Developing Nonresponse Weighting Adjustments for Population-Based HIV Impact Assessments Surveys in Three African Countries

Tien-Huan Lin[1], Natalia Weil[1], Ismael Flores Cervantes[1]
Suzue Saito[2,3]
[1]Westat, 1600 Research Blvd, Rockville, MD 20850
[2]ICAP at Columbia University, New York, NY 10032
[3] Department of Epidemiology, Mailman School of Public Health of Columbia University, New York NY 10032

Abstract
In collaboration with national Ministries of Health, the Centers for Disease Control and Prevention, and other partners, ICAP at Columbia University is conducting Population-based HIV Impact Assessment (PHIA) surveys in 12 sub-Saharan African countries. We use data from the first three PHIA surveys—in Malawi, Zambia, and Zimbabwe—to study the effect of survey nonresponse weighting adjustments on HIV prevalence estimates. In developing the nonresponse adjustments, decisions made about the variables to use in the adjustments, and about the formation of nonresponse cells, may affect survey estimates and their standard errors. The nonresponse adjustments in PHIA surveys are made in three stages, first to adjust for nonresponse to the household interview, then for person interview nonresponse in responding households, and finally for failure to obtain analyzable blood samples among persons responding to the interview. A sizable number of variables is available for use in making the nonresponse adjustments for the personal interview and blood sample nonresponse. This paper describes the use of the LASSO regression and CHAID for variable selection in making these nonresponse adjustments. It then examines how the use of different sets of variables employed in the nonresponse adjustments affects the surveys' HIV prevalence estimates.

Key Words: Survey weights, nonresponse weighting adjustment, variable selection, CHAID, LASSO regression.

1. Introduction to the Nonresponse Weighting Adjustments in PHIA Surveys

The expansion of anti-retroviral treatment (ART) to more than 12.1 million people in sub-Saharan Africa is one of the most successful global public health programs ever undertaken (UNAIDS, 2016). It is by far the largest initiative for a single disease, with the United States alone investing over 70 billion dollars since 2002 (Avert, 2016). After a decade of the anti-retroviral therapy scale-up, the Population-based HIV Impact Assessment (PHIA) Project, implemented by ICAP at Columbia University in collaboration with the Ministries of Health, the US Centers for Disease Control and Prevention (CDC) and other partners, is assessing the status of the HIV epidemic in 12 sub-Saharan Africa countries by means of nationally representative surveys that measure HIV prevalence, incidence, and viral load suppression. To date, the PHIA Project has completed seven national household-based bio-behavioral surveys in sub-Saharan Africa. This paper utilizes the data from the first three countries—Malawi, Zambia, and Zimbabwe—whose surveys were concluded in the summer of 2016.

The sample design for PHIA surveys is typically a multistage sample in which the primary sampling units (PSUs) are enumeration areas (EAs) as defined for the last population census, and households are selected at the second stage. The sample individuals includes all eligible adults in all of the selected households, and all children in a random subsample (usually one-half) of the selected households. There are three stages in the data collection process:

1. The first stage consists of a household questionnaire that collects a household roster, including the age and sex of each household member, as well as responses to a range of items about the household (e.g., whether the household possesses certain household items, animals, and types of vehicles; what the main material of exterior walls/floor/roof is; and the number of households using the same toilet facility).

2. The second stage comprises personal interviews with each eligible adult and interviews with a parent of each eligible child. These interviews cover an extensive set of topics such as sexual activity, male circumcision, female reproduction, and HIV/AIDS-related knowledge and attitudes.

3. At the third stage, at the end of the personal interviews, respondents are asked to provide blood samples for HIV testing. The blood samples are then collected by trained phlebotomists.

Nonresponse occurs at each of these three stages of data collection. Various weighting adjustment methods can be used in attempting to compensate for nonresponse (Kalton & Flores Cervantes, 2003). In PHIA, we use standard weighting class adjustment methods (Lessler & Kalsbeek, 1992) at each stage of data collection. The weighting adjustment for household nonresponse in the PHIA surveys is relatively straightforward to implement because the only relevant information about the nonresponding households is geographic location of the EAs. In contrast, there is a great deal of data available for the interview nonrespondents and the blood draw nonrespondents from the data collected at the previous stage or stages. This paper describes the methods used to incorporate these data in the development of the nonresponse adjustments in the PHIA surveys conducted in Malawi, Zambia, and Zimbabwe, and compares them with two other methods with respect to their effect on the HIV prevalence estimates.

The paper is organized as follows. In Section 2, we describe the three PHIA surveys and their sample designs. Section 3 outlines the weighting adjustment methods employed for these surveys and the two alternative methods for creating weighting classes for nonresponse adjustment. Section 4 evaluates the properties of the weights and the estimates produced by the three methods based on measures such as the design effect, and the Akaike Information Criterion (AIC), and the survey estimates of HIV prevalence. We conclude in Section 5 with a discussion of our results and ideas for future research.

2. The Population-based HIV Impact Assessment Surveys

The three PHIA surveys employ nationally representative stratified multistage probability sample designs, with strata defined as the provinces or regions within each country. The primary sampling units (PSUs) were sampled with probability proportional to estimated size (PPES) within each stratum. The measures of size used for the PPES selections are the number of households in the EAs based on the last census. The second-stage sampling units were dwelling units/households that were listed by trained staff for each of the sampled PSUs. Upon completion of the listing process, a systematic random sample of dwelling units/households was selected from each PSU at rates designed to yield a self-weighting (i.e., equal probability) sample within each stratum to the extent feasible. However, sampling rates did differ (sometimes considerably) across strata because of the need to produce regional estimates of HIV prevalence rates and the extent of viral load suppression rates. Finally, in the last stage, all eligible adults ages 15 and over within sampled households were selected for an interview and for a blood test. Eligible children were selected for the blood tests in a random one-half subsample of households. This paper considers only the adult samples.

To be eligible for the PHIA sample, sampled persons needed to meet the PHIA age and residency requirements. In Malawi, an upper age limit of 64 years was specified; in Zambia an upper age limit of 59 years was specified; and no upper age limit was specified for Zimbabwe. The residency requirement was that the person slept in the household the night before the interview (i.e., the “de facto” population). The weighting adjustment methods for adults that are described here were applied to all eligible adults. However, as is widely done in HIV research on adults, a common age range of 15 to 49 years has been used for the analyses unless indicated otherwise. The number of eligible sampled 15 to 49 years old adults in the first three PHIA surveys varied by country from 20,000 to 23,000 persons. Of these, between 17,400 and 19,500 adults responded to the personal interview, and between 15,200 and 17,500 adults provided analyzable blood.

As described in Section 1, nonresponse in PHIA can occur at the three stages of data collection (i.e., household, interview, and blood test). The overall response rates and response rates by sampling stage are shown in Figure 1 for each country. The figure shows that between 84 and 90 percent of the sampled households completed the household questionnaire and provided roster information, between 86 and 88 percent of eligible adults 15-49 years of age in the participating households responded to the individual questionnaire, and between 87 and 91 percent of those adults who responded to the interview provided analyzable blood tests. The overall response rate[1] among adults 15-49 years of age was around 68 percent for all three countries.

[1] The overall response rate is computed as the product of household response rate, interview response rate, and blood test response rate.

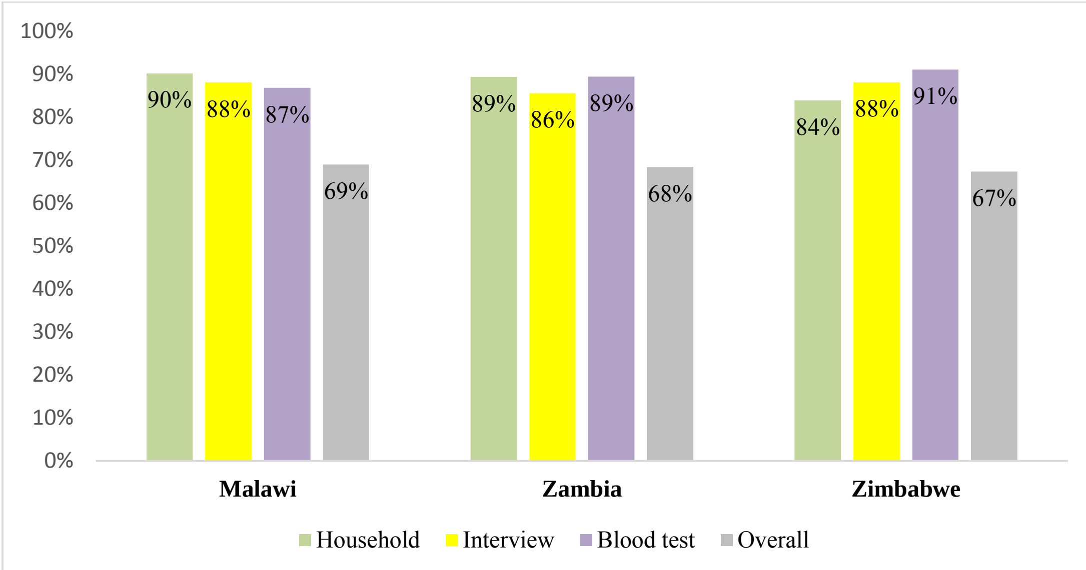


Figure1: Conditional and overall response rates for adults age 15 to 49

3. Weighting Overview in PHIA Surveys

In this section, we describe the weighting adjustment method employed for the PHIA surveys and two alternative methods. The main difference among these methods is the way in which the variables are chosen and used for the creation of the weighting classes.

3.1 General Procedure

Two sets of weights are produced for the analysis of the individual participant level data in PHIA. The first set of weights—*interview weights*—is used to produce population estimates from the person interview data. The second set of weights—*blood test weights*—is used to produce population estimates from both the person interview and the laboratory results of the blood sample. In particular, the latter weights are used to estimate HIV prevalence rates, viral load suppression rates, and HIV incidence rates. The blood test weights are developed by first adjusting the person-level design weights for interview nonresponse, and then adjusting nonresponse-adjusted interview weights for nonresponse to the blood draw. The aim of the latter adjustment is to reduce potential nonresponse bias resulting from missing data for interview respondents who did not consent to the blood test, or when the blood specimen did not yield a successful test result.

Figure 2 shows the weighting steps for the PHIA surveys. The weights are created by sequentially adjusting for nonresponse at each stage of data collection. We start with the household initial weights, computed as the product of the PSU initial weights and the inverse of the household probability of selection within the PSU. The household initial weights are then adjusted to compensate for household interview nonresponse. For adults, the initial interview weights correspond to the nonresponse adjusted household weights since all adults in sampled household are eligible for the person interview. For children, however, the inverse of the subsampling rate of child eligible households is applied to the nonresponse adjusted households weights to obtain the initial interview weights. The interview initial weights are then adjusted for interview nonresponse. Subsequently, the nonresponse adjusted interview weights are adjusted for those who did not consent to the blood draw. Finally, nonresponse adjusted interview weights and nonresponse adjusted blood test weights are poststratified to population control totals to produce the final interview analysis weights and blood test analysis weights.

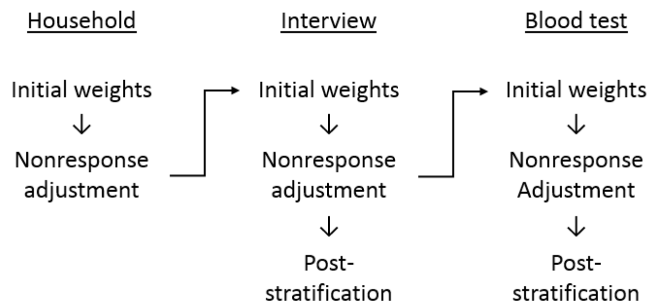


Figure 2: Weighting adjustments in PHIA surveys

The expression for the nonresponse adjusted blood test weight w_{ijk} for person k in household j in EA i can be written as

$$w_{ijk} = w_{ij}^{HH} f_{ij}^{HH} w_{ijk}^{Per} f_{ijk}^{c1} f_{ijk}^{c2} f_{ijk}^{pstc},$$

where w_{ij}^{HH} is the household initial weight, f_{ij}^{HH} is the household nonresponse adjustment factor, w_{ijk}^{Per} is the person interview initial weight, f_{ijk}^{c1} is the person interview nonresponse adjustment factor in weighting class $c1$, f_{ijk}^{c2} is the blood test nonresponse adjustment factor in weighting class $c2$, and f_{ijk}^{pst} is the poststratification factor for poststratum $pstc$.

Our research focuses on the blood test nonresponse adjusted weights, comprising the interview and blood test nonresponse adjustments, prior to poststratification. We compare three methods to compute the nonresponse adjustment factors, f_{ijk}^{c1} (interview) and f_{ijk}^{c2} (blood test), using different definitions of the weighting classes $c1$ and $c2$. The analyses presented in this paper exclude the effect of the poststratification factor.

The nonresponse adjustment factors are computed as the inverse of the weighted response rate in the weighting classes (Brick & Kalton, 1996). These inverses correspond to the nonresponse adjustment factors f_{ijk}^{c1} (interview) or f_{ijk}^{c2} (blood test) described above, and their values depend on how the weighting classes $c1$ and $c2$ are formed. The nonresponse adjustments increase the weights of respondents to represent the nonrespondents in the weighting class with the aim of reducing the bias from nonresponse.

3.2 Two-step Method

In PHIA, at each stage of the nonresponse adjustment process the weighting classes are created through a *two-step* procedure primarily to reduce time and labor. The first step is "feature selection" or variable selection. The "feature selection," or "attribute selection," as known in machine learning, includes techniques for selecting a subset of relevant features or variables for use in model construction. It helps to better explain the impact of the predictors on the dependent variable, in this case response status (James, Witten, Hastie,

& Tibshirani, 2013). Feature selection is also used to enhance generalization by reducing overfitting (Bermingham et al., 2015). The second step employs a tree classification algorithm for creating the weighting classes.

For the PHIA surveys, the feature selection step is performed through Least Absolute Shrinkage and Selection Operation (LASSO) regression, which is a penalized or regularized regression from the field of machine learning (Tibshirani, 1996). The LASSO regression shrinks nonsignificant regression coefficient estimates to zero and produces a simpler model that includes only a subset of the predictors. The main assumption in the LASSO regression is that response status has a sparse model and can be explained by a small subset of predictors. Only the variables predictive of the response variable identified by LASSO are retained for the subsequent step.

In the second step, weighting classes are created using the Chi-squared Automatic Interaction Detector (CHAID) tree classification algorithm (Magidson, 2005). CHAID uses a weighted log-linear modeling algorithm for the computation of chi-square statistics associated with each predictor to classify the sampled individuals (i.e., the respondents and nonrespondents) into cells. The cells are formed sequentially by successive partitions with the aim of forming cells within which individuals have similar response propensities. Nonresponse adjustment factors, computed as the inverses of the weighted response rates within the cells, are applied to the weights of the respondents as described above. CHAID performs a second stage of variable selection because not all the variables entered into the CHAID analysis from the LASSO regression step are used to form a cell (i.e., weighting class). The trees were allowed to grow up to five levels (i.e., depth limit = 5), and the minimum subgroup size (unweighted) required to allow splitting and the minimum terminal node size was set to 50 observations. Cells that either had fewer than 30 respondents or had a weighted response rate of 50 percent or less (or, equivalently, an adjustment factor larger than 2), were combined with neighboring cells.

To limit the creation of sparse cells, some variables were processed prior to the LASSO and CHAID analyses. For example, some derived variables that combine multiple survey responses were created (e.g., a variable that defines the type of construction material of a house—an indicator of economic status—is composed of three variables that describe roof, floor, and wall materials). Other variables were top-coded (e.g., number of rooms in a household or number of households that use one toilet facility) or categorized (e.g., age). Combined variables were created before LASSO, and top-coding and categorizing were usually done after LASSO.

The two-step procedure expedites the creation of the nonresponse adjusted weights. The CHAID analysis with many auxiliary variables is an operationally labor-intensive and possibly error-prone process because each variable needs to be manually classified by type (i.e., continuous, categorical, or ordered variable) and the presence of missing values in the predictors affects this classification. The feature selection procedure prior to CHAID analysis shortens processing time and minimizes manual errors.

In the PHIA surveys, the two-step approach to develop the interview and blood test nonresponse adjustments is implemented separately for different age groups (i.e., adults 15 and over; adolescents 10 to 14; and children 0 to 9). In the CHAID analysis for interview nonresponse, the variable for sex was forced as the first split of the tree for adults and adolescents because males and females tend to have different response propensities in these surveys. In the CHAID analysis for blood test nonresponse, males and females were treated as independent analysis groups because they tend to have different response propensities and because some questionnaire items are only asked of one of the sexes (e.g., male circumcision and female reproduction). All variables from the household roster and questionnaire were used in the LASSO regression analysis of interview nonresponse. The same variables, together with variables from the individual interview questionnaire, were used in the LASSO analysis of blood test nonresponse.

Figure 3 shows the total number of auxiliary variables available for forming the weighting classes and the number of variables selected by LASSO and CHAID for all adults (i.e., not restricted to age 15 to 49) for the two-step interview nonresponse adjustment, by country. Figure 4 provides the same information for the blood test nonresponse adjustment, by sex and by country. Within each country, the left-hand column represents the total number of variables that could be used for constructing weighting classes.

In Figure 3, the left-hand column within a country indicates the total number of variables available from the roster and the household questionnaire for the LASSO regression for the interview nonresponse adjustment. The middle bar (i.e., "After LASSO") provides the number of significant predictors identified by LASSO. The third bar shows the final number of variables selected to create weighting classes, after the CHAID analysis. As an example, for adults in Zimbabwe, 50 roster and household questionnaire variables entered in LASSO, which identified 28 variables for the CHAID analysis (i.e., a 44 percent reduction of predictors of response propensity). Among the countries, the LASSO regression reduced the number of variables by 34 to 44 percent for adults.

In the same figure, the last column within a country (i.e., "After CHAID") shows the final number of variables selected to create weighting classes after the CHAID analysis. Within the CHAID parameters (i.e., depth limit=5, minimum cell size=50), the number of predictors ultimately used to construct the interview response models (i.e., predictive of nonresponse) is on average 30 percent of the total number of available auxiliary variables, but varied by country.

Figure 4 shows the same information except for the blood test nonresponse analysis. It includes the additional variables for modeling from the individual questionnaire and displays the variable selection for adult males and females separately. The extensive amount of data available for modeling comes from the individual questionnaire. For the adult male blood test nonresponse adjustment, there were between 130 and 160 variables available for the LASSO regression. The number of available variables for the adult female nonresponse was between 160 to 210 variables. The number of potential predictors was notably larger for adult females because of the additional questions asked only of female respondents (e.g., reproductive health, cervical cancer screening).

As an example, for adult males in Zimbabwe, 50 roster and household interview variables and 94 individual questionnaire variables were entered into the LASSO regression for the blood response propensity. Out of these 144 variables, LASSO identified a subset of only 42 variables (a 71 percent reduction) as significant predictors of response propensity. Among the countries, the LASSO regression reduced the number of variables by 53 to 71 percent for adult males and by 55 to 73 percent for adult females. A considerable amount of time and resources were saved by employing this method.

The CHAID analysis for the blood nonresponse analysis used the same parameters as the interview CHAID analysis. In the case of Zimbabwe, 15 variables were used to create the adult male weighting classes while 22 variables were used for females. These represent 10 percent of all the adult male variables available for the blood nonresponse adjustment and 13 percent of the available variables for adult females.

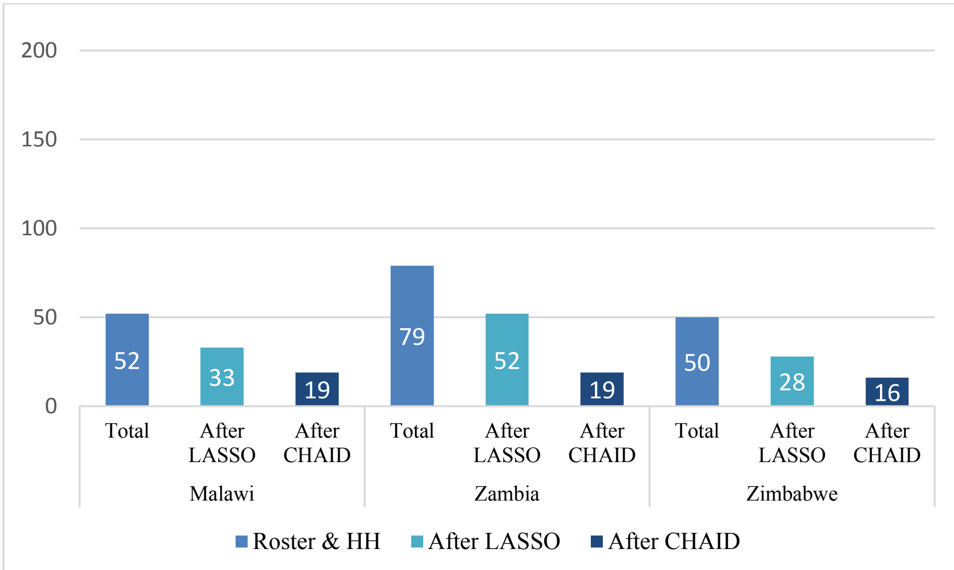
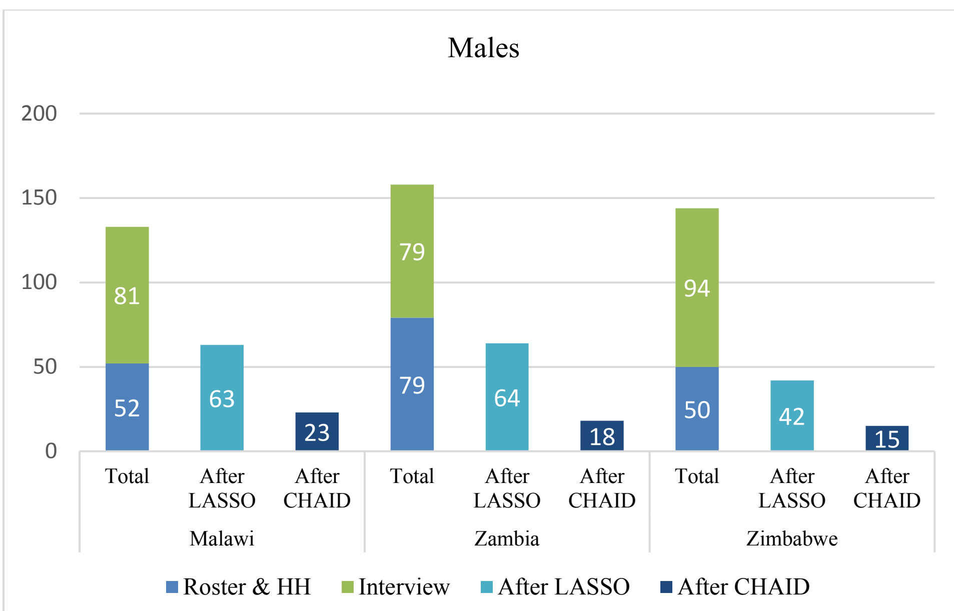


Figure 3: Number of variables selected by LASSO for the interview nonresponse adjustment for adults 15 years or older by country



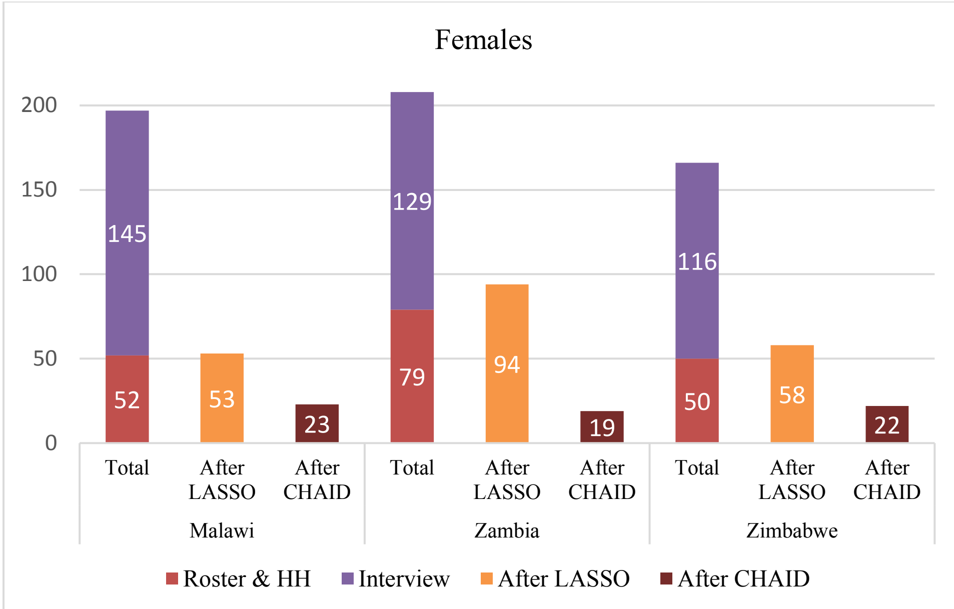


Figure 4: Number of variables selected by LASSO and CHAID for the interview nonresponse adjustment for adults 15 years or older by country

3.3 One-Step Method

We refer to the first alternative method for the creation of the weighting classes as the *one-step* method. This method omits the LASSO feature selection step and relies on only the CHAID algorithm for creating the weighting classes. This method has long been used in health related studies such as the Medical Expenditure Panel Survey (Wun et al., 2004). The setup for the CHAID analysis is the same as in the two-step method described in the previous section. The overall nonresponse adjustment factor in the one-step procedure is the same as in the two-step method, namely it is the product of the interview nonresponse and blood test nonresponse adjustment factors.

Although the one-step method excludes the LASSO step, it is not operationally faster than the two-step method because conducting a CHAID analysis with a large number of auxiliary variables becomes more labor intensive as the number of variable increases.

The comparison between the one- and two-step methods enables us to (a) examine the effect of the LASSO variable selection algorithm prior to running CHAID, (b) determine if the two- and one-step methods identify similar sets of variables to explain nonresponse, and (c) to compare their effects on HIV prevalence estimates.

3.4 Basic Method

The third alternative for forming weighting classes relies on a predetermined set of auxiliary variables. We call this method *basic*, and for this evaluation, the weighting classes are defined by a cross-tabulation of variables for region, urban/rural, and sex, as is done in the Demographic and Health Surveys (ICF International, 2012). In the basic method, the weighting classes do not depend on the response outcomes observed in the selected sample.

Another difference in implementation between the one- and two-step modeling methods and the basic method is that there are two stages of nonresponse adjustments for the one- and two-step methods (i.e., first for interview nonresponse and second for blood test nonresponse), whereas the basic method adjusts for blood sample nonrespondents among all eligible sampled persons in a single adjustment.

An advantage of the basic method is that the nonresponse adjusted weights can be generated quickly with the basic approach because no analysis and review of predictors are required. The method generally produces a smaller number of weighting classes and the weighting adjustments are likely to be less variable, thus resulting in survey estimates with smaller variances. However, if the missing at random assumption does not hold reasonably well within the smaller number of weighting classes, the weighting adjustments may not adequately compensate for nonresponse bias in the survey estimates.

4. Evaluation of Methods for Adjusting for Nonresponse

In this section, we compare various characteristics of the weights created using the two-step, one-step, and basic methods. First, we examine the set of variables used by each method for the creation of the weighting classes. We then compare some statistical properties associated with the nonresponse-adjusted weights produced by these methods: the distributions of the weights and their correlations; design effects; and the Akaike information criterion (AIC). In the last part of the analysis, we examine the differences among the estimates of HIV prevalence rates produced by these methods. As noted earlier, for evaluation purposes, the analyses are restricted to adults ages 15 to 49 as this age group is of particular interest in PHIA, and is common to the three surveys.

4.1 Variables Selected for the Nonresponse Adjustments

Since the variables identified as predictive of response propensity differed among the three methods, the nonresponse-adjusted weights created using the weighting classes formed based on these predictors are different. Table 1 shows the number of variables predictive of blood test nonresponse along with the number of overlaps for each method by sex and country. Of the three variables used in the basic method, sex and region appeared to be predictive of response propensity in the two-step method while urban/rural status did not. While the basic method is able to adjust for a differential response pattern by sex and region, it fails to account for more complicated nonresponse patterns.

Within countries, the predictors selected by both one and two-step methods ranged from 5 to 11 for adult males and 1 to 5 for adult females. These numbers appear low, considering the large number of predictors selected by each method. However, from this analysis alone, it is unclear if the response models from these methods are very different, or if there a high degree of collinearity among the selected predictors in the respective models.

Table 1: Number of variables and classes selected for blood test nonresponse by method and country, all adults

		Nonresponse Adjustment Method							
		Basic		One-step		Two-step		No. overlapping Variables	
Sex	Country	No. Var	No. Classes	No. Var	No. Classes	No. Var	No. Classes	Basic vs. Two-step	One-step vs. Two-step
Male	Malawi	2	12	15	32	23	43	1	5
	Zambia	2	20	18	25	18	23	1	11
	Zimbabwe	2	20	16	25	15	32	1	7
Female	Malawi	2	12	18	41	23	37	1	5
	Zambia	2	20	10	17	19	32	1	2
	Zimbabwe	2	20	11	21	22	44	1	1

4.2 Distribution and Correlation Among the Nonresponse Adjustments

As outlined in the previous section, the two-step, one-step, and basic methods used different predictors for the creation of weighting classes, and there are few overlapping variables between these methods. In this section, we examine the statistical properties of the adjustments produced by the three methods. Figure 5 shows box plots with graphical presentation of the distribution of the overall adjustment factors by country and by method. The overall response adjustment factor for the one- and two-step methods is the product of the interview adjustment and the blood test adjustment. In contrast, the adjustment factor for the basic method comes from the blood nonresponse adjustment alone (i.e., there is no separate interview nonresponse adjustment). We chose to study the adjustment factors instead of the adjusted weights because comparisons of the distribution of the weights are confounded by differences in the sample designs employed by the three countries, which contribute to considerable variation in base weights across regions within country.

Two observations can be drawn from this figure. First, the means of the nonresponse adjustments are almost identical for one- and two-step methods while the means of the basic methods are generally larger. Furthermore, the factors for the basic method are more concentrated, while the one-step and two-step factors are more wide spread and include some large adjustment factors (i.e., greater than 3). This observation is a direct result the number of weighting classes created by each method given in Table 1, indicating that the modeling methods can better account for subgroups that have drastically different response propensity.

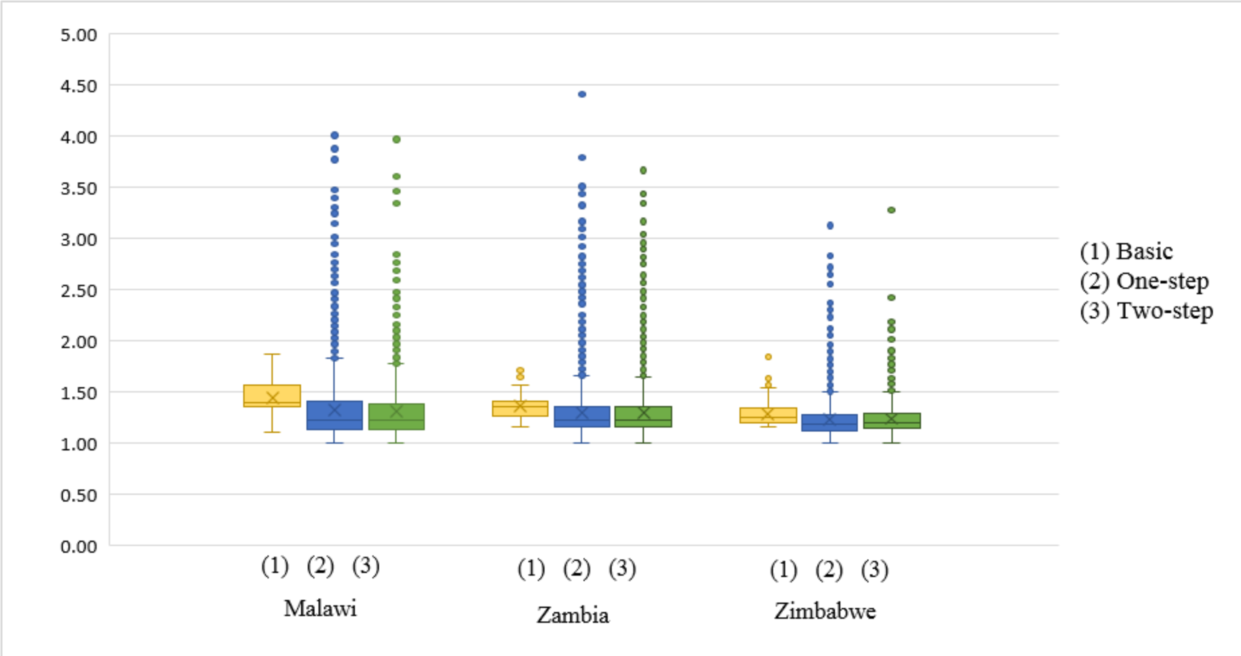


Figure 5: Box-Plot of the nonresponse adjusted weights by method and country

Figure 6 shows the scatter plots of nonresponse adjustment factors produced by the one-step and two-step methods. The one-step factors and the two-step factors are highly correlated in all three countries, and they tend to form a spray of lines. Each line corresponds to a subset of respondents, receiving different adjustment factors from the two methods. The one-step and two-step adjustment factors for Zambia and Zimbabwe differ less than those for Malawi, indicating there is more difference between the response models developed by the two methods in Malawi. Figure 7 shows the scatter plots of the nonresponse adjustment factors produced by the basic and two-step methods. The figure demonstrates that the adjustment factors for the basic method are much more constrained than those for the other methods. The scatter plot consists of a set of horizontal lines that correspond to each of the weighting classes in the basic method. The nonresponse adjustments computed using the basic method and two-step methods are dramatically different.

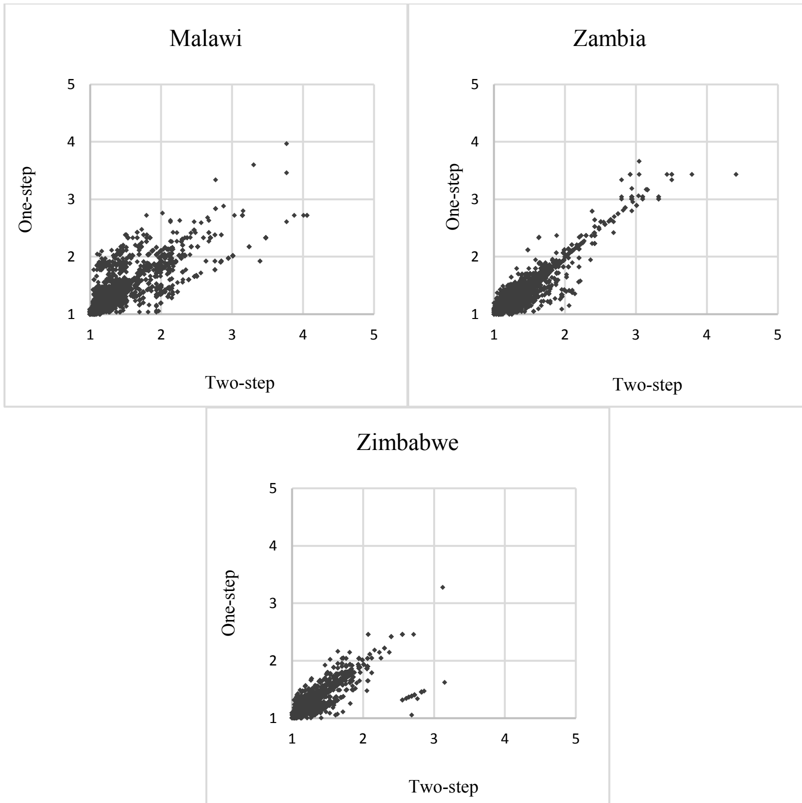


Figure 6: Scatter plots of nonresponse adjustment factors for one-step vs. two-step blood test weights

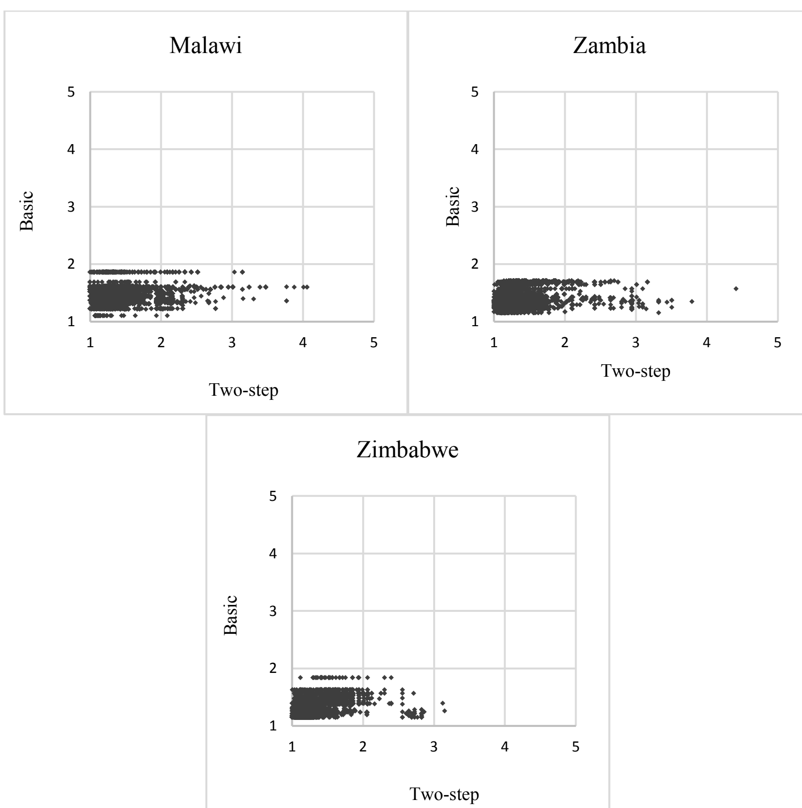


Figure 7: Scatter plots of nonresponse adjustment factors for blood test weight by basic method vs. two-step method

Table 2 shows the correlations of the adjustment factors between the different methods by country. The correlations of the adjustment factors between the one-step method and the two-step method are high, ranging from 0.83 to 0.93, despite the differences in variables identified to be predictive of response status. The correlations of the basic and two-step adjustment factors are much lower, ranging from 0.27 to 0.48.

Table 2: Correlations of the adjustment factors produced by one-step, basic, and the two-step methods

Comparison	Malawi	Zambia	Zimbabwe
One-step vs. Two-step	0.83	0.93	0.86
Basic vs. Two-step	0.27	0.41	0.48

4.3 Design Effects

In this section, we evaluate the three sets of nonresponse adjusted weights through the design effect or *deff* (Kish, 1992). The design effect due to weighting is defined as $deff = 1 + \mathrm{CV}(w)^2$, where $\mathrm{CV}(w)$ is the coefficient of variation of the weights. The design effect can be interpreted as the increase of variance of an estimate due to variation in sampling weights. The *deff* is also computed when developing analysis weights to determine if the nonresponse adjusted weights have become too variable and should be modified (Valliant, Dever, & Kreuter, 2013). Although larger values of *deff* (which implies highly variable weights) are not desired, a small increase of the *deff* after the nonresponse adjustment may mean that the nonresponse weighting adjustment was not successful in reducing nonresponse bias. As expected, the *deffs* of the one-step and two-step weights are larger than the *deffs* for the basic weights; reflecting the fact that the one-step and two-step weights procedures were more effective in capturing variation in response propensity than the basic weights. The *deffs* for the one-step and two-step methods are almost identical, as expected given the similar variation in the adjustment factors for these methods.

Table 3: Design effects for blood test nonresponse adjusted weights by method and county

Country	Base weights	Basic	One-step	Two-step
Malawi	1.29	1.34	1.34	1.35
Zambia	1.06	1.07	1.12	1.12
Zimbabwe	1.10	1.12	1.16	1.17

4.4 Akaike Information Criterion (AIC)

Another criterion for evaluating the nonresponse adjustments is how effective they are in identifying weighting classes with differing response rates. The measure that we use for this purpose is the Akaike information criterion or AIC (Akaike, 1981), a measure that is commonly used to assess the goodness of fit (or quality) of a statistical model in describing the observed data (Greene, 2008). In this evaluation, we use a version of the AIC applicable for survey data to assess the quality of the model that describes the response mechanism in the nonresponse weighting adjustment (Lumley & Scott, 2015). The implicit model behind the weighting class adjustments is a logit model that predicts the propensity to respond (either to the interview or to provide a blood sample) based on predictor variables that are indicator variables for the weighting classes. Smaller values of the AIC that indicate a better model fit to the observed data are desirable.

For this analysis, we compare the AIC of the blood test response model. A more complete approach requires computing the AIC for the combined interview and blood test response model. However, the construction of such a model with two levels of response is beyond the scope of this paper. To make the comparable comparisons, for this analysis alone we applied a two-stage adjustment (i.e., first for interview nonresponse and second for blood test nonresponse) to the basic method using the same cross-tabulation of variables (i.e., region, urban/rural, and sex) for defining weighting classes at both stages and computed the AIC that reflects only the blood test response model. These AIC values are shown in Table 4, by method and by country. These values indicate that the model based on the weighting classes created using the one-step and two-step methods have better fit than the basic model (i.e., smaller AIC values). Therefore, it is expected that the one- and two-step methods may perform better in nonresponse bias reduction for those estimates that are correlated with their weighting classes. However, the table shows mixed results when the models for the one-step and the two-step method are compared. For Zimbabwe, the AIC values of both models are the same, which indicates that the models from the two methods have the same fit. In contrast, the one-step model for Malawi has a better fit while the two-step model for Zambia is better. Even when the AIC values are different, the differences are only marginal.

Table 4: Akaike information criterion (AIC) of response models by method and county for blood test adjustment

	Nonresponse Adjustment Method		
Country	Basic	One-step	Two-step
Malawi	31,600	26,000	26,100
Zambia	27,500	24,300	24,000
Zimbabwe	24,700	21,600	21,600

4.5 Comparison of HIV Prevalence Rates

Finally, we compare estimates of HIV prevalence rates computed using the nonresponse-adjusted weights produced by the one-step, two-step, and basic methods. These estimates and their corresponding 95 percent confidence intervals are presented in Table 5. The table also shows estimates produced using the base weights without any nonresponse adjustment. Estimates produced by the basic method were virtually identical to base-weighted estimates, indicating that the basic adjustment method had very little effect in adjusting for nonresponse bias in the estimates of HIV prevalence rates. Estimates produced using weights generated by the one-step and two-step methods showed stronger effects; with HIV prevalence estimates being lower than those produced using the base weights and the weights produced by the basic methods.

Differences between estimates produced by the basic method vs. the one-step method, and the basic method vs. the two-step method, were statistically significant for all countries for both sexes at the 0.05 significant level[2]. Estimates produced by the one-step method and the two-step method were surprisingly close despite the differences in variables identified to predict response propensity. The only significant differences were for Zambia females

[2] The statistical tests take into account the correlation between two sets of weights with similar nonresponse adjustments and estimates from the same data.

and Zambia overall (i.e., largely driven by females). While this analysis suggests differences in the extent of nonresponse bias adjusted by the basic method versus the modeling methods (i.e., one-step and two-step), it does not provide an answer as to whether residual nonresponse bias remains after the adjustments.

It should be noted that the HIV prevalence rates given in Table 6 are not the official estimates, because they were calculated with the blood test nonresponse adjusted weights but without the poststratified adjustments that are used in forming the final weights for computing the official estimates.

Table 5: HIV prevalence estimates and 95 percent confidence intervals, and tests of differences between methods, by country and method, by sex and overall, for adults 15 to 49

Country	Group	Estimates / 95% Confidence Interval				*p*-value of test of difference of HIV prevalence		
		Base weights	Basic	One-Step	Two-Step	Basic vs. One-step	Basic vs. Two-step	One-step vs Two-step
Malawi	Male	8.2 (7.4, 9.0)	8.5 (7.7, 9.3)	8.0 (7.3, 8.8)	8.0 (7.3, 8.8)	*	*	0.89
	Female	13.4 (12.4, 14.4)	13.4 (12.5, 14.4)	12.5 (11.6, 13.4)	12.5 (11.5, 13.4)	*	*	0.71
	Overall	11.3 (10.6, 12.0)	11.3 (10.6, 12.0)	10.5 (9.9, 11.2)	10.5 (9.9, 11.2)	*	*	0.78
Zambia	Male	8.8 (8.1, 9.5)	8.9 (8.1, 9.6)	8.6 (7.9, 9.3)	8.6 (7.9, 9.3)	0.0005	0.0005	0.98
	Female	15.2 (14.4, 16.1)	15.3 (14.4, 16.1)	14.3 (13.5, 15.1)	14.6 (13.7, 15.4)	*	*	*
	Overall	12.5 (11.8, 13.2)	12.4 (11.7, 13.1)	11.7 (11.0, 12.3)	11.8 (11.2, 12.5)	*	*	*
Zimbabwe	Male	11.5 (10.6, 12.4)	11.5 (10.6, 12.4)	11.0 (10.1, 11.8)	10.9 (10.1, 11.7)	*	*	0.14
	Female	17.2 (16.4, 18.0)	17.4 (16.5, 18.2)	16.3 (15.5, 17.1)	16.4 (15.6, 17.1)	*	*	0.14
	Overall	14.8 (14.1, 15.6)	14.8 (14.1, 15.6)	13.9 (13.3, 14.6)	13.9 (13.2, 14.6)	*	*	0.86

*P<0.0001

5. Conclusions and Future Research Directions

This research compares three alternative methods for creating weighting classes for nonresponse adjustments for the PHIA surveys: (1) the two-step method that first reduces the number of predictor variables using LASSO regression followed by CHAID analysis to create weighting classes using the subset of variables identified in the first step; (2) a one-step method that that uses CHAID to create the weighting classes but does not subset the variables prior to analysis as in the two-step method; and (3) a basic method based on a small set of predetermined geographic and demographic variables.

The one-step method was the most operationally cumbersome to implement, because over 200 potential predictors had to be reviewed and processed. The two-step method was less labor intensive than the one-step method while producing similar HIV prevalence estimates. It maintained the same level of variation in weights, as well as the goodness of fit of the response propensity model as the one-step method, despite the difference in the set of variables identified to be predictive of nonresponse. More research is needed to study the collinearity of the selected variables in the one- and two-step methods and the impact of collinearity on the weighting adjustments.

The basic method was by far the easiest and quickest method to implement. However, it was not as effective as the other methods in compensating for nonresponse bias. The one- and two-step methods gave rise to greater variation in the weight adjustments than the basic method. Although this greater variation did result in some loss in the precision in the survey estimates, the loss was modest. The goodness of fit of the response propensity models for the one- and two-step models were better than that of the basic method, suggesting that these methods better accounted for differential nonresponse. The weights produced by the one- and two-step methods had a stronger impact on adjusting for nonresponse bias in HIV prevalence rates.

Acknowledgements

Special thanks to Graham Kalton for his support and comments.

References

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics,* 16(1), 3-14. doi:10.1016/0304-4076(81)90071-3.

American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys.* AAPOR. Retrieved September 26, 2017, from https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx.

Avert. (2016). *Funding for HIV and AIDS*. Retrieved September 26, 2017, from https://www.avert.org/professionals/hiv-around-world/global-response/funding.

Bermingham, M.L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., . . . Haley, C.S. (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific Reports, 5*, Article number: 10312. doi:10.1038/srep10312.

Brick, J.M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5(3), 215-238. doi:10.1177/096228029600500302.

Greene, W.H. (2008). *Econometric analysis* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

ICF International. (2012). Sampling and household listing manual: Demographic and Health Surveys methodology. Calverton, MD: ICF International.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

Kalton, G., & Flores Cervantes, I. (2003). Weighting methods. Journal of Official Statistics, 19(2), 81-97.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistics*, 8, 183-200.

Lessler, J.T., & Kalsbeek, W.D. (1992). Nonsampling errors in surveys (1st ed.). New York: John Wiley and Sons.

Lumley, T., & Scott, A. (2015). AIC and BIC for modeling with complex survey data. *Journal of Survey Statistics and Methodology*, 3(1), 1-18.

Magidson, J. (2005). SI-CHAID 4.0 user's guide. Belmont, Massachusetts. Retrieved September 26, 2017, from https://www.statisticalinnovations.com/wp-content/uploads/SICHAIDusersguide.pdf.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*, 58(1), 267-288.

UNAIDS. (2016). *Global AIDS update*. Geneva: Joint United Nations Programme on HIV/AIDS. Retrieved September 26, 2017, from http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf.

Valliant, R., Dever, J.A., & Kreuter, F. (2013). *Practical tools for designing and weighting survey samples.* New York: Springer.

Wun, L., Ezzati-Rice, T., Baskin, R., Greenblatt, J., Zodet, M., Potter, F., . . . Touzani, M. (2004). *Using propensity scores to adjust weights to compensate for dwelling unit level nonresponse in the Medical Expenditure Panel Survey*. Agency for Healthcare Research and Quality Working Paper. Retrieved from https://www.ahrq.gov/.