling units/households in PSUs that have experienced growth in population since the time of the latest census projections, and fewer than the desired number of dwelling units/households in PSUs that have declined in population.

The calculation of the required within-PSU sampling rates proceeded as follows. First, the target overall sampling rate for stratum h = 1, 2, ..., 8, was computed as:

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

where

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

m_h = number of sample PSUs in stratum *h*;

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

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

The total *expected* number of listings to be selected across all eight strata is $\sum_{h=1}^{8} T_h = 7{,}000$ (see Table 2-2). To obtain an equal probability sample within stratum *h*, the required within-PSU sampling rate for PSU *i* in stratum *h* was then computed as:

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

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

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

To reduce the variation in workload across the sampled PSUs, the maximum number of dwelling units/households to be selected in any PSU was capped at 70 and the minimum number was set to 15. Inspection of the values of $\mathrm{E}(n_{hi})$ indicated that the expected sample sizes for one PSU would fall below 15, and one would exceed 70. The difference between the number of dwelling units/households that would have been selected using the rates, f_{hi}^{within}, and the specified maximum and minimum number was then re-distributed to the other PSUs in the same stratum so as to maintain as closely as possible the desired total sample size for the stratum. The within-PSU sampling rates, f_{hi}^{within}, were therefore adjusted to account for the redistribution of the sample within the stratum. The adjusted within-PSU sampling rate used to select the sample of dwelling units/households, $f_{hi}^{adj(w)}$, was calculated as:

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

where the adjustment factors, A_{hi}, were determined such that

$$\mathrm{L} \leq N_{hi} A_{hi} f_{hi}^{within} \leq \mathrm{U},$$

L = 15 = the minimum PSU sample size,

U = 70 = the maximum PSU sample size, and

$\sum_{i=1}^{m_h} A_{hi} f_{hi}^{within} = T_h$.

To achieve a geographical ordering of the listed dwelling units/households, the dwelling unit/household records in each PSU were sorted by a proximity variable that indicated the geographic distance between the listed dwelling unit/household and the dwelling unit/household closest to the centroid of the PSU. 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.

2.4.5 Results of Second-Stage Sampling

Table 2-5 summarizes the numbers of dwelling units/households selected for the study and the minimum and maximum PSU sample size by stratum. The last column shows the unequal weighting (UEW) design effects (DEFF) to be expected for the selected sample. The UEW DEFF provides a measure of the increase in the variance of a sample-based estimate resulting from the use of variable overall sampling rates within a stratum (e.g., see Kish, 1965, page 403). With an equal-probability sample within each stratum, the DEFFs would ordinarily equal 1.0. Variable sampling rates within stratum will increase the DEFF, which would arise, for example, from the capping of sample sizes that is done to control workload across EAs. However, since the extent of the capping and redistribution of the sample described previously was moderate, the corresponding increase in the variation of the overall sampling rates was small, resulting in stratum-level UEW DEFFs that range from 1.00 to 1.06 (Table 2-5). The overall UEW DEFF is higher at 1.10 because it reflects total variation in weights within and across strata.

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

Stratum code	Stratum name	Number of PSUs	Number of sampled DUs/HHs	Minimum number of DUs/HHs per PSU	Maximum number of DUs/HHs selected per PSU	Unequal weighting DEFF
1	Hhohho Rural	39	1,365	17	53	1.00
2	Hhohho Urban	18	630	27	53	1.00
3	Lubombo Rural	35	1,224	21	53	1.00
4	Lubombo Urban	7	246	22	61	1.00
5	Manzini Rural	38	1,331	29	53	1.00
6	Manzini Urban	28	979	15	67	1.01
7	Shiselweni Rural	31	1,086	20	70	1.06
8	Shiselweni Urban	4	139	31	38	1.00
All	Eswatini	200	7,000	15	70	1.10[1]

DU = dwelling unit; HH= household

[1] Overall DEFF reflects total variation in weights within and across strata.

Table 2-6 summarizes the distribution of the sampled dwelling units/households by final dwelling unit/household response status. Of the 7,000 sampled dwelling units, 492 (7.0%) were determined during data collection to be vacant/unoccupied, 23 (0.3%) were those for which eligibility for the survey (i.e., occupancy status) could not be established, 1,072 (15.3%) were determined to be eligible for the study (i.e., contained eligible household members) but did not complete the household interview, and 5,413 (77.3%) completed the household interview. Excluding the ineligible cases, the overall unweighted household response rate was 83.2%.

Table 2-6 Distribution of dwelling unit/household sample by stratum and response status

Stratum code	Stratum name	Number of sampled DUs/HHs	Number of ineligible DUs[1]	Number of DUs/HHs with unknown eligibility[2]	Number of responding households[3]	Number of eligible non-responding households[4]	Unweighted response rate[5]
1	Hhohho Rural	1,365	111	1	1,043	210	0.832
2	Hhohho Urban	630	68	1	378	183	0.673
3	Lubombo Rural	1,224	80	3	968	173	0.846
4	Lubombo Urban	246	9	2	171	64	0.722
5	Manzini Rural	1,331	50	8	1,116	157	0.871
6	Manzini Urban	979	68	3	753	155	0.827
7	Shiselweni Rural	1,086	92	4	893	97	0.899
8	Shiselweni Urban	139	14	1	91	33	0.729
All	Eswatini	7,000	492	23	5,413	1,072	0.832

DU = dwelling unit; HH= household

[1] Vacant dwelling units, nonresidential units, and units located outside the sampled PSU, as determined during data collection.

[2] Sampled dwelling units/households for which existence of eligible households could not be ascertained.

[3] Households completing the household interview.

[4] Occupied households that did not complete the household interview.

[5] Computed as R/ [R + N + U*{(R + N)/(R + N + I)}], where R = number of households completing interview; N = number of eligible nonresponding households; I = number of ineligible dwelling units, and U = number of dwelling units with unknown eligibility.

2.5 Selection of Individuals

The selection of individuals for SHIMS3 2021 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. As noted below, only those individuals who were present (i.e., slept) in the household on the night prior to the time the household roster was compiled (i.e., the *de facto* population) were eligible for data collection and 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. Included on the roster were 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 not present during that time. The information recorded for each rostered individual included sex, age, relationship to head of household, residency status (i.e., whether a usual resident), and physical presence in household (i.e., slept in household the night prior to interview). Table 2-7 summarizes the number of households completing the roster and the corresponding number of rostered individuals by sampling stratum and resident status.

Table 2-7 Distribution of households completing rosters and corresponding numbers of rostered persons by resident status and stratum

Stratum code	Stratum name	Number of households completing interview	Rostered persons by resident status[1]				
			Usual resident/did not sleep here[2]	Usual resident/ slept here	Nonresident/ slept here	Nonresident/ did not sleep here[2]	Total rostered persons
1	Hhohho Rural	1,043	325	4,550	105	272	5,252
2	Hhohho Urban	378	26	1,143	34	27	1,230
3	Lubombo Rural	968	472	4,318	103	182	5,075
4	Lubombo Urban	171	8	422	18	9	457
5	Manzini Rural	1,116	598	4,839	231	446	6,114
6	Manzini Urban	753	184	1,765	58	60	2,067
7	Shiselweni Rural	893	323	4,024	127	206	4,680
8	Shiselweni Urban	91	18	224	11	4	257
All	Eswatini	5,413	1,954	21,285	687	1,206	25,132

[1] Counts include persons of all ages.

[2] Not eligible to be surveyed for SHIMS3 2021.

2.5.2 Selecting Individuals for Data Collection

All individuals listed in the household rosters who were 15 years of age and older and were present (slept in the household) on the night prior to the household interview were eligible for data collection. Excluded are usual residents and any rostered nonresidents who were not present in the household on the night prior to the interview. Table 2-8 summarizes the number of individuals eligible for data collection by stratum, age group, and resident status.

Table 2-8 Number of individuals eligible for data collection

Stratum code	Stratum name	Persons 15-49 years[1]			Persons 50 years or older[1]		
		Usual resident/ slept here	Nonresident/ slept here	Total sampled persons[2]	Usual resident/ slept here	Nonresident/ slept here	Total sampled persons[2]
1	Hhohho Rural	2,163	63	2,226	633	13	646
2	Hhohho Urban	685	23	708	125	2	127
3	Lubombo Rural	1,977	69	2,046	579	13	592
4	Lubombo Urban	287	16	303	36	0	36
5	Manzini Rural	2,314	146	2,460	743	18	761
6	Manzini Urban	1,140	40	1,180	138	4	142
7	Shiselweni Rural	1,927	81	2,008	594	17	611
8	Shiselweni Urban	124	10	134	34	0	34
All	Eswatini	10,617	448	11,065	2,882	67	2,949

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

[2] Eligible persons selected for data collection based on information reported in roster.

2.5.3 Distribution of Sampled Persons

Table 2-9 summarizes the number of individuals selected for data collection and the corresponding numbers completing the interview and blood test by age group and stratum. Note that the age classification in this table is based on rostered age. Interview respondents are those persons who met the criteria for completing the individual interview. Among the interview respondents, the blood test respondents are those persons who provided analyzable blood test results (i.e., had a final HIV status determination). The criteria used to define the interview and blood test respondents are given in Appendix B.

Table 2-9 Distribution of sampled persons by age group, response status, and stratum

Stratum code	Stratum name	Persons 15-49 years[1]			Persons 50 years or older[1]		
		Selected for data collection	Interview respondents[2]	Blood test respondent[3]	Selected for data collection	Interview respondents[2]	Blood test respondent[3]
1	Hhohho Rural	2,226	1,863	1,757	646	577	556
2	Hhohho Urban	708	604	525	127	111	96
3	Lubombo Rural	2,046	1,696	1,615	592	535	511
4	Lubombo Urban	303	246	226	36	32	30
5	Manzini Rural	2,460	2,106	1,971	761	697	662
6	Manzini Urban	1,180	1,054	932	142	130	118
7	Shiselweni Rural	2,008	1,695	1,564	611	553	512
8	Shiselweni Urban	134	115	102	34	29	22
All	Eswatini	11,065	9,379	8,692	2,949	2,664	2,507

[1] Age recorded in household roster. In a small number of instances, the actual confirmed age at interview may be different.

[2] Persons who completed all relevant modules of the individual interview (see Appendix B.2).

[3] Subset of interview respondents with confirmed results of blood tests (see Appendix B.3).

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 computing an appropriate sampling weight for 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, defined as 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 SHIMS3 2021, nonresponse can occur at different stages of data collection. For example: (1) after the selection of the sample of PSUs at the first stage; (2) after the enumeration of dwelling unit/household in each PSU at the second stage; (3) after household enumeration and selection of persons but before completion of the individual interview; and (4) 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 3.4.3.3 and 3.4.4.3 are used to calibrate the weighted sample counts to available population projections.

3.1 Overview of the Weighting Process

The overall weighting approach for SHIMS32021 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. All of the subsequent weighting steps described below are applied to the full sample, and to each of the jackknife replicates.

Calculation of PSU Weights: The weighting process begins with the calculation and checking of the sample PSU (EA) base weights as the reciprocals of the overall PSU probabilities of selection. The PSU base weights are adjusted to account for nonresponding eligible PSUs, if there are any. This adjustment is generally made within the sampling stratum in which the PSUs are located. The resulting weight is the final PSU weight.

Calculation of Household Weights: The next step is to calculate household weights. The household base weights are calculated as the nonresponse adjusted PSU 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 (see Table 2-6) and then the responding households have their weights adjusted to account for nonresponding eligible households. This adjustment is made within the PSU or groups of PSUs in which the households are located. The resulting weight is the final household weight.

Calculation of Person-Level Interview Weights: Once the household weights are determined, they become the individual base weights for individuals found from the household roster to be eligible for the survey. Similar to the household weights, the first phase of individual weight adjustment is for any individuals whose eligibility is unknown. Eligibility is unknown when age was not confirmed at the interview stage. These adjusted individual weights are then adjusted for nonresponse among the eligible individuals, with a final poststratification adjustment for the individual weights to compensate for undercoverage in the sampling process by adjusting the weighted frequencies to correspond to 2021 population projections.

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

Four 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 SHIMS3 2021 survey data used to construct the sampling weights are contained in the following data files.

- **sz_CFF_hh_int_STAT_20211220**: A household (HH) file that contains the household data collected in the HH questionnaire.
- **sz_CFF_roster_STAT_20211220**: A file that contains the roster of household members collected in the HH questionnaire with a record for each rostered person.
- **sz_CFF_ind_int_STAT_20211220**: 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.
- **SZ2Biomarker20211221**: 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.

Each of these data files except the Biomarker file contains records for all sampled or collected cases, irrespective of response and eligibility status. However, for weighting purposes, a subset of the roster file was created with only "roster eligible" cases: these are person-level records from a

responding household with a roster age of 15 or older and who were identified on the roster as having slept in the household the night before the interview. At the time of creating weight delivery files the "roster ineligible" cases were returned to the delivery files; however they have missing values for the weight variables.

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 steps include:

- Check identification variables, 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 HH selection process.)
- Check counts of sampled and responding HHs against what was expected, overall and by sampling stratum.
- Adjust for substitution of EAs, 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) sampled households, (b) sampled individuals, and (c) individuals selected for blood draws.

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 SHIMS3 2021: the jackknife replication and Taylor Series methods. The jackknife replication variance estimation method is a widely used method for producing variance estimates using data from a complex survey. This method can correctly account for the stratification, clustering, and sample weighting, including nonresponse and poststratification weighting adjustments, from SHIMS3 2021 complex sample design. Taylor 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 Series

method, the required variables are those that indicate the "variance stratum" and the "variance unit" to which each sampled respondent belongs.

3.3.1 Jackknife Replication

To permit the calculation of 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 2.3.2, the PSUs were selected systematically from a list of PSUs that had been ordered geographically within sampling stratum. To take account of the precision benefits of implicit stratification as fully as possible, the sampled PSUs within each sampling stratum were paired off in the systematic order in which they were selected, treating each pair as a variance-estimation stratum. When there was an odd number of sampled PSUs in a sampling stratum, one of the variance-estimation strata was defined to contain three sampled PSUs. To fully reflect the sample design, the formation of the variance-estimation strata was applied to all 200 of the sampled PSUs.

For SHIMS3 2021, 98 variance-estimation strata were created. 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. For a variance-estimation stratum consisting of a pair of PSUs, the weight of the retained PSU within the variance-estimation stratum k was doubled. For a variance-estimation stratum consisting of three PSUs, the weight of the two retained PSUs within the variance-estimation stratum were increased by 1.5 (see Section 3.4.1). The process was repeated for all k = 1, 2, 3, … 98 variance-estimation strata, resulting in a total of 98 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 code	Sampling stratum name	Sampled PSUs[1]	Variance strata consisting of pairs	Variance strata consisting of triplets	Number of jackknife replicates
1	Hhohho Rural	39	18	1	19
2	Hhohho Urban	18	9	0	9
3	Lubombo Rural	35	16	1	17
4	Lubombo Urban	7	2	1	3
5	Manzini Rural	38	19	0	19
6	Manzini Urban	28	14	0	14
7	Shiselweni Rural	31	14	1	15
8	Shiselweni Urban	4	2	0	2
All	Eswatini	200	94	4	98

[1] Includes nonresponding and ineligible PSUs if applicable.

3.3.2 Taylor Series

Even though jackknife replication is the recommended method for variance estimation, not all software packages have a replication option to produce variance estimates. Therefore, information for producing Taylor Series estimates of variance is included in the SHIMS3 2021 data files.

The full-sample weight (see Section 3.4) is used as the weight to compute Taylor Series variance estimates. The variable **VarStrat** indicates the variance-estimation stratum and the variable **VarUnit** indicates the PSU within the variance-estimation stratum. This pair of variables allows the analyst to produce variance estimates if their software does not easily accommodate replication methods but does have a Taylor Series capability.

3.4 Development of Weights

3.4.1 PSU Weights

The initial weighting step after the jackknife replicates were defined was to calculate PSU base weights for the full sample and the replicates.

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 sampling stratum *h*. Using the PSU weights defined above, the sampled PSUs weight up to the numbers shown in the fourth column of Table 3-2.

As described in Section 3.3.1, 98 jackknife replicates were formed from the 200 sampled PSUs. For variance estimation, replicate-specific PSU weights, $W_{(r)bi}^{(1)}$, *r* = 1, 2, ..., 98 were created to provide the basis for calculating the required replicate weights in subsequent stages of the weighting process. Let *b* denote one of the variance-estimation strata created for jackknife replication within sampling stratum *h* (Section 3.3.1) and let *i* denote the PSU within variance-estimation stratum *b*. For a given jackknife replicate, *r* = 1, 2, ..., 98, the corresponding replicate-specific PSU base was computed as

$W_{(r)bi}^{(1)}$ =	$a\,W_{hi}^{(1)}$	if *b* = *r* and PSU *i* in variance-estimation stratum *b* is included in replicate *r*
=	0	if *b* = *r* and PSU *i* in variance-estimation stratum *b* is not included in replicate *r*
=	$W_{hi}^{(1)}$	if *b* ≠ *r*

$$W_{(r)bi}^{(1)} = \begin{cases} aW_{hi}^{(1)} \\ 0 \\ W_{hi}^{(1)} \end{cases}$$

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

Since there are no nonresponding PSUs in SHIM3 2020, no nonresponse adjustment is needed. The variables in the file that carry the nonresponse adjusted PSU weights are set equal to the PSU base weights.

Table 3-2 Number of PSUs and corresponding weighted counts by sampling stratum

Sampling stratum Code	Sampling stratum Name	Number of sample EAs (PSUs)	Weighted number of EAs (PSUs) [1]	Weighted measure of size (MOS) [2]
1	Hhohho Rural	39	425.82	50,327
2	Hhohho Urban	18	198.54	23,800
3	Lubombo Rural	35	371.05	38,498
4	Lubombo Urban	7	85.18	11,360
5	Manzini Rural	38	404.31	44,674
6	Manzini Urban	28	273.06	52,461
7	Shiselweni Rural	31	427.22	38,213
8	Shiselweni Urban	4	30.57	5,523
All	Eswatini	200	2,216	264,856

[1] Weights are the PSU base weights, $W_{hi}^{(1)}$. The weighted count provides an estimate of the number of PSUs in the sampling frame but differs from the true total due to random sample variation.

[2] The measure of size used to select the sample of PSUs; the PSU Measure of Size (MOS) equals the number of households in the frame. Weights are the PSU base weights, $W_{hi}^{(1)}$.

3.4.2 Dwelling Unit/Household Weights

3.4.2.1 Dwelling Unit/Household Base Weights

The household weighting process starts by calculating the dwelling unit/household level base weights. These are the product of the PSU weight adjusted for nonresponse (described in Section 3.4.1 as equal to the PSU base weight for SHIMS3 2021) and the reciprocal of the within-PSU dwelling unit/household selection probability; i.e., the dwelling unit/household base weight for sampled dwelling unit/household *j* in PSU *i* in sampling stratum *h* was computed as:

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

where

$W_{hi}^{(1A)}$ = the nonresponse-adjusted weight for PSU *i* in PSU sampling stratum *h*

$P_{j|hi}^{DU}$ = the conditional probability of selecting dwelling unit/household *j* in PSU *i* in sampling stratum *h*.

The corresponding weights for jackknife replicate $r = 1, 2, \ldots, 98$ were computed as:

$$W_{(r)bij}^{(2)} = W_{(r)bi}^{(1A)} / P_{j|hi}^{DU},$$

where $W_{(r)bi}^{(1A)}$ is the PSU nonresponse-adjusted weight for PSU i in variance estimation stratum b 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. The specific rules used to classify dwelling units into the response status groups are given in Appendix B. In Table 3-4, we show the weighted counts of dwelling units/households by response status and sampling stratum using the dwelling unit/household base weights described above. The characteristics of the dwelling unit/household base weights 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, and the minimum and maximum values of the weights, both overall and by sampling stratum.

Table 3-3 Distribution of sampled dwelling units/households by response status

Response status group[1]	Description	Number of sampled dwelling units/households
1	Respondent (household with completed household interview)	5,413
2	Nonrespondent (household without a completed household interview)	1,072
3	Ineligible (dwelling units with no households)	492
4	Unknown eligibility (not known if dwelling unit contains household)	23
All	—	7,000

[1] See Appendix B for definitions.

Table 3-4 Weighted counts of dwelling unit/household base weights by response status and sampling stratum

Sampling stratum code	Sampling stratum name	Response status[1]				Total groups 1-4
		Group 1: responding household	Group 2: nonresponding household	Group 3: ineligible dwelling unit	Group 4: unknown eligibility	
1	Hhohho Rural	46,766	9,416	4,977	45	61,204
2	Hhohho Urban	16,789	8,128	3,020	44	27,981
3	Lubombo Rural	39,367	7,036	3,253	122	49,778
4	Lubombo Urban	13,355	4,999	703	156	19,213
5	Manzini Rural	47,287	6,652	2,119	339	56,397
6	Manzini Urban	63,955	13,375	5,868	259	83,456
7	Shiselweni Rural	41,188	4,512	4,777	174	50,650
8	Shiselweni Urban	4,178	1,515	643	46	6,381
All	Eswatini	272,884	55,632	25,359	1,185	355,061

[1] See Table 3.3. Counts given in table are weighted counts using the dwelling unit base weights, $W_{hij}^{(2)}$ described in Section 3.4.2.1.

3.4.2.2 Adjustment for Dwelling Unit/Household Nonresponse

The general approach for handling dwelling unit/household nonresponse was to increase the weights of responding households so that they represent the nonresponding dwelling units/households in the same PSU. Because such nonresponse could occur before establishing whether or not a sampled dwelling unit/household is eligible for the study (i.e., whether or not the associated dwelling unit/household contains persons eligible for SHIMS3 2021), the nonresponse adjustment was implemented in two phases. In the first phase of adjustment, the base weights were adjusted to compensate for sampled dwelling units/households 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 dwelling units/households among those dwelling units/households known to be eligible for the study.

To account for variation in response rates across different types of PSUs, the dwelling unit/household nonresponse adjustments were made within weighting cells defined by the individual PSUs or group of PSUs. The procedures used to compute the nonresponse-adjusted dwelling unit/household weights are described below.

Phase 1 Adjustment

To account for those dwelling units/households in which eligibility status is unknown, in the first phase of adjustment, the weights of the dwelling units/households where eligibility status is known (response status groups 1, 2, and 3) were inflated by the inverse of the (weighted) rate of known eligibility status in the PSU weighting cell after eliminating the dwelling units with eligibility status unknown (i.e., response-status group 4). As indicated above, the weighting cells for the dwelling unit/household nonresponse adjustments are either the individual PSUs or a group of PSUs. Let n_{hf}^{DU} denote the number of sampled dwelling units/households in PSU weighting cell *f* in sampling stratum *h*. Note that n_{hf}^{DU} is the sum of the sample sizes in each of the four response status groups defined in Table 3-3, i.e.,

$$n_{hf}^{DU} = n_{hf}^{(1)} + n_{hf}^{(2)} + n_{hf}^{(3)} + n_{hf}^{(4)}$$

where

$n_{hf}^{(1)}$ = the number of responding dwelling units/households (i.e., dwelling units/households with a completed dwelling unit/household interview) in PSU weighting cell *f* in sampling stratum *h*

$n_{hf}^{(2)}$ = the number of eligible nonresponding dwelling units/households (i.e., dwelling units/households without a completed dwelling unit/household interview) in PSU weighting cell *f* in sampling stratum *h*

$n_{hf}^{(3)}$ = the number of known ineligible dwelling units/households (i.e., dwelling units known to contain no households) in PSU weighting cell *f* in sampling stratum *h*

$n_{hf}^{(4)}$ = the number of sampled dwelling units for which it is not known whether a household is present in PSU weighting cell *f* in sampling stratum *h* .

The first-phase nonresponse adjustment factor for PSU weighting cell *f* in sampling stratum *h* was computed as the ratio:

$$A_{hf}^{(DU1)} = \sum_{j=1}^{n_{hf}^{DU}} W_{hij}^{(2)} \; / \; \sum_{j=1}^{n_{hf}^{(1)}+n_{hf}^{(2)}+n_{hf}^{(3)}} W_{hij}^{(2)}$$

where $W_{hij}^{(2)}$ is the base weight for dwelling unit/household *j* in PSU *i* in PSU weighting cell *f* in sampling stratum *h*, and where the sum in the numerator extends over the entire sample of dwelling units/households in PSU weighting cell *f* in sampling stratum *h*, while the sum in the denominator extends over the first three response status groups of dwelling units/households.

The corresponding replicate-specific first-phase dwelling units/households nonresponse adjustment factor for cell *f* were similarly computed for jackknife replicate *r* = 1, 2, ..., 98.

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 *i* in PSU weighting cell *f* in sampling stratum *h* was then computed as:

$$W_{hij}^{DU1} = A_{hf}^{(DU1)} \, W_{hij}^{(2)}$$

The corresponding replicate weights for replicate r = 1, 2, …, 98 were computed in similar fashion as:

$$W_{(r)bij}^{DU} = A_{(r)hi}^{(DU)} W_{(r)bij}^{(2)},$$

where

$$A_{(r)hf}^{(DU)} = \sum_{j=1}^{n_{(r)hf}^{DU}} W_{(r)bij}^{(2)} / \sum_{j=1}^{n_{(r)hf}^{(1)}+n_{(r)hf}^{(2)}+n_{(r)hf}^{(3)}} W_{(r)bij}^{(2)}$$

Note that for the dwelling units in response-status group 4 (dwelling units/households of unknown eligibility), $W_{hij}^{DU} = W_{(r)bij}^{DU} = 0$ for r = 1, 2, …, 98.

The effect of this adjustment is to distribute the total weight of the unknown-eligibility cases (i.e., the estimated 1,185 dwelling units/households 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}^{DU} as computed above are summarized in Table 3-5.

Table 3-5 Weighted counts of dwelling units/households adjusted for unknown eligibility

Sampling stratum code	Sampling stratum name	Response status				Total households: groups 1-2
		Group 1: responding household	Group 2: nonresponding household	Group 3: ineligible dwelling unit	Total status 1-3	
1	Hhohho Rural	46,794	9,426	4,985	61,204	56,220
2	Hhohho Urban	16,817	8,140	3,024	27,981	24,957
3	Lubombo Rural	39,466	7,052	3,260	49,778	46,518
4	Lubombo Urban	13,462	5,045	706	19,213	18,507
5	Manzini Rural	47,560	6,703	2,133	56,397	54,264
6	Manzini Urban	64,168	13,413	5,875	83,456	77,581
7	Shiselweni Rural	41,352	4,515	4,783	50,650	45,867
8	Shiselweni Urban	4,208	1,526	647	6,381	5,734
All	Eswatini	273,827	55,821	25,413	355,061	329,648

Note: Counts in table are weighted counts using first-phase adjusted household weights, W_{hij}^{DU1}.

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 *i* in PSU weighting cell *f* in sampling stratum *h* was computed as the ratio:

$$A_{hf}^{(HH)} = \sum_{j=1}^{n_{hf}^{(1)}+n_{hf}^{(2)}} W_{hij}^{DU} \; / \; \sum_{j=1}^{n_{hf}^{(1)}} W_{hij}^{DU}$$

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

The weighted household interview response rate for cell *f* is $R_{hf}^{(HH)} = 1/A_{hf}^{(HH)}$.

The corresponding replicate-specific interview nonresponse adjustment factor for cell *f* were similarly computed for jackknife replicate *r* = 1, 2, ..., 98.

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

$$W_{hij}^{(2A)} = A_{hf}^{(HH)} \, W_{hij}^{DU}.$$

The corresponding replicate weights for replicate *r* = 1, 2, …, 98 were computed in similar fashion as:

$$W_{(r)bij}^{(2A)} = A_{(r)hf}^{(HH)} \, W_{(r)bij}^{DU},$$

where

$$A_{(r)hf}^{(HH)} = \sum_{j=1}^{n_{(r)hf}^{(1)}+n_{(r)hf}^{(2)}} W_{(r)bij}^{DU} \; / \; \sum_{j=1}^{n_{(r)hf}^{(1)}} W_{(r)bij}^{DU}.$$

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

In this section, we detail the calculation of person-level sampling weights to be used to analyze the individual interview responses in the SHIMS3 2021 data files. First, we define the initial person-level (interview) base weights in Section 3.4.3.1. Next, to compensate for interview nonresponse, the person base weights are adjusted within cells defined by variables available for both the responding and nonresponding individuals. Like the dwelling unit/household nonresponse adjustments described previously, this person-level nonresponse adjustment was implemented in two phases.

3.4.3.1 Person Base Weights

All persons included on the rosters provided by responding households initially receive a person-level base weight equal to the final nonresponse-adjusted household weight, $W_{hij}^{(2A)}$. That is, the base weight for rostered person *k* in household *j* in PSU *i* in sampling stratum *h* was computed from the formula

$$W_{hijk}^{(base)} = W_{hij}^{(2A)} .$$

The corresponding replicate base weights, $W_{(r)bijk}^{(base)}$, for $r = 1, 2, \ldots, 98$ were computed in an analogous manner, with $W_{hij}^{(2A)}$ replaced by $W_{(r)bij}^{(2A)}$ in the above formula.

3.4.3.2 Adjustment of Person Weights for Interview Nonresponse

Since the final eligibility of a rostered person cannot be determined until after the actual age is confirmed during the interview, the person-level base weights were adjusted in two phases. Table 3-6 summarizes the distribution of the rostered persons by the five response-status groups specified for the first-phase adjustment. Response status groups 4 and 5 are the cases determined to be ineligible for the study because they were either under 15 years old, or because they were neither present in the household nor a usual resident of the household at the time the household roster was compiled. All of these cases are treated as "known ineligible" cases and are excluded from the first-phase adjustment. The cases in response-status group 3 are cases for which final eligibility for the study is not known because actual age was not obtained. The combined weight of these individuals

was distributed to the cases in response-status groups 1 and 2 within weighting classes defined by sex and age group as described below.

Table 3-6 Distribution of rostered persons by age group and de facto status and eligibility

First-phase eligibility group[1]	Resident status and age based on roster	Confirmed age based on interview	Eligibility Status	Number of rostered persons	Weighted number of rostered persons[2]
1	De facto person 15 years or older	15+	Eligible	14,010	807,225.07
2	De facto person 15 years or older	Under 15	Ineligible	0	0.00
3	De facto person 15 years or older	Unknown	Eligibility Unknown	4	234.36
4	Non de facto persons 15 years or older	NA	Ineligible	2,569	137,541.12
5	Persons under 15 years	NA	Ineligible	8,549	467,294.33
All	—	—		25,132[3]	1,412,295[3]

[1] See Appendix B for definitions of response status categories.

[2] Weighted by the person-level base weight, $W_{hijk}^{(base)}$.

[3] Of the 25,132 rostered persons, 1,206 were those that neither slept in the household nor were usual residents (see Table 2-7). On a weighted basis, these 1,206 persons account for 63,333 of the total weighted count of 1,412,295 rostered persons.

Phase 1 Adjustment

The procedure for computing the first phase adjustment was as follows. For each of the sex-age weighting classes specified for the adjustment, the first-phase interview nonresponse adjustment factor for cell c is, $A_c^{(1)}$, was computed as

$$A_c^{(1)} = \left(\sum_{i=1}^{n_c^{(1)}} W_{ck}^{(base)} + \sum_{i=1}^{n_c^{(2)}} W_{ck}^{(base)} + \sum_{i=1}^{n_c^{(3)}} W_{ck}^{(base)}\right) / \left(\sum_{k=1}^{n_c^{(1)}} W_{ck}^{(base)} + \sum_{i=1}^{n_c^{(2)}} W_{ck}^{(base)}\right)$$

where c denotes the first-phase adjustment cell, $W_{ck}^{(base)}$ is the base weight for person k in cell c, and $n_c^{(a)}$ = the number of cases in eligibility-status group a = 1, 2, 3 (Table 3-6) in weighting class c.

The corresponding replicate-specific first-phase interview nonresponse adjustment factors for cell c were similarly computed for jackknife replicate r = 1, 2, ..., 98.

The first-phase weighted interview response rate for cell c is $R_c^{(1)} = 1/A_c^{(1)}$ for the full sample, and $R_{(r)c}^{(1)} = 1/A_{(r)c}^{(1)}$ for jackknife replicate r = 1, 2, ..., 98.

The full-sample first-phase nonresponse-adjusted weight for person k in cell c was then computed as

$$W_{ck}^{(3)} = A_c^{(1)} W_{ck}^{(base)},$$

and the corresponding jackknife replicate weights for replicate $r = 1, 2, ..., 98$ were similarly computed as

$$W_{(r)ck}^{(3)} = A_{(r)c}^{(1)} W_{(r)ck}^{(base)}.$$

Phase 2 Adjustment

Table 3-7 summarizes the unweighted and weighted counts of eligible sample persons by sex and interview response status. The weights used to derive the weighted counts in this table are the first-phase person-level nonresponse-adjusted weights, $W_{ck}^{(3)}$. To compensate for interview nonresponse, the first-phase nonresponse-adjusted weights, $W_{ck}^{(3)}$, were further 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 sampling 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) was used for initial variable selection, and the Chi-square Automatic Interaction Detector (CHAID) was used to form the final weighting cells for nonresponse adjustment.

Table 3-7 Unweighted and weighted counts of eligible sample persons by sex and interview response status

Sex/Age group[1]	Interview response status[2]	Unweighted sample size	Weighted count[3]
Male 15 or older	Eligible respondent	5,057	292,049
	Eligible nonrespondent	1255	71,388
	All response statuses	**6312**	**363,437**
Female 15 or older	Eligible respondent	6986	403,112
	Eligible nonrespondent	712	40,911
	All response statuses	**7698**	**444,023**
Total 15 years or older	Eligible respondent	12043	695,161
	Eligible nonrespondent	1967	112,298
	All response statuses	**14,010**	**807,459**

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

[2] See Appendix B for definitions of the interview response status categories.

[3] Weighted by the first-phase adjusted person weight, $W_{ck}^{(3)}$.

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

There are 43 variables from the household questionnaire and EA sampling frame that could potentially be used for nonresponse adjustment. The LASSO regression was used to reduce the number of variables to a manageable subset that would subsequently be entered into the CHAID algorithm to define the final nonresponse adjustment weighting cells. 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 not significant have their regression coefficients set to zero (Hastie, Tibshirani, and Friedman, 2009).

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 base weight adjusted for unknown eligibility, $W_{ck}^{(3)}$. The final model was selected on the basis of cross validation with observations in the input data set partitioned into disjoint subsets, 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 so that if the program had to be re-run, the same results would be produced. Of the 43 variables used in the initial model, the LASSO identified 18 variables as significant predictors of response.

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 weighting cells based on information available for all sampled persons. The cells are formed in such a way that persons belonging to the same cell are expected to have similar propensities for responding to the study. 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 first-phase nonresponse-adjusted weight, $W_{ck}^{(3)}$. 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).

There are four different populations of inference at the individual level – males and females divided by age group (under and over 18 years of age). In the individual questionnaire, males and females in the specified age groups received different questions. To create the CHAID tree, gender (variable SEX) and an indicator of whether or not the individual was under 18 years of age (H_AGETEENYEARS) were forced into the model to make the initial splits. By forcing gender and age group as the first variables in the CHAID tree, the tree model essentially creates four distinct CHAID trees. After forcing these two variables into the model, the tree was then allowed to grow freely. The CHAID algorithm identified 13 variables to create the weighting classes for nonresponse adjustment. Table 3-8 lists the variables that were included in the final CHAID models. The final trees produced by the CHAID algorithm are documented in Appendix C.1. The corresponding nonresponse-adjustment classes used to adjust the person-level base weights are given in Appendix C.2.

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

Variable number	Variable name	Description
1	DEATHS	Has Any Usual Resident Of Your Household Died Since January 1, 2019?
2	ECONSUPCOVID	Economic Support - Covid19: Were Any Of These Economic Support Related To Covid-19?
3	H_AGETEENYEARS	Teen Indicator: 1 – 15-17 Years Old; 2 – Otherwise; Based On Ageyears (Roster)
4	H_AGEYEARS	Age (Categorical), Based On Roster Age. Matches Poststratification Cells
5	H_COOKINGFUEL	1. Electricity; 2. Gas/Kerosene/Coal; 6. Charcoal From Wood; 96. Other
6	H_HH_SIZE_C	1-9, Where 9 Includes All Hhs With 9 Or More People
7	H_RELATTOHH	1-Head, 2-Wife/Husband/Partner, 3-Son Or Daughter, 4-Son-In-Law/Daughter-In-Law, 5-Grandchild, 6-Parent, 7-Parent-In-Law, 8-Brother/Sister, 9-Co-Wife, 10-Other
8	H_WATERSOURCE	What Is The Main Source Of Drinking Water For Members Of Your Household?
9	LIVEHERE	Usually Live Here?
10	SEX	Male Or Female?
11	SICK_HOUSEHOLD	Any Member Of The Household Has Answered That They Are Sick On The Sick3mo Question
12	STRATA	Numeric Code For Ea Sampling Stratum

13	URBAN_RURAL	1=urban, 2=rural

Calculation of Phase 2 Nonresponse-Adjusted Person Weights

The general approach for computing the second-phase nonresponse-adjusted person-level interview weights was as follows. Within each of the final adjustment cells specified in Appendix C.2, the interview nonresponse adjustment factor for cell m $A_m^{(int)}$, was computed as

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

where m denotes the adjustment cell, $W_{mk}^{(3)}$ is the first-phase nonresponse-adjusted 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 interview nonresponse adjustment factor for cell m were similarly computed for jackknife replicate r = 1, 2, ..., 98 as

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

The weighted interview response rate for cell m is $R_m^{(int)} = 1/A_m^{(int)}$ for the full sample, and $R_{(r)m}^{(int)} = 1/A_{(r)m}^{(int)}$ for jackknife replicate r = 1, 2, ..., 98.

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, ..., 98 were similarly computed as

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

A summary of selected features of the nonresponse adjustment process is given in Table 3-9.

Table 3-9 Summary of the interview nonresponse adjustment process

Characteristic	Total sample
Number of variables in initial model	43
Number of variables selected by LASSO	18
Number of variables selected by CHAID	13
Number of final nonresponse-adjustment cells	32
Number of interview respondents	12,043
Minimum adjustment factor	1.00
Maximum adjustment	1.86
Weighted count of respondents before adjustment[1]	695,161
Weighted count of respondents after adjustment[2]	807,459

[1] Weight is the first-phase nonresponse-adjusted person weight, $W_{mk}^{(3)}$.

[2] Weight is the second-phase nonresponse-adjusted person weight, $W_{mk}^{(int)}$.

3.4.3.3 Poststratification Adjustment

The final step in computing the individual interview weights was to adjust the nonresponse-adjusted interview weights 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. For example, undercoverage can occur:

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

To compensate for the types of coverage problems indicated above, the nonresponse-adjusted person weights were ratio-adjusted so that the resulting weighted sample counts match the population control totals indicated in Table 3-10. The population control totals given in this table are projected 2021 national population projections by gender and five-year age groups provided by the CSO. The poststratified interview weights were computed as follows.

Let N_{ga}^{2021} denote the 2021 Eswatini 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:

$$T_{ga}^{2021} = N_{ga}^{2021} / \sum_{k=1}^{n_{ga}^{resp}} W_{gak}^{(int)},$$

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:

$$T_{(r)ga}^{2021} = N_{ga}^{2021} / \sum_{k=1}^{n_{(r)ga}^{resp}} W_{(r)gak}^{(int)}$$

for the *r* = 1, 2, …, 98 jackknife replicates.

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

$$W_{gak}^{(ps-int)} = T_{ga}^{2021} W_{gak}^{(int)},$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-int)} = T_{(r)ga}^{2021} W_{(r)gak}^{(int)}$$

for *r* = 1, 2, …,98.

Table 3-10 provides the population control totals, weighted counts of the respondents before poststratification, and the ratio of the control totals to the nonresponse adjusted weights (poststratification adjustment factor) by age and gender.

.

Table 3-10 2021 Eswatini population projections and weighted counts before poststratification

Age group	Male			Female			Total		
	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]
15-19	64,516	65,508	0.985	62,417	64,351	0.970	126,933	129,858	0.977
20-24	56,144	56,077	1.001	55,559	61,893	0.898	111,703	117,970	0.947
25-29	46,087	48,359	0.953	48,945	59,693	0.820	95,032	108,052	0.880
30-34	39,666	41,525	0.955	43,232	53,517	0.808	82,898	95,042	0.872
35-39	35,682	38,315	0.931	37,671	47,547	0.792	73,353	85,863	0.854
40-44	29,497	29,360	1.005	31,231	32,681	0.956	60,728	62,040	0.979
45-49	21,155	22,200	0.953	23,367	26,322	0.888	44,522	48,522	0.918
50-54	16,522	15,206	1.087	19,974	21,612	0.924	36,496	36,818	0.991
55-59	12,777	12,177	1.049	17,746	22,299	0.796	30,523	34,477	0.885
60-64	10,356	14,339	0.722	14,575	16,539	0.881	24,931	30,877	0.807
65+	22,723	20,389	1.114	35,730	37,551	0.951	58,453	57,940	1.009
Total 15+	355,125	363,453	0.977	390,447	444,006	0.879	745,572	807,459	0.923

[1] Source: Central Statistics Office (CSO).

[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 provided a useable blood sample. Thus, a separate set of weights is required for analysis of the blood test results. Similar to the construction of the interview weights described previously, development of the final blood test weights involves adjustments for nonresponse and poststratification to 2021 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 $W_{(r)bijk}^{(int)}$ (for replicate $r = 1, 2, \ldots, 98$), respectively, where k denotes the interview respondent, h denotes the sampling stratum (b denotes the variance stratum), i denotes the PSU, and j denotes the household. These weights have been adjusted for interview nonresponse, and thus act as the "base" weights for developing nonresponse adjustments for the blood test weights. Table 3-11 summarizes the counts of individuals by sex, age group and blood test response status, and the corresponding weighted counts using the nonresponse person-level interview weights, $W_{hijk}^{(int)}$.

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

Age group[1]	Sex	Blood test response status[2]	Unweighted sample size	Weighted count[3]
15 to 49 years	Male	Eligible respondent	3,771	273,643
		Eligible nonrespondent	341	27,700
	Female	Eligible respondent	4,915	320,465
		Eligible nonrespondent	346	25,539
50 years or older	Male	Eligible respondent	883	57,978
		Eligible nonrespondent	63	4,133
	Female	Eligible respondent	1,630	91,722
		Eligible nonrespondent	94	6,280
15 years or older	Male	Eligible respondent	4,654	331,621
		Eligible nonrespondent	404	31,832
	Female	Eligible respondent	6,545	412,187
		Eligible nonrespondent	440	31,819

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

[2] Status among the interview respondents. See Appendix B for definitions of the response status groups.

[3] Weighted count of interview respondents using final nonresponse-adjusted person-level interview weight, $W_{hijk}^{(int)}$.

3.4.4.2 Nonresponse Adjustment of Blood Test Weights

To compensate for blood test nonresponse, the nonresponse-adjusted person-level interview weights were further adjusted within cells defined by variables available for both the responding and nonresponding individuals (i.e., individuals completing the interview who may or may not have a final HIV status determination). These variables included data from the household roster and other information collected in the household questionnaire, selected PSU characteristics such as sampling stratum and urban/rural status, and the individual interview. The age and sex variables used to make the nonresponse adjustments are those reported in the interview.

For males, 77 potential predictor variables were available for initial selection. For females, 77 potential predictor variables were available for initial selection. The LASSO procedure was used to identify a reduced set of predictor variables to be used in the CHAID algorithm. From these initial sets of variables, the LASSO regression identified 30 significant variables for males and 46 significant variables for females. The selected variables were then input into the CHAID program to create the final weighting cells for nonresponse adjustment.

The CHAID algorithm identified 17 variables for males and 15 variables for females that were then used to create weighting classes for nonresponse adjustment. Table 3-12 lists the variables that were included in the final CHAID models. The final trees produced by the CHAID algorithm are documented in Appendix C.1. The corresponding nonresponse-adjustment classes used to adjust the person-level base weights are given in Appendix C.2.

Table 3-12 Variables selected by CHAID to produce classes for blood test nonresponse adjustment

Sex	Variable number	Variable name	Description
Male	1	ADDISHIV	Prevention Intervention: Have You Ever Discussed HIV With Your Parents Or Guardian?
	2	AT_BESTAGE_C	CATEGORICAL AGE BASED ON INTERVIEW AGE (CONFAGEY)
	3	AT_FIRSTSXAGE	AGE OF FIRST SEXUAL ACTIVITY - LIMITED TO MAXIMUM OF 21
	4	AT_LIFETIMESEX	IN TOTAL, WITH HOW MANY DIFFERENT PEOPLE HAVE YOU HAD SEX IN YOUR LIFETIME? - LIMITED TO MAXIMUM OF 10
	5	AT_PART12MONUM	How many different people have you had sex with in the last 12 months? - LIMITED TO MAXIMUM OF 8
	6	AVOIDPREG	Reproduction: Are You Or Your Partner Currently Doing Something Or Using Any Method To Delay Or Avoid Getting Pregnant?
	7	CONDOMGET	Prevention Intervention: If You Wanted A Condom, Would It Be Easy For You To Get One?
	8	CURMAR	What Is Your Marital Status Now: Are You Married, Living Together With Someone As If Married, Widowed, Divorced, Or Separated/Single?
	9	DEPRESSED	TB And Other Health Issues: Over The Past Two Weeks, How Often Have You Felt Down, Depressed Or Hopeless?
	10	ECONSUPCOVID	HH Economic Support - Covid19: Were Any Of These Economic Support Related To Covid-19?
	11	HFLAST12MO	HIV Testing: Have You Seen A Health Care Provider In A Health Facility In The Last 12 Months?
	12	MATEXWALLS	HH Characteristics: Main Material Of Exterior Walls
	13	PARTLASTCNDM1	Sexual Activity: The Last Time You Had Sex With Partner, Was A Condom Used?
	14	PARTLASTETOH1	Sexual Activity: The Last Time You Had Sex With Partner, Did Either Of You Drink Alcohol Beforehand?
	15	SCHCOM	Background: What Is The Highest Level Of School You Have Completed: Primary, Secondary, Or Higher?
	16	URBAN_RURAL	1=Urban, 2=Rural
	17	WORKIND	Background: What Is Your Occupation? That Is, What Kind Of Work Do You Mainly Do?
Female	1	AT_BESTAGE_C	Categorical age based on interview age (CONFAGEY)
	2	AT_LIFETIMESEX	In total, with how many different people have you had sex in your lifetime? - limited to maximum of 10
	3	AT_LIVEB	How many times have you had a pregnancy that resulted in a live birth? - limited to maximum of 14
	4	AT_PART12MONUM	How many different people have you had sex with in the last 12 months? - limited to maximum of 8
	5	CURMAR	What Is Your Marital Status Now: Are You Married, Living Together With Someone As If Married, Widowed, Divorced, Or Separated/Single?
	6	ECONSUPCOVID	HH Economic Support - Covid19: Were Any Of These Economic Support Related To Covid-19?

Table 3-12 Variables selected by CHAID to produce classes for blood test nonresponse adjustment (continued)

Sex	Variable number	Variable name	Description
Female	7	HIVSELFTST	HIV Testing: Have You Ever Tested Yourself For HIV Using A Self-Test Kit?
	8	MATEXWALLS	HH Characteristics: Main Material Of Exterior Walls
	9	MATFLO	HH Characteristics: Main Material Of Floor
	10	PARTLASTCNDM1	Sexual Activity: The Last Time You Had Sex With Partner, Was A Condom Used?
	11	ROOMSLEEP	HH Characteristics: How Many Rooms Are Used For Sleeping?
	12	SCHCOM	Background: What Is The Highest Level Of School You Have Completed: Primary, Secondary, Or Higher?
	13	STRATA	Numeric code for EA sampling stratum
	14	URBAN_RURAL	1=Urban, 2=Rural
	15	WORKIND	Background: What Is Your Occupation? That Is, What Kind Of Work Do You Mainly Do?

Calculation of Nonresponse-Adjusted Blood Test Weights

The general approach for computing the nonresponse-adjusted blood test weights was as follows. Within each of the final adjustment cells specified in Appendix C.2 for blood-test nonresponse adjustment factor for cell *m*, $A_m^{(BT)}$, was computed as

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

where *m* denotes the adjustment cell, $W_{mk}^{(int)}$ is the final nonresponse-adjusted person-level 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 nonresponse adjustment factor for cell *m* were similarly computed for jackknife replicate *r* = 1, 2, ..., 98.

The weighted blood test response rate for cell *m* is $R_m^{(BT)} = 1/A_m^{(BT)}$ for the full sample, and $R_{(r)m}^{(BT)} = 1/A_{(r)m}^{(BT)}$ for jackknife replicate *r* = 1, 2, ..., 98.

The full-sample nonresponse-adjusted blood test weight for 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 $r = 1, 2, ..., 98$ were similarly computed as

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

A summary of selected features of the blood-test nonresponse adjustment process is given in Table 3-13.

Table 3-13 Summary of the blood test nonresponse adjustment process

Characteristic	Male	Female
Number of variables in initial model	77	77
Number of variables selected by LASSO	30	46
Number of variables selected by CHAID	17	15
Number of final nonresponse-adjustment cells	37	35
Number of interview respondents	4,654	6,545
Minimum adjustment factor	1.00	1.00
Maximum adjustment	1.27	1.54
Weighted count of respondents before adjustment[1]	331,621	412,187
Weighted count of respondents after adjustment[2]	363,453	444,006

[1] Weight is nonresponse-adjusted person-level interview weight, $W_{mk}^{(int)}$.

[2] Weight is nonresponse-adjusted blood test weight, $W_{mk}^{(BT)}$.

3.4.4.3 Poststratification Adjustment

Like the nonresponse-adjusted interview weights described previously, the nonresponse-adjusted blood test weights were poststratified to projected 2021 Eswatini population counts within classes defined by gender and five-year age group.

Let N_{ga}^{2021} denote the 2021 Eswatini population control total for gender *g* and (five-year) age group *a* as given in Table 3-14. The poststratification ratio adjustment factor used to adjust the blood test weights for gender *g* and age group *a* was computed as:

$$T_{ga}^{2021} = N_{ga}^{2021} / \sum_{k=1}^{n_{ga}^{BT}} W_{gak}^{(BT)},$$

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:

$$T_{(r)ga}^{2021} = N_{ga}^{2021} / \sum_{k=1}^{n_{(r)ga}^{BT}} W_{(r)gak}^{(BT)}$$

for the *r* = 1, 2, …, 98 jackknife replicates.

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

$$W_{gak}^{(ps-BT)} = T_{ga}^{2021} W_{gak}^{(BT)},$$

and the corresponding poststratified replicate weights were computed as:

$$W_{(r)gak}^{(ps-BT)} = T_{(r)ga}^{2021} W_{(r)gak}^{(BT)}$$

for *r* = 1, 2, …, 98.

Weighted counts of the blood test respondents before and after poststratification (namely, the population control totals) are summarized in Table 3-14.

Table 3-14 2021 Eswatini population projections and weighted counts of blood test respondents before and after poststratification

Age group	Male			Female			Total		
	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]	Population control total[1]	Weighted count before post-stratification[2]	Poststrat-ification ratio[3]
15-19	64,516	65,775	0.981	62,417	65,191	0.957	126,933	130,966	0.969
20-24	56,144	55,776	1.007	55,559	62,274	0.892	111,703	118,050	0.946
25-29	46,087	48,881	0.943	48,945	59,564	0.822	95,032	108,445	0.876
30-34	39,666	40,490	0.980	43,232	53,270	0.812	82,898	93,760	0.884
35-39	35,682	38,175	0.935	37,671	47,447	0.794	73,353	85,622	0.857
40-44	29,497	29,868	0.988	31,231	33,053	0.945	60,728	62,921	0.965
45-49	21,155	21,422	0.988	23,367	26,149	0.894	44,522	47,570	0.936
50-54	16,522	15,113	1.093	19,974	21,219	0.941	36,496	36,331	1.005
55-59	12,777	12,559	1.017	17,746	21,820	0.813	30,523	34,379	0.888
60-64	10,356	14,739	0.703	14,575	16,409	0.888	24,931	31,148	0.800
65+	22,723	20,657	1.100	35,730	37,611	0.950	58,453	58,267	1.003
Total 15+	355,125	363,453	0.977	390,447	444,006	0.879	745,572	807,459	0.923

[1] Source: Central Statistics Office (CSO).

[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, $W_{gak}^{(BT)}$.

References

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics.

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

2017 – 2038 Population Projections: Based on the 2017 Eswatini Population and Housing Census, National Statistical Office (CSO), August 2020

Appendix A

Definition of Eligibility for Dwelling Unit/Household Sampling

Appendix A - 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 39,631 dwelling unit/household records in the listing file, 13 were classified as ineligible for sampling based on the structure type, vacancy status, and residential status. Thus, a total of 39,618 records in the Master Listing File were eligible for household sampling.

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

Structure type (STOBS_D)	Vac/Constr. Status (STVAC_D)	Resid. Status (RESYN_D)	ELIG_FLAG	Total in master file	Eligible
Cases with no GPS information			N	0	0
1 = Single House / compound of houses	1 = Not Vacant and not under Construction	1 = Yes	Y	21,862	21,862
1 = Single House / compound of houses	1 = Not Vacant and not under Construction	2 = No	Y	112	112
1 = Single House / compound of houses	2 = Vacant	1 = Yes	Y	58	58
1 = Single House / compound of houses	2 = Vacant	2 = No	Y	326	326
1 = Single House / compound of houses	3 = Under Construction	1 = Yes	Y	161	161
1 = Single House / compound of houses	3 = Under Construction	2 = No	Y	288	288
2 = Flat/Block/Apartment building	1 = Not Vacant and not under Construction		Y	1	1
2 = Flat/Block/Apartment building	1 = Not Vacant and not under Construction	1 = Yes	Y	15,845	15,845
2 = Flat/Block/Apartment building	1 = Not Vacant and not under Construction	2 = No	Y	36	36
2 = Flat/Block/Apartment building	1 = Vacant	1 = Yes	Y	32	32
2 = Flat/Block/Apartment building	2 = Vacant	2 = No	Y	430	430
2 = Flat/Block/Apartment building	3 = Under Construction		Y	28	28
2 = Flat/Block/Apartment building	3 = Under Construction	1 = Yes	Y	27	27
2 = Flat/Block/Apartment building	3 = Under Construction	2 = No	Y	157	157
3 = Church/Mosque/Temple	1 = Not Vacant and not under Construction	1 = Yes	Y	27	27
3 = Church/Mosque/Temple	1 = Not Vacant and not under Construction	2 = No	N	5	-
3 = Church/Mosque/Temple	2 = Vacant	1 = Yes	Y	-	-
3 = Church/Mosque/Temple	2 = Vacant	2 = No	N	2	-
3 = Church/Mosque/Temple	3 = Under Construction	1 = Yes	Y	2	2
3 = Church/Mosque/Temple	3 = Under Construction	2 = No	N	1	-
4 = Shop/office/bus. cntr/comm. bldg.	1 = Not Vacant and not under Construction	1 = Yes	Y	45	45
4 = Shop/office/bus. cntr/comm. bldg.	1 = Not Vacant and not under Construction	2 = No	N	2	-
4 = Shop/office/bus. cntr/comm. bldg.	2 = Vacant	1 = Yes	Y	-	-
4 = Shop/office/bus. cntr/comm. bldg.	2 = Vacant	2 = No	N	1	-
4 = Shop/office/bus. cntr/comm. bldg.	3 = Under Construction	1 = Yes	Y	-	-
4 = Shop/office/bus. cntr/comm. bldg.	3 = Under Construction	2 = No	N	-	-

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

Structure type (STOBS_D)	Vac/Constr. Status (STVAC_D)	Resid. Status (RESYN_D)	ELIG_FLAG	Total in master file	Eligible
5 = School/University	1 = Not Vacant and not under Construction	1 = Yes	Y	157	157
5 = School/University	1 = Not Vacant and not under Construction	2 = No	Y	-	-
5 = School/University	2 = Vacant	1 = Yes	Y	-	-
5 = School/University	2 = Vacant	2 = No	Y	2	2
5 = School/University	3 = Under Construction	1 = Yes	Y	-	-
5 = School/University	3 = Under Construction	2 = No	N	-	-
6 = Clinic/hospital/Doctors office	1 = Not Vacant and not under Construction	1 = Yes	Y	17	17
6 = Clinic/hospital/Doctors office	1 = Not Vacant and not under Construction	2 = No	N	-	-
6 = Clinic/hospital/Doctors office	2 = Vacant	1 = Yes	Y	-	-
6 = Clinic/hospital/Doctors office	2 = Vacant	2 = No	N	-	-
6 = Clinic/hospital/Doctors office	3 = Under Construction	1 = Yes	Y	-	-
6 = Clinic/hospital/Doctors office	3 = Under Construction	2 = No	N	-	-
7 = Community Center/CBO	1 = Not Vacant and not under Construction	1 = Yes	Y	4	4
7 = Community Center/CBO	1 = Not Vacant and not under Construction	2 = No	N	1	-
7 = Community Center/CBO	2 = Vacant	1 = Yes	Y	-	-
7 = Community Center/CBO	2 = Vacant	2 = No	N	-	-
7 = Community Center/CBO	3 = Under Construction	1 = Yes	Y	-	-
7 = Community Center/CBO	3 = Under Construction	2 = No	N	-	-
96 = Other	1 = Not Vacant and not under Construction	1 = Yes	Y	1	1
96 = Other	1 = Not Vacant and not under Construction	2 = No	N	-	-
96 = Other	2 = Vacant	1 = Yes	Y	-	-
96 = Other	2 = Vacant	2 = No	N	1	-
96 = Other	3 = Under Construction	1 = Yes	Y	-	-
96 = Other	3 = Under Construction	2 = No	N	-	-
				39,631	39,618

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

Structure type (STOBS_D)
1 - Single House/compound of houses
2 - Flat/Block/Apartment building
3 - Church/Mosque/Temple
4 - Shop/office/business cntr/commercial bldg.
5 - School/University
6 - Clinic/hospital/Doctors office
7 - Community Center/CBO
96 – 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

Appendix B - 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 Survey Status for Household: HH_STATUS

B.1.1 Summary

HH_STATUS is defined for all sampled dwelling units. First, the variable UPCODE_RESLTNDT is derived using RESULTNDTOTHR. Next, the questionnaire completion variable and the upcoded RESULTNDT are used to calculate UPCODE_STAT_HH. Lastly, HH_STATUS is set equal to UPCODE_STAT_HH when the Data Lock files are delivered.

HH_STATUS	Description
1	Responding household (completed household interview)
2	Nonresponding in-scope household
3	Household not in scope for the survey
4	Household whose survey eligibility could not be determined

B.1.2 SAS code defining HH_STATUS

HH_STATUS = UPCODE_STAT_HH;

Definition for responding household:

UPCODE_STAT_HH = 1 if:

- RESULTNDT is NULL and (STARTINT = 1 AND HHELIG = 1 AND HHCONSTAT = 1 AND HHQDTHSINS is NOT NULL AND ROSTER_START and HHRINS3 is NOT NULL AND HHQINSHH is NOT NULL AND HHQASSIGN_INST is NOT NULL) OR

- RESULTNDT is NULL and (STARTINT = 4 and HHRINS3 is NOT NULL and HHQINSHH is NOT NULL AND HHQASSIGN_INST is NOT NULL)

Definitions for household without completed questionnaire:

The table below shows the values for RESULTNDT on the data file:

CANNOT COLLECT CSPRO CODE (RESULTNDT)	Map to UPCODE_STAT_HH
1 = HH NOT AVAILABLE AT ALL VISIT ATTEMPTS	2 = NONRESPONDING HH
2 = REFUSED	2 = NONRESPONDING HH
3= DWELLING VACANT OR ADDRESS NOT A DWELLING	3 = INELIGIBLE HH
4= DWELLING DESTROYED	3 = INELIGIBLE HH
5= DWELLING NOT FOUND	4 = UNKNOWN STATUS HH
6= HOUSEHOLD ABSENT FOR EXTENDED PERIOD OF TIME	3 = INELIGIBLE HH
96 = OTHER	Will be upcoded to UPCODE_RSLTNDT

ELSE assign UPCODE_STAT_HH to 2, 3 or 4 using rules shown below.

UPCODE_STAT_HH = 2 if

- RESULTNDT OR UPCODE_RESLTNDT = 1 or 2 or 7 or 8 or 9
- If RESULTNDT=NULL, then
  - If HHELIG = 2 OR
  - (HHCONSTAT = 2 or 3) or
  - HHELIG = 1 AND HHCONSTAT=NULL OR
  - STARTINT = 4 and HHQASIGN_INST is NULL or
  - (HHQINSHH is NOT NULL) and HHQASSIGN_INST is NULL

UPCODE_STAT_HH = 3 if

- RESULTNDT OR UPCODE_RESLTNDT = 3 or 4 or 6 or 91

UPCODE_STAT_HH = 4 if

- RESULTNDT or UPCODE_RESLTNDT = 5, 89, 90, 92, 93, 94, 95, 97 or 99 or
- STARTDWELL = 10 and the record does not meet the criteria for 1, 2, or 3

UPCODE_RESLTNDT categories for RESULTNDT = '96' cases, based on text in RESULTNDOTH

Text meaning	UPCODE_RESLTNDT	UPCODE_STAT_HH
HOUSEHOLD NOT AVAILABLE AT ALL VISIT ATTEMPTS	1	2
REFUSED	2	2
DWELLING VACANT OR ADDRESS NOT A DWELLING	3	3
DWELLING DESTROYED	4	3
DWELLING NOT FOUND	5	4
HOUSEHOLD ABSENT FOR EXTENDED PERIOD OF TIME	6	3
Bereavement related	7	2
No capable Head of Household available to do survey	8	2
Out of Scope	91	3
COVID Delay – Unknown Eligibility	94	4
Cannot Trace	95	4
Recorded in another HH or tablet (discrepant record)	99	4

Table of examples for RESULTNDOTH upcoding

RESULTNDOTH	UPCODE_ RESLTNDT	UPCODE_ STAT_HH
Not available at three occasions	1	2
HOUSEHOLD HEAD TOO BUSY TO ACCOMODATE SURVEY		
HOUSEHOLD HEAD NOT AVAILABLE FOR AN EXTENDED PERIOD OF TIME		
HOUSEHOLD HEAD IS AWAY IN SOUTH AFRICA AND WIFE IS NOT ABLE TO MAKE DECISIONS OR GIVE PERMISSION		
HHH IS AN ARTISAN MINOR HE COMES BACK AROUND 10 PM AND GOES VERY EARLY IN THE MORNING AROUND 4 AM		
KEPT GIVING APPOINTMENTS BUT WAS NOWHERE TO BE FOUND ON LAST DAY		
PARTICIPANT 'S WORK SHIFTS COULD NOT ACCOMMODATE SURVEY ACTIVITIES TO BE CONDUCTED.		
Refusing Behavior	2	2
COULD NOT ACCOMODATE SURVEY DUE TO RELIGIOUS AFFILIATION.THEY ARE FROM THE JOHANNE MARANGE CHURCH		
DATA CANNOT BE COLLECTED DUE TO STRONG RELIGOUS BELIEF		
HEAD OF HOUSE STATED THAT IF THERE ARE NO MONETARY BENEFITS HIS HOUSEHOLD SHOULD NOT BE INCLUDED		
PARTICIPANT REFUSED TO PARTICIPATE IN THE SURVEY AND THE REASON BEING DOMESTIC ISSUES.		
THE FAMILY WAS RECENTLY ATTACHED AND ROBBED BY ARMED ROBBERS AT GUN POINT. WRONG TIMING		
HH HEAD LISTED AGREED HOWEVER THE SON IS NOT ALLOWING THE PROCEDURES TO BE DONE		
Vacant or not a dwelling	3	3
STRUCTURE UNDER CONSTRUCTION STILL AT FOUNDATION LEVEL		
NO ONE SLEEPS AT THE HOUSE		
HOUSEHOLD HEAD DECEASED. DWELLING VACANT		
VACANT		
DWELLING IS A BOTTLESTORE		
Household absent for extended period of time	6	3
MEMBERS OF THE HOUSEHOLD HAVE TRAVELLED FOR A LONG PERIOD OF TIME		
THE INDIVIDUAL STAYS ALONE AND HE HAS TRAVELLED TO ARGENTINA AND THERE IS NOONE STAYING AT THE HOUSE		
Death/Funeral	7	2
SHE LOST HER BOYFRIEND WHO WAS BURIED LAST SUNDAY. HE DIED OF LIVER PROBLEMS IN SOUTH AFRICA		
FUNERAL AT THE HOUSEHOLD		
GRIEVING.SHE RECENTLY LOST A SON AND MOURNERS ARE STILL GATHERED		
NOT IN AN EMOTIONAL STATE TO PARTICIPATE, HH MISSING, DEATH OF A GRANDCHILD AND BIRTH OF CHILD		
CLOSE RELATIVE (DAUGHTER-IN-LAW) TO THE DECEASED BURIAL SCHEDULED		

Table of examples for RESULTNDOTH upcoding (continued)

RESULTNDOTH	UPCODE_ RESLTNDT	UPCODE_ STAT_HH
Participant/Household Head unable to do survey (incapacitated, language barrier, under age)	8	2
HOUSEHOLD HEAD INCAPACITATED MENTALLY CHALLENGED		
THE PARTICIPANT IS INCAPACITATED -DEAF		
SINGLE HOUSEHOLD MEMBER WHO IS TOO OLD AND INCAPACITATED		
HH IS 14 YEARS OLD SO PARTICIPANT IS INELIGIBLE		
HOUSEHOLD HEAD UNABLE TO SPEAK ANY OF THE SURVEY LANGUAGES		
THE HOUSEHOLD HEAD PASSED ON IN BULAWAYO ON THE 3RD DAY VISIT. NO ONE TO CONSENT FOR THE HOUSEHOLD		
HOUSEHOLD HEAD INVOLVED IN A CAR ACCIDENT THEREFORE CANNOT ACCOMODATE AN INTERVIEW		
MEMBERS OF THE HOUSEHOLD HAVE TRAVELLED FOR A LONG PERIOD OF TIME		
THE INDIVIDUAL STAYS ALONE AND HE HAS TRAVELLED TO ARGENTINA AND THERE IS NOONE STAYING AT THE HOUSE		
Out of Scope	91	3
COVID Delay – Unknown Eligibility	94	4
Cannot Trace	95	4
Recorded in another HH or tablet (discrepant record)	99	4

B.2 INDIV_STATUS

B.2.1 Summary

INDIV_STATUS is defined for all final roster records. This variable is derived when the Data Lock files are delivered.

INDIV_STATUS	Description
1	Respondent
2	Eligible nonrespondent
3	Roster eligible but confirmed age <15
4	Roster eligible but no confirmed age
5	Roster ineligible (roster age < 15 or SLEEPHERE=2, except cases in status 9)
6	Rostered case from household with no questionnaire data
9	DeJure ineligible (SLEEPHERE = 2, LIVEHERE = 1 and roster age >=15)

B.2.2 SAS Code for INDIV_STATUS

First create a variable to designate whether the case is survey eligible based on the roster:

```
label roster_elig = "Flag for roster eligible";

if hh_status ^= 1 then roster_elig = 2;
```

```
else
  if sleephere = 1 and
    ageyears => 15 then roster_elig = 1;
  else
    roster_elig = 0;
```

Next, combine Roster_Elig with endmsg1 and Confagey to create INDIV_STATUS (endmsg1 = 'A' indicates a completed Individual questionnaire)

```
label INDIV_STATUS = "Individual Response Status";

if roster_elig = 2 then indiv_status = 6;
else
  if roster_elig = 0 then do;

    if sleephere = 2 and
      livehere  = 1 and
      ageyears >= 15 then indiv_status = 9;
    else
      indiv_status = 5;
end;

else
  if confagey => 15 and
     endmsg1 = "A" then indiv_status = 1;
  else
    if confagey => 15 and
       endmsg1 = " " then indiv_status = 2;
    else
      if confagey ^= .  and
         confagey < 15 then indiv_status = 3;
      else
        if confagey = . then indiv_status = 4;
run;
```

B.3 BT_STATUS

B.3.1 Summary

BT_STATUS is only defined for cases where INDIV_STATUS = 1. It is based on information from the Biomarker data set.

BT_STATUS	Description
1	Blood test respondent (Interview respondent with valid HIV lab result)
2	Blood test nonrespondent (Interview respondent with no valid HIV lab result)

B.3.2 SAS Code for BT_STATUS

```
ATTRIB BT_STATUS LABEL="Blood test disposition code: 1 = Valid lab results, 2 = No valid
lab results or didn't do BT;
        IF HIV1statusfinalsurvey IN ("Positive" "Negative") THEN BT_STATUS=1;
        ELSE BT_STATUS=2;
```

Note: BT_STATUS = 2 is used for cases with no blood sample taken and also for cases where the blood sample did not result in a definite outcome.

Appendix C

CHAID Trees and Definition of Final Nonresponse-Adjustment Weighting Cells

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

Individual Interview

AD_INDIV_STATUS.pdf (Persons 15+ years)

Blood Test

AM_BT_STATUS.pdf (Males 15+ years)

AF_BT_STATUS.pdf (Females 15+ 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 three 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).

Individual Interview

MW_AD_INDIV.xlsx (Persons 15+ years)

Blood Test

MW_AM_BT.xlsx (Males 15+ years)

MW_AF_BT.xlsx (Females 15+ years)