

Appendix C
Sample Weighting

APPENDIX C. SAMPLE WEIGHTING

After the data collection and editing phases of the NSV 2000 were completed, we began the sample weighting phase of the project. We constructed the sampling weights for the data collected from the veterans who responded so they would represent the entire veteran population. The weights were the result of calculations involving several factors, including original selection probabilities, adjustment for nonresponse, households with multiple residential telephones, and benchmarking to veteran population counts from external sources. The weighting process would also correct for noncoverage and help reduce variance of estimates. We produced a separate set of weights for the List and the RDD Samples and then combined them to produce the composite weights for use with the combined sample.

We also constructed a set of replicate weights for each respondent veteran and appended them to each record for use in estimating variances. This chapter describes the calculation of the full sample composite weights and replicate composite weights. We start with a description of the List and RDD Sample weights because the two sets of weights were constructed independently.

C.1 List Sample Weights

The List Sample weights are used to produce estimates from the List Sample that represent the population of veterans who are on the list frame. The steps involved in constructing the List Sample weights are the calculation of a base weight, poststratification adjustment to known list frame population counts, and adjustments to compensate for veterans with unknown eligibility, and for nonresponse. These steps are summarized below.

Calculation of List Sample Base Weights

The base weight for each veteran is equal to the reciprocal of his/her probability of selection. The probability of selection of a veteran is the sampling rate for the corresponding sampling stratum. If

n_h out of N_h veterans are selected from a stratum denoted by h , then the base weight assigned to the veterans sampled from the stratum was obtained as

$$w_{hi} = \frac{N_h}{n_h}; \quad i \in h.$$

Properly weighted estimates using the base weights above would be unbiased if the eligibility status of every sampled veteran could be determined and every eligible sampled veteran agreed to participate in the survey. However, the eligibility status of each and every sampled veteran could not be determined and some were not even located. Moreover, nonresponse is always present in any survey operation. Thus, weight adjustment was necessary to minimize the potential biases due to unknown eligibility and nonresponse. In order to improve the reliability of the estimates we also applied a poststratification adjustment. Normally, the poststratification adjustment is applied after applying the nonresponse adjustment, but we carried this out before the nonresponse adjustment because determining the eligibility status of every veteran on the list frame was not possible.

Poststratification Adjustment

Poststratification is a popular estimation procedure in which the base weights are adjusted so that the sums of the adjusted weights are equal to known population totals for certain subgroups of the population. We defined the poststrata to be the cross classification of three age categories (under 50, 50-64, over 64), gender (male, female), and census regions (Northeast, Midwest, South, and West), which resulted in 24 poststrata. The advantage of poststratified weighting is that the reliability of the survey estimates is improved.

The minimum sample size for poststratification cells was set at 30 veterans. For 2 out of the 24 poststrata, the sample sizes were fewer than 30 veterans so we collapsed these two cells in order to achieve sufficient sample size. Thus, the poststratified weights were computed using population counts from the list frame for 23 poststrata.

Adjustments for Unknown Eligibility and Nonresponse

The List Sample cases can be divided into respondents and nonrespondents. Further, the respondents can be either eligible or ineligible (out of scope) for the survey. The eligibility of the nonrespondent veterans could not always be determined. For example, a sampled veteran who could not be located could have been deceased and hence ineligible for the survey. Therefore, the nonrespondents were classified into two categories: (1) eligible nonrespondents and (2) nonrespondents with unknown

eligibility. In order to apply the adjustments for unknown eligibility and nonresponse, the List Sample cases were grouped into four response status categories:

- Category 1: **Eligible Respondents.** This group consists of all eligible sampled veterans who participated in the survey, namely those who provided usable survey data.
- Category 2: **Ineligible or Out of Scope.** This group consists of all sampled veterans who were ineligible or out of scope for the survey, such as veterans who had moved abroad and were therefore ineligible for the survey.
- Category 3: **Eligible Nonrespondents.** This group consists of all eligible sampled veterans who did not provide usable survey data, but information provided proved they were eligible.
- Category 4: **Eligibility Unknown.** This group consists of all sampled veterans whose eligibility could not be determined.

We used the final List Sample extended interview result codes and other information to assign the sampled veterans to one of the four response categories defined above.

The nonresponse adjustment was applied in two steps. In the first step the poststratified weights of the veterans with unknown eligibility (Category 4) were distributed proportionally over those with known eligibility (Categories 1, 2, and 3). In the second step, we calculated an adjustment factor to account for the eligible nonrespondent veterans.

The final List Sample weight for each eligible respondent was computed by multiplying the base weight by the appropriate nonresponse adjustment factor as defined above. The final List Sample weight for the eligible nonrespondent veterans was set to zero. The final List Sample weight of the out-of-scope/ineligible veterans is the weight obtained after applying the adjustment factor for unknown eligibility. The weights for the out-of-scope/ineligible veterans could be used to estimate the ineligibility rate of the list frame that we used to select the List Sample.

C.2 RDD Sample Weights

The calculation of the RDD Sample weights consisted of five main steps. The steps included computing the base weight and various adjustments at the screener interview level and the extended interview level. In summary, we:

- Computed base weight as the inverse of the probability of selection of the telephone number associated with the household;
- Applied an adjustment to account for household level nonresponse during screening;
- Applied an adjustment for multiple telephone lines as the reciprocal of the number of “regular residential” telephone numbers used by the household;
- Applied an adjustment to correct for the nonresponse to the extended interview; and
- Benchmarked to known veteran population counts from the Census 2000 Supplementary Survey (C2SS) that the U.S. Bureau of the Census conducted.

The final RDD Sample weights were obtained as the product of the base weight and the various adjustments applied to the base weights. The steps involved in computing these weights are summarized below.

RDD Sample Base Weights

The RDD Sample selected included members from the list-assisted RDD sampling methodology and the Puerto Rico RDD Sample. The base weights for the two RDD Samples were defined accordingly.

List-assisted RDD Sample Base Weights

The base weight is defined as the reciprocal of the probability of selection. With the list-assisted RDD methodology, the telephone numbers were selected with equal probabilities of selection. We used a systematic sampling scheme to select telephone numbers, and the probability of selecting a telephone number when n telephone numbers from a pool of N numbers is selected is given by $f = n/N$. Because the national RDD Sample was selected from two RDD frames constructed at two different times we also had to take this into consideration during the process.

Puerto Rico Sample Base Weights

The Puerto Rico RDD Sample was a pure RDD sample due to the fact that information was not available on the telephones to construct the sampling frame for list-assisted RDD methodology. The base weight was defined to be the inverse of the selection probability.

RDD Sample Weight Adjustments

RDD Sample weight adjustments include weight adjustments for the national (list-assisted) RDD Sample and the Puerto Rico RDD Sample.

List-assisted RDD Sample Weight Adjustments

List-assisted RDD Sample weight adjustments were applied as screener interview nonresponse adjustment, adjustment for multiple telephone lines, and an adjustment for nonresponse at the extended interview.

Screener Nonresponse Adjustment. The base weights were adjusted to account for the households (telephones) with unknown eligibility during the screening interview. The adjustment for unknown eligibility was applied in two separate steps. In the first step, we adjusted for those telephones whose type – residential, business, or nonworking – could not be determined. In the second step, nonworking and business telephone numbers were removed and the weights were adjusted to account for the residential telephone numbers for which the eligibility for the NSV 2000 could not be determined.

Adjustment for Multiple Residential Lines. If every household had exactly one residential telephone number, then the weight for a household would be the same as the base weight of the corresponding telephone number. The adjustment for multiple residential telephone households prevents households with two or more residential telephone numbers from receiving a weight that is too large by reflecting their increased probability of selection. A weighting factor of unity was assigned to households reporting only one telephone number in the household, and an adjustment factor of $\frac{1}{2}$ was assigned to households with more than one residential telephone number.

RDD Extended Interview Nonresponse Adjustment. The RDD Sample required administration of both a household screening questionnaire and the extended NSV 2000 questionnaire, and included the possibility of identifying multiple veterans in a single household. Because the screener survey interview screened for the households with potential veterans, a small fraction of persons who were screened in were not actually eligible for the NSV 2000. Once the extended interview began, it was

still necessary to establish with certainty that the selected person was indeed a veteran. If the responses to the set of eligibility questions during the extended interview indicated that the person was not an eligible veteran, the interview was terminated. Moreover, for some cases that were screened in, no information could be collected from the extended interview to ascertain their eligibility (e.g., the potential veteran could not be contacted for the extended interview). Thus, the screened-in sample contained cases with unknown eligibility as well as eligible and ineligible cases. Further, the eligible cases contained respondents and nonrespondents. Therefore, the screened-in RDD Sample cases were grouped into the same four categories as the List Sample cases.

Category 1: **Eligible Respondents**

Category 2: **Ineligible or out of scope**

Category 3: **Eligible Nonrespondents**

Category 4: **Eligibility Unknown.**

The screened-in sample cases were assigned to the four response categories on the basis of final extended interview result codes and other information. The weights of the cases with unknown eligibility (Category 4) were proportionally distributed over the other 3 categories (Categories 1, 2, and 3) and then adjustment factors were calculated.

The next step in the RDD Sample weighting was the extended interview nonresponse adjustment. The RDD extended interview nonresponse adjustment factor was calculated as the ratio of the sum of weights for eligible RDD extended interview respondents and eligible RDD extended interview nonrespondents to the sum of the weights for only the eligible RDD extended interview respondents.

Puerto Rico Sample Weight Adjustments

We screened 96 households with potentially 102 veterans for which extended interviews were attempted. We completed only 51 extended interviews from the Puerto Rico RDD Sample. The nonresponse adjustment factors for the screener interview and extended interview were computed similarly to those for the national RDD Sample except that the screener nonresponse adjustment was computed separately for two age groups (under 60, over 59) and a single nonresponse adjustment was computed for the extended interviews. This was due to the small sample size for the Puerto Rico RDD Sample.

After applying the screener interview and extended interview nonresponse adjustments, the national (list-assisted) RDD and the Puerto Rico RDD Samples were combined into one RDD Sample. The base weights adjusted for nonresponse were further adjusted in a raking procedure, discussed in a later section. The raked weights were the final RDD Sample weights that were used to compute the composite weights for the combined List and RDD Samples.

Comparison of RDD Estimates with VA Population Model Estimates

As a check, we compared the RDD Sample estimate of number of veterans based on the weights before raking with the estimate from the VetPop 2000 model¹, VA population projection model. The NSV 2000 target population includes only noninstitutionalized veterans living in the U.S. The reference period for the NSV 2000 is the year 2000. The VA population model estimates are also for the year 2000 and these are based on the 1990 Census. These estimates are derived by incorporating survival rates and information on veterans leaving military service. The VA population model estimate for the entire veteran population is 25,372,000 veterans, whereas the estimate from the RDD Sample is 23,924,947 veterans, which is 5.7 percent lower than the VA population model estimate. The difference of 5.7 percent can be attributed to the combination of the differences from exclusion of the institutionalized veterans and RDD undercoverage of nontelephone households and households with unlisted telephone numbers belonging to “zero-listed telephone banks.”

The portion of undercoverage due to nontelephone households and households with unlisted numbers belonging to “zero-listed telephone banks” was addressed with the raking procedure, described in the next section. The control total of veteran population for the raking procedure was 25,196,036 veterans. Thus, the estimated undercoverage due to nontelephone households and households with unlisted telephone numbers belonging to “zero-listed telephone banks” would be only about 5.0 percent. After correcting for the undercoverage from these two sources, the difference between the NSV 2000 and the Vetpop 2000 estimates is less than one percent, which is from institutionalized veterans and veterans living abroad.

Raking Ratio Estimation/Undercoverage Adjustment

The raking ratio estimation procedure is based on an iterative proportional fitting procedure and involves simultaneous ratio adjustments to two or more marginal distributions of the population counts. The purpose of the raking procedure in this survey is to improve the reliability of the survey estimates, and to correct for the bias due to missed households, namely, households without telephones and households with unlisted telephone numbers belonging to “zero-listed telephone banks.”

The raking procedure is carried out in a sequence of adjustments. First, the base weights are adjusted to one marginal distribution and then to the second marginal distribution, and so on. One

¹ The VetPop 2000 is a veteran population projection model developed by the office of the Actuary, Department of Veterans Affairs. It is the official VA estimate and projection of the number and characteristics of veterans as of September 30, 2000. Details of all aspects of the development and content of the model are available from the office of the Actuary, Department of Veterans Affairs, 810 Vermont Avenue NW, Washington DC 20420.

sequence of adjustments to the marginal distributions is known as a cycle or iteration. The procedure is repeated until convergence is achieved.

We used a two-dimensional raking procedure for the RDD Sample. The first dimension was formed from the cross classification of three age categories (under 50, 50-64, over 64) with four education levels (no high school diploma, high school diploma, some college, bachelor's degree or higher) and four race categories (Hispanic, Black, Other, and White), resulting in 48 cells. The second dimension was formed from the cross classification of gender (male, female) and the four census regions (Northeast, Midwest, South, and West), resulting in 8 cells. (The above variables were chosen as the raking variables due to significant differences in the telephone coverage by categories of these variables, and hence maximum bias reduction would be achieved.)

We used the Census 2000 Supplementary Sample (C2SS) data from the U.S. Bureau of the Census to define the control totals for the raking procedure. We also included the Puerto Rico RDD Sample in the raking procedure. Because the C2SS did not include Puerto Rico in the survey target population, we estimated the Puerto Rico veteran population counts for the year 2000 from the Census 1990 population counts based on a model.

We applied the convergence criteria in terms of percent absolute relative difference, which was specified to be no more than 0.01 percent for all marginal population counts. The raking procedure converged in 8 iterations.

C.3 Composite Weights

Integration of samples from multiple frames into a single micro-data file with a single weight requires, at a minimum, the ability to tell which of the veterans had more than one chance of selection. This is enough to create unbiased weights. The Social Security numbers (SSNs) of all the veterans on the list frame were known. To identify the RDD Sample veterans on the list frame, we needed to obtain their SSNs during data collection so that the overlap RDD Sample would be identified by matching the SSNs of the veterans in the RDD Sample with the list frame. However, out of 12,956 completed extended RDD interviews (including Puerto Rico), we were able to obtain an SSN from only 6,237 veterans, which is 48.1 percent of the RDD completed extended interviews. The veterans sampled as part of the RDD Sample could thus only be categorized as belonging to the overlap RDD Sample or nonoverlap RDD Sample if the SSN was reported. For others (those who did not report their Social Security numbers), we used a prediction model to impute the overlap status. (See Chapter 6 of the Methodology Report for details on how this imputation was carried out.)

The veterans in the overlap RDD Sample (including the imputed cases) also had a chance of being selected in the List Sample, and hence, had an increased chance of selection. These RDD cases are referred to as the overlap sample because they represent the portion of the RDD frame that overlaps with the list frame. A composite weight was created for the identified overlap RDD Sample (both observed and imputed) and List Sample cases using the principles of composite estimation so that the combined RDD and List Sample file could be used for analysis.

Calculation of Composite Weights

Composite weights were calculated using an approach that would take into account the design effects of the RDD and List Sample designs when combining the two samples. The List (7,092 completed interviews) and RDD (12,956 completed interviews) Samples were combined into one file resulting in a combined sample of 20,048 completed extended interviews.

In composite estimation, the estimates being combined are assumed to be independent, and are unbiased estimates of the same population parameter. In other words, the List Sample and the overlap RDD Sample cases theoretically represent the same population (i.e., veterans on the list frame). Therefore, a linear combination of the two independent estimates would also produce an unbiased estimate.

The composite weights were calculated by computing the estimates of proportions and their variances for 16 statistics identified as key variables by the VA for the List Sample and the overlap portion of the RDD Sample. These variables are listed in Table 1.

Table 1. VA key variables

MB24:	Combat or War Zone Exposure (Yes/No)
DIS1:	Ever Applied for VA Disability Benefits (Yes/No)
HB21:	Currently Covered by Medicare (Yes/No)
HC1:	Emergency Room Care During Last 12 Months (Yes/No)
HC4a:	VA Paid for Emergency Room Care (Yes/No)
HC5:	Outpatient Care During Last 12 Months (Yes/No)
HC6:	VA Facility for Outpatient Care (Yes/No)
HC9:	Hospitalized Overnight in a VA Hospital (Yes/No)
SD14d:	VA Service Connected Disability Compensation in 2000 (Yes/No)
SD14e:	VA Non-Service Connected Pension in 2000 (Yes/No)
SD14j:	Income Source: Public Assistance in 2000 (Yes/No)
ET1:	Ever Received Any VA Education or Training Benefits (Yes/No)
ML3a:	Ever Used VA Loan Program to Purchase Home (Yes/No)
ML3b:	Ever Used VA Loan Program for Home Improvement (Yes/No)
ML3c:	Ever Used VA Loan to Refinance Home (Yes/No)
PRIORITY:	Priority Group (Mandatory/Discretionary)

Raked Composite Weights

The composite weights obtained by combining the List and RDD Samples were also raked using the same two dimensional raking procedure that was used for the RDD sample raking. The RDD Sample was raked mainly to correct for undercoverage because of nontelephone households and households with unlisted numbers in the “zero-listed telephone banks” that were missed in the list-assisted RDD sampling methodology. The composite weights were raked to achieve consistency with the C2SS estimates, and to improve the precision of the survey estimates.

C.4 Replicate Weights

A separate set of replicate weights was created for the RDD Sample and the List Sample. These were then combined to construct the preliminary composite replicate weights. The final composite replicate weights were obtained by using the same two dimensional raking procedure with the preliminary composite replicate weights as the input weights that were used for the composite full sample weights.

List Sample Replicate Weights

A set of 51 Jackknife 1 (JK1) replicate weights was created for the List Sample for use in variance estimation. The replicate weights were designed for the JK1 replication method. To create the replicate weights, the entire List Sample, including ineligible and nonresponding veterans, was sorted by the twelve sampling strata, and by the order of selection within strata.

The same adjustments applied to the full List Sample base weights to obtain the full List Sample final weights were applied to the replicate base weights to obtain the List Sample replicate final weights. This included poststratification and the extended interview nonresponse adjustments that were recalculated for each replicate, so that the sampling variability in the response rates would be captured in the replicate weights.

RDD Sample Replicate Weights

A set of 51 JK1 replicate weights was also created for the veterans identified from the RDD Sample. Once the he replicate base weights were calculated, they were adjusted following the same steps as those applied to the full sample base weights. These included the screener level nonresponse adjustment, adjustment for multiple residential telephone lines, extended interview level nonresponse adjustment, and raking to the external veteran population counts obtained from the Census 2000 Supplementary Survey. By raking the replicate weights in the same manner as the full sample weights,

the sampling variability in the raking adjustment factors would be reflected in the replicate weights, and hence included in the overall variance estimate. The raking procedure was carried out on the combined national and Puerto Rico RDD Samples.

If there were two or more veterans in a household, each respondent in the household received the same set of replicate base weights but the adjusted weights could differ because they could belong to different adjustment cells.

Composite Replicate Weights

Composite replicate weights were created for each replicate weight from the List Sample and overlap RDD Sample cases. The remaining RDD Sample cases were assigned composite replicate weights equal to their original RDD Sample replicate weights. Finally, the composite replicate weights were raked to the veteran population counts estimated from the C2SS in a two dimensional raking procedure as was done for the composite full sample weights. The convergence criteria for the composite replicate weights was modified so that the percent absolute relative difference was no more than 0.1 percent for all marginal population counts.

C.5 Reliability of the Survey Estimates

Estimates obtained from the sample data are different from figures that would have been obtained from complete enumeration of the entire veteran population. Results are subject to both sampling and nonsampling errors. Nonsampling errors include biases from inaccurate reporting, processing, and measurement, as well as errors from nonresponse and incomplete reporting. These types of errors cannot be measured readily. However, to the extent possible, each error has been minimized through the procedures used for data collection, editing, quality control, and nonresponse adjustment. The variances of the survey estimates are used to measure sampling errors. The variance estimation methodology used for NSV 2000 is discussed below.

Estimation of Variances of the Survey Estimates

The variance of an estimate is inversely proportional to the number of observations in the sample. In other words, as the sample size increases, the variance decreases. For the NSV 2000 the variance estimation methodology for estimates of totals, ratios (or means) and difference of ratios is based on the JK1 replication method. Additionally, we have constructed the composite full sample and composite replicate weights for the combined List and RDD Samples corresponding to the JK1 replication

methodology. The WesVar² variance estimation system can be used to produce the survey estimates based on the composite full sample weights and the corresponding variances of these estimates.

Construction of Confidence Intervals

Each of the survey estimates has an associated standard error, which is defined as the square root of the variance of the estimate. Consider the example of estimating the proportion of veterans with a certain characteristic, such as a service-connected disability. We denote by \hat{p} the estimated proportion of veterans with the particular characteristic of interest and let $v(\hat{p})$ be the corresponding variance estimate. Then the standard error of the estimated proportion \hat{p} is given by

$$se(\hat{p}) = \sqrt{v(\hat{p})}.$$

The 95 percent confidence interval is the interval such that the unknown proportion p would have a 95 percent probability of being within the interval. The 95 percent confidence interval is given by

$$\hat{p} \pm t_{(0.025,50)} \times se(\hat{p}).$$

The factor $t_{(0.025,50)}$ is the t -value at $\alpha = 0.025$ with 50 degrees of freedom, which is approximately equal to 2.0. The smaller the half-width of the confidence interval, the more precise is the survey estimate.

Alternatively, the precision of the survey estimate can also be expressed in terms of the coefficient of variation (cv) of the estimate. The cv of an estimate is defined as the ratio of the standard error of the estimate and the magnitude of the estimate expressed in percent. Thus, the cv of the estimated proportion \hat{p} is given by

$$cv(\hat{p}) = 100.0 \times \frac{se(\hat{p})}{\hat{p}},$$

where $se(\hat{p})$ is the standard error of the estimated proportion \hat{p} . The smaller the cv of the estimate, the more precise is the estimate. The percent margin of error at the 95 percent confidence level can also be obtained by multiplying the cv of the estimate by the factor $t_{(0.025,50)}$.

² WesVar is software for analyzing data from complex surveys. The software was developed by Westat and can be downloaded from Westat's website (www.westat.com/wesvar) for a 30-day free trial.

C.6 Bias and Precision in the Combined Sample

We investigated two main issues associated with the use of the combined sample versus the separate RDD and List Samples. These were: (1) potential biases incurred in the estimates as a result of the matching involved in creating the composite weights, and (2) the gains in precision from the increased sample sizes of the combined sample. The reason that both of these issues are important is that the total mean square error (MSE) of a survey estimate is equal to the sum of its variance and the square of the bias, ($MSE = Variance + (Bias)^2$). In surveys with large sample sizes, the MSE may be dominated by the bias term. When sample sizes are small, the variance may be a greater cause for concern.

To address the first issue of bias, the potential risk of bias would be due mainly to imputing the overlap status of those RDD sample respondents who did not provide their Social Security numbers. The question arises as to whether the cases that reported an SSN are different from those that did not. To answer this question, statistical comparisons were made for the two groups to see whether their distributions differed with respect to age and to other key statistics. It was determined that the risk of potential bias was minimized due to imputing the overlap status within homogenous imputation cells.

The precision of the estimates can be evaluated by comparing the standard errors (SEs) of the estimates from the combined sample with those from the RDD Sample alone. In this situation, the population of analytical interest is the population of all noninstitutionalized veterans living in the U.S. The statistics of interest for the purpose of this analysis are proportions for various key statistics identified by the VA. This analysis showed the increased sample sizes of the combined sample always result in a significant reduction in sampling variability. The standard errors of the combined estimates are always lower than the standard errors of the corresponding estimates from the RDD Sample alone.

The efficiency of the combined sample can also be defined as the ratio of the corresponding variances expressed as percentage. We denote by $Eff\left(Combined\ vs.\ RDD\right)$ the efficiency of the combined sample as compared with the RDD Sample alone, then

$$Eff\left(Combined\ vs.\ RDD\right) = 100 \times \frac{\text{var}(RDD)}{\text{var}(Combined)},$$

where $\text{var}(RDD)$ and $\text{var}(Combined)$ are, respectively, the variances of the RDD Sample alone and the combined sample. Efficiency values of more than 100 percent imply that the combined sample estimates are more efficient than the estimates based on the RDD Sample alone. After calculations, efficiencies are greater than 100 percent for all variables of interest in this analysis and the efficiency values range from 104 percent to 316 percent. Thus, the combined sample with the corresponding composite weights should be used for all VA analyses.