Efficient Estimation of Response Rates when a Small Subsample of Nonrespondents is Selected for Follow-Up Conversion

Avinash C. Singh[1], Vincent G. Iannacchione[2*], Shijie Chen[2] and Jill A. Dever[3]**

[1)] Statistics Canada
Ottawa, Canada. Email: singavi@rogers.com

[2)] RTI international
1615 M ST N.W. Suite 740, Washington, DC 20036
Emails: vince@rti.org[*] and schen@rti.org[**]

[3)] Joint Program in Survey Methodology
1218 LeFrak Hall, University of Maryland, College Park, MD 20742
Email: jdever@survey.umd.edu

Abstract

For surveys with a nonresponse follow-up, the usual double-sampling estimator offers a way to measure the combined response to the initial survey and the follow-up. However, when cost considerations require that the follow-up sample size be small, the double-sampling estimator may be unstable. In this paper, we use dual-frame estimation techniques to obtain a more precise composite response rate (CRR) which combines the unstable but unbiased double-sampling estimator with the stable but biased estimator of response rate from the main survey after being bias-corrected.

Under a population response model, each unit in the population is assigned a random response indicator of 1 or 0 before the actual sampling takes place. Suppose the parameter of interest is the total number of respondents if the survey were administered to the entire population. In this case, the response rate parameter is the total number of respondents divided by the population size. For approximately unbiased estimation of this population response rate, it is natural to use the design-weighted estimator especially if the sampling design is non-ignorable.

When a survey is subject to low response, a follow-up of initial nonrespondents can provide a better understanding of bias attributable to nonresponse. In this situation, it is desirable to estimate the response rate if all individuals in the entire population who are nonrespondents to the main survey but subject to possible conversion via a follow-up are also included. For this new parameter, the response rate estimator from the main survey will be biased because of under-reporting because it doesn't use the respondents who could have been converted.

To correct for under-reporting bias among the nonrespondents in the main survey, we propose to use a mass imputation of means under a super-population model which models the probability of an individual responding to the follow-up survey conditional on the individual being nonrespondent to the main survey. We use the sampling weight calibration approach to fit this imputation model using respondents from the main and the follow-up surveys. We illustrate the proposed method of estimating CRRs using data from the 1991 Gulf War Veterans Survey.

Key Words: Response rate as finite population parameter; Response under-reporting; Follow-up surveys; Composite response rate; Dual-frame estimation.

1. The Population Response Model

For surveys with a nonresponse follow-up, the estimated response rate (un-weighted or weighted) based on the combined survey is, by definition, greater than the corresponding estimated response rate from the main survey. This observation leads to some interesting questions: What are the corresponding population parameters? Do they bear the same relationship as the estimates? These questions can be addressed by considering the difference between the corresponding finite population parameters under a population response model. Under this model, specific to the survey instrument and conditions, each unit in the population is assigned random response indicators of 1 or 0 before actual sampling takes place; see e.g., Fay (1991).

For surveys with a nonresponse follow-up, each unit $k \in U$ is assigned two response indicators r_{1k} and r_{2k} where

$$r_{1k} = \begin{cases} 1, \text{ if k responds to the main survey} \\ 0, \text{ otherwise} \end{cases}$$

and

$$r_{2k} = \begin{cases} 1, \text{ if k responds to the follow-up given that } r_{1k} = 0 \\ 0, \text{ otherwise.} \end{cases}$$

Clearly, r_{2k} is needed only for the domain $\{k \in U : r_{1k} = 0\}$. We can assume $r_{2k} = 1$ if $r_{1k} = 1$. Let $u_k = r_{1k} + (1 - r_{1k}) r_{2k}$. The parameter of interest is $N^{-1} T_r$ where $T_r = \Sigma_{k \in U} u_k$ i.e., response rate when a person is respondent to either the main or the follow-up survey. Note that the follow-up survey is administered to nonrespondents of the main survey.

Let

S_A = full sample for the main survey (respondents and nonrespondents) , and

S_F = full sample for the follow-up survey.

S_B = respondents to the main survey and full sample for the follow-up survey.

Also let,

R_A = respondents to the main survey, and

R_F = respondents to the follow-up survey.

R_B = respondents to the main survey and respondents to the follow-up survey.

We note that $S_B = R_A \cup S_F$ and $R_B = R_A \cup R_F$. u_k is observed in R_A and S_F , but not for nonrespondents to the main survey (S_A) who are also not selected for the follow-up survey (S_F), i.e., $S_A \setminus S_B$.

2. The Composite Response Rate Estimator

First, we perform Hàjek ratio adjustments to each full-sample (S_A, S_B) and then use the Generalized Exponential Model (GEM) approach (Folsom and Singh 2000) to post-stratify the two full samples, S_A and S_B . The Hàjek adjustment helps to make the GEM post-stratification closer to optimal. Now, one estimator of T_r is the double-sampling estimator based on S_B , i.e.,

$$\hat{T}_{r(B)} = \Sigma_{k \in S_B} u_k w_{kB,ps} ,$$

where $w_{kB,ps}$ denotes the GEM post-stratified weight with respect to sample S_B .

However, the double-sampling estimator is not expected to be precise when the follow-up sample is small. An alternate estimator $\hat{T}_{r(A)}$ can be constructed from S_A as

$$\hat{T}_{r(A)} = \Sigma_{k \in S_A} \tilde{u}_k w_{kA,ps}$$

where $w_{kA,ps}$ denotes the GEM post-stratified weight with respect to sample S_A ,

$$\tilde{u}_k = \begin{cases} u_k \text{, if } r_{1k}=1 \text{ or } r_{1k}=0 \text{ and } r_{2k} \text{ is observed} \\ u_k^I \text{, otherwise} \end{cases},$$

and u_k^I is the imputed value of u_k. In S_A, u_k is imputed only in $S_A \setminus S_B$.

Now, assume an imputation model ξ: $r_{2k} \mid r_{1k=0} = \mu_k(\lambda)$, where $\mu_k(\lambda)$ is modeled by the GEM. In order to use this model to impute r_{2k} for nonrespondents $(r_{1k}=0)$ in S_A, we need to assume that the model holds for the nonrespondents (i.e., $r_{1k}=0$) in *U*. In other words, the nonresponse mechanism (for r_{1k}) is ignorable for the model ξ. To estimate λ, we will use data from the follow-up. If we only use the follow-up sample S_F, then we also need to assume that the super-population model ξ continues to hold for the $\{r_{1k}=0\}$ subset of S_A, i.e., the first-phase (or main survey) sample design is ignorable for ξ. To avoid this assumption, we should work with the full sample S_B involving both the first- and second-phase samples, and not just the follow-up sample in fitting model ξ. This can be done as follows.
With predictors x for model ξ, we use GEM calibration equations to estimate λ.

$$\Sigma_{R_A} x_{kA} w_{kB,ps} + \Sigma_{R_F} x_{kF} w_{kB,ps} a_k(\lambda) = \Sigma_{S_A} x_{kA} w_{kA,ps},$$

where $a_k(\lambda) = 1/\mu_k(\lambda)$.

Also, *x*-predictors should include variables observed in the main survey (such as easy or difficult to contact) which are expected to be good predictors, and are available for S_A— the full first-phase sample. Besides avoiding the assumption of first-phase design ignorability for model ξ, the above calibration equation uses controls based on the larger sample S_A rather than the usual full follow-up sample S_F (i.e., r_{2k} respondents and nonrespondents). Since control totals based on S_A are more reliable than those based on S_F, we expect λ-estimates to be more precise with the above calibration approach.

Note that GEM is only used to fit ξ as an estimation tool, i.e., we don't use the calibration weights produced by GEM. Once λ is estimated, we impute u_k by its mean $\mu_k(\lambda)$ by substituting the appropriate values of the covariates $\underset{\sim}{x}_k$ in the model. The estimator $\hat{T}_{r(A)}$ can now be constructed. It provides a $\pi\xi$-consistent estimator of T_r because when λ is known, we have

$$E_\pi\left(\hat{T}_{r(A)}\right) = \Sigma_{\substack{k\in U \\ r_{1k}=1}} u_k + \Sigma_{\substack{k\in U \\ r_{1k}=0}} \mu_k(\lambda)$$
$$= \Sigma_{k\in U} u_k + \Sigma_{\substack{k\in U \\ r_{1k}=0}} \left(\mu_k(\lambda) - \mu_k\right).$$

Therefore, $E_{\xi\pi}\left(\hat{T}_{r(A)}\right) = \Sigma_{k\in U}\mu_k(\lambda) + 0$ because ξ is assumed to hold for the subset $\{r_{1k}=0\}$ of *U*.
Now, if λ is estimated,

$$N^{-1}\hat{T}_{r(A)}\left(\hat{\lambda}\right) = N^{-1}\hat{T}_{r(A)}(\lambda) + O_p\left(\frac{1}{\sqrt{n_F}}\right)$$

and so, at least asymptotically, $\hat{\lambda}$ contribution to the bias is negligible. However, in finding the variance of the response rate estimator, $\hat{\lambda}$ variability needs to be taken into account (this is automatic in jackknife or bootstrap estimation).

Now, we use DFC-methodology (Singh and Wu, 2003 and Singh, Iannacchione and Dever, 2003) to obtain $\hat{T}_{r(A\square B)}^{DFC}$ from the two expansion estimates $\hat{T}_{r(A)}$ and $\hat{T}_{r(B)}$. In DFC with GEM with pre-specified parameters:

$$\ell_{kA} < c_{kA} < u_{kA} \quad \text{and} \quad \ell_{kB} < c_{kB} < u_{kB},$$

the adjustment factors are modeled as

$$a_{kA} = \frac{\ell_{kA}(u_{kA} - c_{kA}) + u_{kA}(c_{kA} - \ell_{kA})\exp_A}{(u_{kA} - c_{kA}) + (c_{kA} - \ell_{kA})\exp_A}, \text{ and } a_{kB} = \frac{\ell_{kB}(u_{kB} - c_{kB}) + u_{kB}(c_{kB} - \ell_{kB})\exp_B}{(u_{kB} - c_{kB}) + (c_{kB} - \ell_{kB})\exp_B}$$

where

$$\exp_A = \exp[\eta_A^{-1} A_{kA}(\mathbf{x}'_{kA}\boldsymbol{\lambda}_{\mathbf{x}A} + r'_{kA}\boldsymbol{\lambda}_{\mathbf{z}})], \text{ and } A_{kA} = \frac{(u_{kA} - \ell_{kA})}{(u_{kA} - c_{kA})(c_{kA} - \ell_{kA})}$$

Here, we use several zero functions corresponding to r-variables for various study domains. That is the reason we prefer to use the DFC method rather than the normal optimal regression approach. In DFC, we need to choose η_A, a parameter representing the effective sample size of design *A* with respect to design *B,* either for each variable or as a common value for all. Here, we can choose for each r-variable, because we don't need to produce one set of final weights.

We implement the following steps to calculate the response rate estimator:

<u>Step I:</u> Hàjek –ratio adjustment to design weights from S_A and S_B.
<u>Step II:</u> Apply GEM poststratification adjustment to S_A and S_B.
<u>Step III:</u> Fit $\xi : E\left(r_{2k} = 1 \mid r_{1k} = 0\right) = \dfrac{1}{a_k(\lambda)}$ using GEM for nonresponse adjustment with a centering factor of c = inverse of response propensity to the follow-up.
<u>Step IV:</u> use $\hat{\lambda}$-estimates from step III to impute means $\mu_k(\lambda) = a_k^{-1}(\lambda)$ for all $k \in \{r_{1k} = 0 \text{ in } S_A\}$.
<u>Step V:</u> Perform DFC for the concatenated samples $\{S_A^* \cup S_B^*\}$ with several zero-controls and poststratification controls for given η_A for each grid γ in (0, 1) in increments of 0.05.
<u>Step VI:</u> Find optimal η_A for each r-variable and the corresponding variance estimate (using Jackknife method).
<u>Step VII:</u> Report CRR estimates $N^{-1}\hat{T}^{DFC}_{r(A\cup B)}$ for each r-variable and the corresponding standard error (SE) and confidence interval (CI).

3. Application

The Tenth Anniversary Gulf War Veterans Health Survey (GWHS) is a national probability-based survey of men and women who served in the 1991 Persian Gulf War within all branches of the U.S. Armed Forces. The primary objectives of the study are (1) to provide national estimates of Gulf War veterans who report significant health concerns and (2) to model the key correlates of those health concerns. Other objectives include comparisons between active-duty military and reservists, and the development of separate explanatory models for the occurrence of health concerns in male and female veterans. The objective of the sample design for this study was the selection of a probability sample of veterans from the target population of sufficient size to support these analytic objectives.

The *target population* for the GWHS is the more than 685 thousand men and women who served in the 1991 Persian Gulf War with all branches of the U.S. Armed Forces. We selected a stratified systematic sample of 10,301 veterans from the sampling frame maintained by Defense Manpower Data Center. We defined four primary strata by subdividing active-duty military and reservists by gender. Within each primary stratum, veterans who had registered with Department of Defense's Gulf War Comprehensive Clinical Evaluation Program (CCEP) and received a medical diagnosis based on International Classification of Diseases, 9th Revision were over-sampled to obtain a sufficient number of veterans reporting significant health concerns. Additionally, the frame was sorted by race/ethnicity to ensure a representative sample.

The survey originally was implemented as a mail survey in 2001. An overall response rate of 54.4 percent (RR3 definition from AAPOR 2004) was achieved after three mailings of the instrument, as well as a reminder post card, and a reminder telephone call. Response rates to the mail survey were highest among females, reservists, and those who had been evaluated by the CCEP.

The response rate to the mail survey was 20 percentage points lower than expected. In an effort to reduce the potential bias associated with nonresponse to the mail survey, the project team decided to conduct a telephone follow-up of a sub-sample of nonrespondents to the mail survey. We based the follow-up sub-sample size of 1,000 mail nonrespondents (about one-fifth of all mail nonrespondents) on funding available to the study.

We allocated the follow-up sample inversely proportional to the mail response rates of each stratum. Prior to selection, each mail nonrespondent was classified as probable 'easy to contact' or 'difficult to contact' based on whether an interviewer had made contact with someone in the veteran's household during calls made to prompt the return of the mail survey. Mail nonrespondents classified as 'easy to contact' were over-sampled to increase the expected effective sample size of the follow-up. To decrease response burden, the telephone follow-up obtained information on 69 of the 151 questions included in the mail survey.

As **Table 1** illustrates, the un-weighted (i.e. RR3) response rates to the Gulf War survey do not account for differential selection rates that were used to select the sample. These un-weighted response rates sometimes can lead to biased conclusions about the response propensity of the surveyed population. For example, the 54.4 percent response rate (using the AAPOR RR3) to the mail portion of the survey was almost seven percentage points higher than the corresponding design-weighted response rate (WRR) of 47.6 percent. The disparity was the result of over-sampling females and reservists, two groups who responded at noticeably higher rates than males and active-duty personnel. For the telephone follow-up, we achieved a 55.1 percent response rate (AAPOR RR3) which was noticeably lower than the WRR of 50.0 percent because we over-sampled mail nonrespondents with updated contact information. The response patterns for the follow-up were similar to the mail survey although the largest increase in response rate occurred among active-duty males not evaluated by the CCEP.

The WRR estimates the response rate that would be obtained if everyone in the target population were selected for interview. In addition, treating the response rate as a population parameter enables us to calculate the sampling variance of the WRR in a design-consistent fashion. Associating a measure of precision with response rates provides important information about their reliability and provides a statistical basis for comparing response rates among important subgroups.

A total of 5,709 eligible sample members responded to either the mail survey or the telephone follow-up, another 441 were found to be ineligible (i.e., non Gulf War veterans); and the remaining 4,151 were nonrespondents. If the RR3 formulation is used, the overall response rate for the survey is 59.7 percent. However, the RR3 formulation is misleading because only a sub-sample of initial nonrespondents was selected for follow-up. In this situation, the WRR in the form of a double-sampling estimator (DSE) provides a design-consistent estimator of the combined response to the initial survey and the follow-up.

While the usual double-sampling estimator provides an unbiased estimate of the combined response rate, it may be unstable especially if the follow-up sub-sample is small. Again using the survey of Gulf War veterans as an example, the double-sampling estimator of the combined WRR among female reservists was 81.2 percent compared to a combined WRR of 74.4 percent among females on active duty. However, the 6.8 percentage point difference was not statistically significant because of the large sampling errors associated with the DSE estimators.

We used dual-frame calibration (DFC) to develop a composite estimator of the overall response rate for surveys with a nonresponse follow-up. To correct for under-reporting bias among the nonrespondents in the main survey, we used a mass imputation of means under the super-population model which models the probability of an individual responding to the follow-up survey conditional on the individual being nonrespondent to the main survey. Then, we use the GEM weight calibration approach to fit this imputation model using respondents from the main and the follow-up surveys.

We used the steps described in **Appendix A** to develop a composite response rate (CRR) that strikes a balance between variance and bias. We used a grid search of the relative effective sample size parameters η_A and η_B (=1–η_A) in the interval (0, 1) to determine the optimal (i.e., minimum variance) CRR. A comparison of weighted response rates using the design weights and the post-stratified weights is shown along with the CRRs for various study domains in **Table 2**. In general, the CI half-widths associated with the CRRs are smaller than those associated with the post-stratified DSEs reflecting the improved efficiency of dual-frame calibration.

4. Summary and Discussion

For surveys with a nonresponse follow-up, a composite response rate (CRR) estimator (a type of dual-frame calibration estimator) was proposed to improve the efficiency of the usual double-sampling estimator which is a weighted response rate. Using the population response model, we demonstrate that the weighted response rate provides an unbiased estimate for non-ignorable designs unlike the un-weighted response rate obtained using the AAPOR RR3 Definition.

An important underlying notion associated with calculating a CRR for a survey with a nonresponse follow-up is that response is under-reported in the initial or first phase of the survey. For the Gulf War survey, we assumed that under-reporting occurred because at least some of the sample members who did not respond by mail were likely to respond to the telephone follow-up. This notion of under-reported response can be extended to mixed-mode surveys. For example, assume that, instead of selecting a sub-sample of 1,000 initial (mail) nonrespondents to the Gulf War Survey, we select an *independent* sample of 1,000 Gulf War sample members for telephone interview. In this case, the overall response propensity is still under-reported in the initial (i.e., mail) sample because (by the same model) there are sample members who are unlikely to respond by mail but will respond by phone. However, under-reporting of response also can be assumed in the sample selected for telephone interview. That is, there are at least some sample members who are unlikely to respond by telephone but will respond to the mail. Here the issue of households with regular mailing address but without landline telephone or households with telephones but without regular mailing addresses can be dealt with by assuming that each household can be interviewed in either mode, at least conceptually.

This under-reporting of response in both samples motivates the use of mass imputation of response propensities for nonrespondents to both surveys. An interesting implication of this constraint is that the CRR could be greater than either of the individual response rates. For example, if the response rate to the mail portion was 50 percent and the response rate to the telephone portion was 35 percent, the CRR could exceed 50 percent. (The actual magnitude of the CRR would depend on what adjustment factors are needed to satisfy the specified control totals.) Ultimately, a CRR that exceeds the RR of either sample may be reasonable, especially if one is willing to accept the notion that the overall response propensity is under-reported in each sample.

5. References

American Association for Public Opinion Research (2004), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 3rd edition*. Lenexa, Kansas: AAPOR.

Fay, R.E. (1991), "A Design-Based Perspective On Missing Data Variance," *Proceedings of the Annual Research Conference, US Bureau of the Census*, pp. 429-440.

Folsom Jr., R.E., and Singh, A.C. (2000), "A Generalized Exponential Model for Sampling Weight Calibration for a Unified Approach to Nonresponse, Post-stratification, and Extreme Weight Adjustments," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 598-603.

Singh, A.C., and Wu, S. (2003), "An Extension of Generalized Regression Estimator to Dual-Frame Surveys," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 3911-3918.

Singh, A.C., and Wu, S. (1996), "Estimation for Multiframe Complex Surveys by Modified Regression," *Proceedings of the Statistical Society of Canada, Survey Methods Section*, pp. 69-77.

Singh, A.C., Iannacchione, V.G., and Dever, J.A. (2003), "Efficient Estimation for Surveys with Nonresponse Follow-Up Using Dual-Frame Calibration," *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 3919-3930.

Table 1. Comparison of Weighted and Un-weighted Response Rates

	Initial Survey			Follow-Up Survey		
Study Domain	Sample Size	RR3 (%)	WRR*(%)	Sample Size	RR3 (%)	WRR*(%)
Overall	10,301	54.4	47.6 ± 2.9	1,000	55.1	50.0 ± 3.5
Active Males	4,489	50.4	45.9 ± 3.6	472	52.3	47.7 ± 5.2
Active Females	2,311	54.7	53.5 ± 6.0	218	54.6	47.6 ± 7.8
Reserve Males	2,310	58.4	53.3 ± 4.0	208	58.7	53.9 ± 7.9
Reserve Females	1,191	60.9	60.0 ± 5.3	102	61.8	58.1 ± 8.9
Males	6,799	53.1	47.0 ± 3.1	680	54.3	49.6 ± 4.4
Females	3,502	56.8	55.4 ± 4.5	320	56.9	50.8 ± 6.1
Active Duty	6,800	51.8	46.4 ± 3.4	690	53.0	47.7 ± 4.3
Reserves	3,501	59.2	54.1 ± 3.5	310	59.7	55.2 ± 6.0
On CCEP	4,274	59.9	58.2 ± 1.7	372	57.8	53.5 ± 5.4
Not on CCEP	6,027	50.5	47.0 ± 3.0	628	53.5	48.0 ± 4.6
NH Black	3,062	42.8	31.6 ± 3.6	364	51.1	46.4 ± 5.5
NH White	6,104	60.0	53.2 ± 3.0	512	55.3	50.3 ± 4.9
Other Race	1,135	55.1	46.4 ± 4.9	124	66.1	60.1 ± 9.4

*Weighted response rates and 95% confidence intervals calculate using design weights.

Table 2. Comparison of Weighted Response Rates*

	Design Weights		Post-Stratified Weights*		DFC Weights	
Study Domain	Initial Survey	Initial and FU Surveys	Initial Survey	Initial and FU Surveys	CRR	η_A
Overall	47.6 ± 2.9	72.6 ± 3.5	TBD	72.6 ± 3.5	TBD	TBD
Active Males	45.9 ± 3.6	71.4 ± 4.3		71.4 ± 4.3		
Active Females	53.5 ± 6.0	74.4 ± 6.4		74.4 ± 6.4		
Reserve Males	53.3 ± 4.0	77.4 ± 5.3		77.7 ± 5.2		
Reserve Females	60.0 ± 5.3	81.2 ± 6.6		80.3 ± 7.0		
Males	47.0 ± 3.1	72.3 ± 3.8		72.3 ± 3.8		
Females	55.4 ± 4.5	76.4 ± 4.9		76.1 ± 5.0		
Active Duty	46.4 ± 3.4	71.6 ± 4.1		71.6 ± 4.1		
Reserves	54.1 ± 3.5	77.9 ± 4.7		78.1 ± 4.6		
On CCEP	58.2 ± 1.7	79.1 ± 3.4		79.0 ± 3.4		
Not on CCEP	47.0 ± 3.0	72.3 ± 3.7		72.3 ± 3.7		
NH Black	31.6 ± 3.6	60.7 ± 7.4		60.6 ± 7.4		
NH White	53.2 ± 3.0	76.3 ± 4.0		76.3 ± 4.0		
Other Race	46.4 ± 4.9	75.0 ± 8.5		74.9 ± 8.6		

*Post-stratified weights for the initial survey include an adjustment for under-reporting.

Appendix A.

Calculation of Composite Response Rates for the Gulf War Veterans Health Survey

1. Define two over-lapping samples.
s_A = 10,301 veterans initially selected for the survey; and,
s_B = 5,599 mail respondents plus the 1,000 mail nonrespondents selected for follow-up.
Note that s_B is a proper subset of s_A.

2. Assign the design weights.
d_{kA} = the inverse of the selection probability assigned to the k^{th} sample member.
$d_{kB} = d_{kA}$, if the k^{th} sample member responded to the mail survey. For the 1,000 mail nonrespondents selected for the follow-up, d_{kB} equals d_{kA} times the inverse of the follow-up selection probability.
Note that $\sum_{s_A} d_{kA} = \sum_{s_B} d_{kB}$ = 685,074 veterans on the sampling frame.

3. Construct variance replicates.
We created 294 variance replicates (a.k.a. random groups) that enable us to combine the data obtained from the mail survey with that obtained from the telephone survey and then use the jackknife method to estimate the variances of survey outcomes in a design-consistent fashion. Within each of the eight first-phase strata, we randomly assigned 35 sample members to each replicate with the requirement that each replicate have approximately equal numbers of mail respondents and at least one follow-up respondent.

4. Obtain post-stratification totals.
A total of 20 control totals T_x were created using combinations of the following variables:

- Component (2) = Active, Reserve
- Gender (2) = Male, Female
- Race/Ethnicity (3) = NH Black, NH White, Other
- CCEP (2) = Present on CCEP, Absent from CCEP

Note that the post-stratification totals sum to 689,183 veterans and reflect slightly more complete totals that those for the sampling frame.

5. Post-stratify the design weights.
We used GEM to calculate Hajek-ratio adjustment factors $a_{kA,PS}$ and $a_{kB,PS}$ for all the selected units that were applied to the design weights to force them to sum to the 20 control totals. The resulting weights are written as

$$w_{kA,PS} = d_{kA}\ a_{kA,PS} \text{ and } w_{kB,PS} = d_{kB}\ a_{kB,PS}.$$

Note: $\sum_{s_A} w_{kA,PS} = \sum_{sB} w_{kB,PS}$ = 689,183 veterans.

6. Impute follow-up response propensity.
Because the response rate is under-reported for s_A, we constrained the adjustment factor ($a_{kA,PS}$) to be greater than one for the mail respondents. Using the standard GEM notation, we have

l_{kA} (=1) < c_{kA} < u_{kA} (= maximum) for the s_A respondents;

Additionally, we set the centering factors (the desired mean of the distribution of the adjustment factors) to the following:

c_{kA} = 1 / WRR for s_F

Then, we imputed the follow-up response propensity by applying the inverse of the adjustment factor to the 3,702 phase 1 nonrespondents not selected for the follow-up.

7. Obtain Response Domains.

We created a total of 14 sample-specific response indicators for the following reporting domains. These are referred to as the zero controls in the text.

- Overall (1)
- Component by Gender (4)
- Component (2)
- Gender (2)
- CCEP (2)
- Race/Ethnicity (3)

8. Calculate DFC adjustment factors.

We used GEM to calculate DFC adjustment factors $a_{kA,DFC}$ and $a_{kA,DFC}$ that were applied to the adjusted design weights so that the difference between the set of weighted response rates for s_A and s_B was zero (14 zero controls) while maintaining the 20 control totals. The resulting weights are written as:

$$w_{kA,DFC} = d_{kA}\ a_{kA,PS}\ a_{kA,DFC}$$

$$w_{kB,DFC} = d_{kB}\ a_{kB,PS}\ a_{kB,DFC}.$$

All the DFC adjustment factors were constrained to be positive; the centering factors were set to one.

Note: $\sum_{s_A} w_{kA,DFC} = \sum_{s_B} w_{kB,DFC}$.

Using the η grid, we determined that a scaling constant (η_A) of 0.80 and higher minimized the variances for all the 14 response domains.

9. Construct replicate sets of DFC weights.

We constructed 294 replicated sets of DFC weights for use in the jackknife method of variance estimation. Each set of weights was constructed by first excluding one variance replicate per stratum, and then adjusting the weights of the remaining replicates within the same stratum for the subsampling, and (3) then applying the DFC methodology (Step 7) to the remaining respondents

This process was repeated until each of the 294 variance replicates had been excluded from one replicated set of weights. For each replicated set of DFC weights, we used GEM to calculate jackknife adjustments factors $a_{kA,AJDFC}$ and $a_{kB,AJDFC}$ that were applied to the design weights from Step 7 so that the DFC constraints were maintained.

$$w_{kA,AJDFC} = d_{kA}\ a_{kA,PS}\ a_{kA,DFC}\ a_{kA,AJDFC}$$

$$w_{kB,AJDFC} = d_{kB}\ a_{kB,PS}\ a_{kB,DFC}\ a_{kB,AJDFC}.$$

10. Determine optimal CRRs.

We calculated CRRs using the DFC adjustment weights ($w_{kA,AJDFC}$, $w_{kB,AJDFC}$). The half-width confidence intervals were calculated for a range of η_A using the Jackknife weights in SUDAAN. Note that the Jackknife point estimates are not equivalent to the DFC point estimate because the calibration adjustment factor is non-linear.