Efficient Estimation for Surveys with Nonresponse Follow-Up Using Dual-Frame Calibration

Avinash C. Singh, Vincent G. Iannacchione, and Jill A. Dever
A.C. Singh, RTI, 3040 Cornwallis Road, Research Triangle Park, NC 27709 asingh@rti.org

Abstract

In surveys where response rates are low, a follow-up survey of nonrespondents may be used to augment the respondents from the main survey. Using the theory of double sampling for stratification, estimates from this combined sample provide a less-biased alternative to nonresponse-adjusted estimates from the main survey. This is due to the bias-correction limitations of the main survey nonresponse model in the presence of high nonresponse. However, when cost considerations require that the follow-up sample size be small, the reduction in bias of estimates obtained from the combined sample may be negated by the increase in sampling variance due to variability in selection probabilities between the main and the follow-up samples. In this situation, a possible solution might be to trim the extreme weights to reduce the mean square error (MSE) associated with key survey estimates. However, it is not clear how to define a model to measure and control bias introduced by trimming.

We present an alternative in which we make more efficient use of information in the data to construct estimates by minimizing MSE under joint design- and model-based randomization. Analogous to the small area estimation setting, our goal is to obtain a composite estimator that strikes a balance between variance of the unstable estimator based on the main and follow-up samples and bias of the stable estimator based on the main sample only. However, this situation it is a large area and not a small area problem, and so the dual frame estimation framework can be used for its formulation. Moreover, composite weights can be obtained from weight calibration with built-in controls for extreme weights while preserving the known population control totals as well as zero control totals for difference estimates from the two samples for a key set of study variables. The proposed method is illustrated for a survey of Gulf War veterans with a nonresponse follow-up survey.

Key Words: dual-frame estimation, extreme weights, mean square error, nonresponse bias, small-area estimation, weight calibration.

1. Introduction

An important application of double sampling for stratification involves the use of a follow-up survey to temper the effect of nonresponse in the main survey (Hansen and Hurvitz, 1946; Cochran p370, 1977). In this situation, nonrespondents are classified as a distinct stratum based on the outcome of the initial (or first-phase) sample attempt. The rationale for the application is based on the population response model (for a recent reference, see e.g., Fay, 1991; and Shao and Steel, 1999) which under given survey conditions assumes that a random response indicator can be assigned to each unit in the population before actual sampling takes place. However, the membership of the units in the nonresponse and response strata is not known until the initial or first phase of data collection has been completed. For the second phase, a sub-sample of the initial nonrespondents is drawn and fielded for follow-up, usually with a more intense effort than was used to field the initial sample.

In theory, follow-up samples can be used to alleviate the limitations of model-based nonresponse adjustments in survey estimates that use only the main sample because bias adjustments via nonresponse modeling may not be adequate for surveys with high nonresponse. However, the reality is that less-than-complete response to a follow-up dilutes the actual amount of bias reduction thus requiring some nonresponse modeling. Cost considerations typically limit the size of the follow-up sample thus increasing the variability of estimates due to the high unequal weighting effect (UWE) in the combined sample. Even so, a nonresponse follow-up can provide important information about nonrespondents and nonresponse bias especially for surveys with low response rates. There is an important additional benefit of a follow-up survey which constitutes the main purpose of this paper. It is shown that unlike the traditional estimator based on double sampling for stratification, a more efficient estimator can be developed by an alternative way of combining information from main and follow-up samples.

For surveys of populations where expected response rates and information about potential nonresponse bias can be obtained from prior studies, an optimum sub-sampling fraction among the initial nonrespondents can be determined and integrated into the double-sampling procedure before the survey begins (Cochran 1977, p331). This allows for efficiency considerations at the design stage. However, despite optimization at the design stage, precision of estimates for certain study variables may not be adequate. The present paper considers the problem of improving efficiency at the estimation stage. Note that for surveys of populations where no information about response rates is available *a priori*, the decision to conduct a follow-up may be made during the survey based on lower-than-expected response rates. In this situation, efficiency considerations at the estimation stage become more important because (1) the sub-sampling fraction applied to the initial nonrespondents may be less than optimum, and (2) the reduction in bias obtained from the follow-up may be offset by the increase in sampling variance due to the UWE.

A possible solution to the UWE problem in the combined sample is to trim the weights to reduce mean square error (MSE) associated with key survey estimates (Potter, 1990). The goal of weight trimming is to reduce sampling variance of an estimate more than enough to compensate for the possible increase in bias caused by the use of trimmed weights. However, it is not clear how to model for measuring and controlling the bias introduced by weight trimming. If the problem of a high UWE is attributed to a few extreme values, then a solution might be to use a weight calibration method with unit-specific range restrictions such as the generalized exponential model (GEM) of Folsom & Singh (2000). GEM is a general unified weight-calibration model that controls for extreme weights by incorporating pre-specified upper- and lower-bounds both for extreme and non-extreme weights into the adjustment factors made for nonresponse and post-stratification. Thus extreme weight values are controlled while desired population control totals are maintained during weight calibration.

As an alternative, we borrow ideas from composite estimation for dual-frame surveys to extract more information from the data. In our case, the composite estimator is a weighted combination of two correlated estimators:

- The first estimator ignores the follow-up and uses only the first-phase sample (adjusted for unit nonresponse). This estimator is expected to be stable but potentially biased in spite of nonresponse model adjustments since the model is limited in bias correction when response rates are low. This can be viewed as a quasi-model based estimator because modeling for nonresponse plays a major role as the relative proportion of nonrespondents in the sample is high.
- The second estimator combines the first-phase sample with the follow-up (or second-phase) sample. This estimator is expected to be relatively unbiased, but unstable especially if the follow-up sample is small. Typically, another model is needed to adjust for nonresponse in the follow-up. This can be viewed as a quasi-design based estimator because here modeling for nonresponse plays a minor role due to the relative small proportion of non-respondents in the sample.

The proposed method is motivated in **Section 2** using a small area estimation analogy, but is formulated using the dual-frame calibration (DFC) method of Singh and Wu (1996, 2003) after suitable modifications for dependence between samples. After providing a brief overview of DFC, we discuss how the GEM calibration method can be used to produce range-restricted (e.g., nonnegative) weights with built-in control on extreme sampling weights. Additionally, we discuss how the method was adapted to include extra controls provided by the two estimators. A step-by-step description of the proposed DFC method is presented in **Section 3**. The problem of variance estimation is considered in **Section 4**. We show that after Taylor linearization, the standard variance estimator for double sampling for stratification can be used. A simpler variance estimate similar to the case of single-phase multi-stage sampling is also proposed when the follow-up sample is nested within primary sampling units (PSUs). We present empirical results based on an application to a survey of Gulf War Veterans in **Section 5**, and we summarize our approach with concluding remarks in **Section 6**.

2. Motivation and Formulation of DFC Method

To improve efficiency of survey estimators, we need to incorporate as much relevant information as possible at the estimation stage. To achieve this end, as mentioned in the introduction, we consider the following two estimators of a population total T_y corresponding to a study variable y.

First, let s_A denote respondents from Phase 1 (i.e., the main survey) and p_1 the corresponding probability sampling design. We define an estimator $\hat{T}_{y(A)}$ suitably adjusted for Phase 1 nonresponse under model ξ_1, where the subscript A signifies that only respondents from Phase 1 are used in the estimation. Clearly for this estimator to be useful in view of anticipated low response to Phase 1, we need to obtain good nonresponse predictors (possibly from administrative sources) to link respondents and nonrespondents under the model ξ_1.

Next, let s_B denote respondents from both Phase 1 and Phase 2 (i.e., the main survey and the follow-up) with corresponding probability sampling designs p_1 and p_2. We define a second estimator $\hat{T}_{y(B)}$ suitably adjusted for nonresponse in Phase 2 under model ξ_2, where the subscript B signifies that all respondents are used in the estimation. This is similar to the usual estimator employed in double sampling for stratification.

The estimator $\hat{T}_{y(A)}$ is approximately unbiased under the design p_1 and the corresponding nonresponse model ξ_1. Similarly, the estimator $\hat{T}_{y(B)}$ is approximately unbiased under design p_1p_2, and model ξ_2. The estimator $\hat{T}_{y(A)}$ is stable but likely to be biased because of the limitations of ξ_1, while $\hat{T}_{y(B)}$ is unstable but likely to be nearly unbiased. Now the problem of combining these two estimators can be motivated by small area estimation (SAE) in that the combined estimator should exhibit a suitable trade-off between bias of $\hat{T}_{y(A)}$ and variance of $\hat{T}_{y(B)}$. As in SAE, a composite estimator can be formed by a convex linear combination such that the MSE defined jointly under model $\xi_1=(\xi_1, \xi_2)$ and design $p=(p_1,p_2)$ is minimized. Note that it is not possible to get a stable estimate of bias without a model. So we assume models ξ_1 and ξ_2 hold true for computing MSE. However, since the model ξ_1 is deemed to be tenuous, the resulting composite estimator is expected to be biased but closer in value to though more stable than $\hat{T}_{y(B)}$.

The above analogy with SAE is only used to explain the trade-off between variance and bias. Our situation is not a SAE problem since there is no need to borrow information from other domains of interest via a separate model. In our case, modeling is only used to connect nonrespondents to respondents as in any large area estimation problem. So, the problem can be formulated in a dual-frame estimation framework by regarding the two samples s_A and s_B as being drawn from two identical frames U_A and U_B, respectively. However, unlike the usual goal of minimizing variance, here the goal is to minimize MSE. Note that under the specified designs and models, MSE can be estimated like the usual variance estimator under the usual assumption that the $V_\xi E_p$ component is negligible compared to the $E_\xi V_p$ component. So for all practical purposes, we can treat the new dual-frame estimation problem as the usual one based on minimizing variance.

Typically in the dual-frame setting the two frames overlap partially and the two samples are independent. In our case, however, the two frames are identical and the two samples are dependent, as they share the same set of respondents s_A from Phase 1. Dual-frame calibration (DFC) for constructing combined estimates as expansion estimators was considered by Singh and Wu (1996, 2003) using linear regression estimation for simultaneous post-stratification of samples from two frames. This is done such that the control totals corresponding to the usual auxiliary variables (x) are satisfied as well as some new control totals of zero corresponding to new auxiliary variables (z). Here the z-variables denote certain key study variables used in the composite estimation method which are collected for both samples. The variable y is used to denote an arbitrary study variable which may be one of the z-variables which are directly controlled in composite estimation.

We will first briefly review the DFC methodology before discussing the necessary modifications needed to account for dependent samples. For simplicity in illustration, we consider how the problem of combining two unbiased and independent estimators, $\hat{T}_{y(A)}$ and $\hat{T}_{y(B)}$, of a common population total T_y can be cast into a calibration problem with a new constraint defined in terms of the study variable y and a control total of zero. Suppose we have two simple random samples (SRS) of sample sizes n_A and n_B (the total sample size is denoted by n) from a population of size N. An optimal linear combination to minimize the variance is given by

$$\hat{T}_{y,opt} = \alpha_{opt}\, \hat{T}_{y(A)} + \left(1-\alpha_{opt}\right)\hat{T}_{y(B)} \qquad (2.1a)$$

where,

$$\alpha_{opt} = V(\hat{T}_{y(B)})\left[V(\hat{T}_{y(A)}) + V(\hat{T}_{y(B)})\right]^{-1} \qquad (2.1b)$$

For SRS, we have

$$\hat{T}_{y(A)} = \sum_{s_A} y_{kA}\, d_{kA}$$

where, $d_{ka} = N/n_A$ and the usual variance estimate is

$$\hat{V}\left(\hat{T}_{y(A)}\right)=\left(1-\frac{n_A}{N}\right)\left(\frac{N}{n_A-1}\right)\sum_{s_A} y_{kA}\, d_{kA}\,\left(y_{kA}-\overline{y}_A\right) \quad (2.1c)$$

The variance estimate for $\hat{V}_{y(B)}$ is defined similarly.

Therefore,

$$1-\hat{\alpha}_{opt}=\frac{\sum_{s_A} y_{kA}d_{kA}\left(y_{kA}-\overline{y}_A\right)\left(1-\frac{n_A}{N}\right)\left(\frac{N}{n_A-1}\right)}{\hat{V}\left(\hat{T}_{y(A)}\right)+\hat{V}\left(\hat{T}_{y(B)}\right)} \quad (2.1d)$$

Now, rewriting $\hat{T}_{y,opt}$ as

$$\hat{T}_{y,opt}=\hat{T}_{y(A)}+(1-\hat{\alpha}_{opt})\left(\hat{T}_{y(B)}-\hat{T}_{y(A)}\right) \quad (2.2a)$$

we have,

$$\begin{aligned}\hat{T}_{y,opt} &= \\ &\sum_{s_A} y_{kA}d_{kA}\left[1+c_{kA}\left(y_{kA}-\overline{y}_A\right)\hat{V}^{-1}\left(0-\left(\hat{T}_{y(A)}-\hat{T}_{y(B)}\right)\right)\right] \\ &\approx \sum_{s_A} y_{kA}d_{kA}[1+(n_A/n)^{-1}(\hat{\lambda}_{yA}/n)(y_{kA}-\overline{y}_A)] \\ &= \sum_{s_A} y_{kA}d_{kA}a_{kA}=\sum_{s_A} y_{kA}w_{kA}\end{aligned} \quad (2.2b)$$

where,

$$\hat{V}=\hat{V}(\hat{T}_{y(A)})+\hat{V}(\hat{T}_{y(B)})\,;$$

$$c_{kA}=\left(1-\frac{n_A}{N}\right)\frac{N}{n_A-1}\approx\frac{N}{n_A}\ ,$$

ignoring the finite population correction (fpc);

$$\hat{\lambda}_{yA}=N\left[\hat{V}\left(\hat{T}_{y(A)}\right)+\hat{V}\left(\hat{T}_{y(B)}\right)\right]^{-1}\left(0-\left(\hat{T}_{y(A)}-\hat{T}_{y(B)}\right)\right);$$

and

$$a_{kA}=1+(n_A/n)^{-1}(\hat{\lambda}_{yA}/n)(y_{kA}-\overline{y}_A)\ .$$

This formula is similar to a linear regression calibration estimator. Note that unlike the usual regression estimation for single frame surveys with auxiliary x-variables, here the λ-parameter in the adjustment factor is scaled by the inverse of the relative effective sample size n_A / n. It may be instructive to note that the larger sample, as expected, tends to have smaller adjustments (i.e., the factors are closer to 1). In other words, weights for each sample are differentially adjusted; the adjustments are smaller for the sample with higher the relative sample size so that the two estimates become identical.

Similarly, we can write $\hat{T}_{y,opt}$ in terms of w_{kB}. We have

$$\hat{T}_{y,opt}=\hat{T}_{y(B)}+\hat{\alpha}_{opt}\left(\hat{T}_{y(A)}-\hat{T}_{y(B)}\right) \quad (2.3a)$$

Therefore,

$$\begin{aligned}\hat{T}_{y,opt} &\approx \sum_{s_B} y_{kB}d_{kB}[1+(n_B/n)^{-1}(\hat{\lambda}_{yB}/n)(y_{kB}-\overline{y}_B)] \\ &= \sum_{s_B} y_{kB}d_{kB}a_{kB}=\sum_{s_B} y_{kB}w_{kB}\end{aligned} \quad (2.3b)$$

Equations (2.2b) and (2.3b) imply that the initial weights d_{kA} for s_A and d_{kB} for s_B are calibrated to w_{kA} and w_{kB}, respectively, such that estimates from each sample are identical and equal to $\hat{T}_{y,opt}$. In other words, after calibration, the difference between the two estimates becomes zero, the new control total.

The above idea of zero control totals for new auxiliary variables defined by the key study variables (henceforth denoted by z), such that differences between the two estimates for the overlapping frame are constrained to be zero, is central to the proposed calibration approach. In our case, the two estimators are dependent because of common samples, and so an appropriate modification of the combination parameter α_{opt} taking account of the covariance needs to be made.

For complex designs, it generally is difficult to write the optimal linear combination in a calibration form. For this reason, Singh and Wu (1996, 2003) used an extension of GREG (generalized regression) estimator to propose a suboptimal composite estimator in the form of a calibration estimator. Forms of the adjustment factors are given by

$$\begin{aligned}a_{kA}&=1+\eta_A^{-1}(\mathbf{x}'_A\lambda_A+\mathbf{z}'_A\lambda_z)\ , \\ a_{kB}&=1+\eta_B^{-1}(\mathbf{x}'_B\lambda_B-\mathbf{z}'_B\lambda_z)\end{aligned} \quad (2.4)$$

where, $\mathbf{x}$ denotes the usual auxiliary covariates with known control totals T_x and $\mathbf{z}$ is the set of key study variables. It may be of interest to note that the above adjustment factors can be obtained by minimizing the following distance function subject to

$$\Delta(w,d)=\eta_A\sum\nolimits_{s_A} d_{kA}(a_{kA}-1)^2+\eta_B\sum\nolimits_{s_B} d_{kB}(a_{kB}-1)^2 \quad (2.5)$$

Note that z appears with different signs in the above two adjustment factors because control totals for z are zero. However, the λ_z parameters are common to both factors. The parameters (or scaling factors) η_A and η_B are pre-specified and signify relative sample sizes depending on the design of each sample. They are analogous to the relative sample sizes in the SRS case. In practice, the η-parameters can be determined via a grid search such that variance of $\hat{T}_{z,comp}$ (or trace of the covariance matrix if z is multivariate) is minimized. Thus the choice of η_A and η_B will automatically reflect the dependence, if any, between the two samples.

In the next section, we discuss our use of the generalized exponential model (GEM), instead of GREG, as it has a built-in control for extreme weights for suitable outlier domains and gives adjustment factors in pre-specified ranges. Also, GEM provides a unified approach for both nonresponse and post-stratification adjustments.

3. Proposed DFC Estimator

3.1 Definition of DFC. For the two overlapping samples of respondents, s_A (Phase 1), and s_B (Phase 1 and Phase 2) defined in Section 2, let $s_{A\|B}$ denote the concatenated sample. Our goal is to develop the following composite estimator of a population total:

$$\hat{T}_{y(A\|B)} = \zeta_A \sum_{s_A} w_{kA}\, y_{kA} + (1-\zeta_A)\sum_{s_B} w_{kB}\, y_{kB} \qquad (3.1)$$

where,

y_{kA} = outcome of interest for k^{th} respondent of s_A
y_{kB} = outcome of interest for k^{th} respondent of s_B
w_{kA} = calibrated weight for the k^{th} respondent in s_A
w_{kB} = calibrated weight for the k^{th} respondent in s_B.

Additionally, we require that the usual controls on $\boldsymbol{x}$ and the new controls of zero on $\boldsymbol{z}$ are satisfied.

The combining factor is given by

$$\zeta_A = \left(\sum_{s_B} w_{kB}^2 - \sum_{s_A} w_{kA} w_{kB}\right)\left(\sum_{s_A} w_{kA}^2 + \sum_{s_B} w_{kB}^2 - 2\sum_{s_A} w_{kA} w_{kB}\right)^{-1} \qquad (3.2a)$$

The parameter ζ_A reflects the dependence between s_A and s_B. The parameter is chosen to minimize the variance when the variance is approximated by a constant times the UWE formula $(1+CV^2(w))$ under a simple super population common mean model (Kish, 1965). As a result, ζ_A is an approximation for a_{opt} given in equation (2.1b). Under this simple superpopulation modeling, the minimum variance of an estimated population mean with optimal ζ_A is proportional to

$$N^{-2}\,\frac{\sum_{s_A} w_{kA}^2 \sum_{s_B} w_{kB}^2 - \left(\sum_{s_A} w_{kA} w_{kB}\right)^2}{\sum_{s_A} w_{kA}^2 + \sum_{s_B} w_{kB}^2 - 2\sum_{s_A} w_{kA} w_{kB}} \qquad (3.2b)$$

For the z-variables, the parameter ζ_A has no impact as the two estimates from s_A and s_B are the same. However, the choice of the scaling factor η_A affects the combination which is implicit in the weight calibration (2.5). The choice is made *a priori* or empirically by a grid search over the range $0<\eta_A<1$ such that variance is minimized. On the other hand, for an arbitrary variable y, the two estimates are not the same, and the parameter ζ_A is needed for their combination. The choice (3.2a) for ζ_A is simple and heuristic. Alternatively, for various variables y, one can choose a suitable ζ_A common for all y's such that it minimizes the generalized variance. In variance computation, the factors η_A and ζ_A are treated as pre-specified under the premise that past data have been used for their estimation.

The composite estimator can also be expressed as:

$$\hat{T}_{y(A\|B)} = \sum_{s_B} w_k^*\, y_{kB} \qquad (3.3)$$

where, w_k^* denotes a single set of final calibrated weights for the full sample s_B and is given by

$$w_k^* = \begin{cases} \zeta_A\, w_{kA} + \zeta_B w_{kB}, & if\ k \in s_A \\ \zeta_B w_{kB}, & if\ k \in (s_B - s_A) \end{cases} \qquad (3.4)$$

The calibrated weights in the above formulas are defined as follows.

$$\begin{aligned} w_{kA} &= d_{kA} a_{kA,nr} a_{kA,ps} \\ w_{kB} &= d_{kB} a_{kB,nr} a_{kB,ps} \end{aligned} \qquad (3.5)$$

where, d_{kA} and d_{kB} are the design weights as defined earlier, and the a-factors denote adjustments for nonresponse (nr) and post-stratification (ps). Note that a Hajek-type (1971, in his comments on Basu's paper) ratio adjustment is useful prior to any weight adjustments (for nr or ps) are made since it attenuates the effect of extreme weights as in the elephant fable of Basu . In fact, it gives rise to the desired centering (Singh and Sarndal, 2003) in regression coefficients in the usual GREG for post-stratification so that the

SRS optimal regression estimator (of Section 2) can be obtained as a special case of GREG.

3.2 Use of GEM Calibration for DFC. As mentioned in Section 2, GEM provides a unified approach to both nonresponse and post-stratification adjustments. We use GEM to determine the weight adjustment factors of the following form (given here for the post-stratification factor as an illustration) for pre-specified bound parameters

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

$$a_{kA} = \frac{\ell_{kA}(u_{kA} - c_{kA}) + u_{kA}(c_{kA} - \ell_{kA})\exp_A}{(u_{kA} - c_{kA}) + (c_{kA} - \ell_{kA})\exp_A}$$

$$a_{kB} = \frac{\ell_{kB}(u_{kB} - c_{kB}) + u_{kB}(c_{kB} - \ell_{kB})\exp_B}{(u_{kB} - c_{kB}) + (c_{kB} - \ell_{kB})\exp_B} \qquad (3.6)$$

where,

$$\exp_A = \exp[\eta_A^{-1} A_{kA}(\mathbf{x}'_{kA}\boldsymbol{\lambda}_{xA} + \mathbf{z}'_{kA}\boldsymbol{\lambda}_z)]$$

$$A_{kA} = \frac{m_{kA}(u_{kA} - \ell_{kA})}{(u_{kA} - c_{kA})(c_{kA} - \ell_{kA})}, \quad m_{kA} = b_{kA} / d_{kA},$$

and the corresponding parameters for sample B are defined analogously.

The parameter b_k (here subscripts A and B are suppressed for convenience) denotes the boundaries for defining the extreme weights. Thus extreme weights after calibration satisfy $b_k \ell_k < w_k < b_k u_k$ or the corresponding adjustment factors satisfy $m_k \ell_k < a_k < m_k u_k$, where $a_k = w_k / d_k$.. In other words, the adjusted extreme weights are not necessarily truncated to the boundaries b_k, but stay in its neighborhood depending on the data and the control totals used in the calibration process. For the non-extreme weights, note that $m_k = 1$.

In actual applications, it is clearly not practical to have all different bounds for different units k. In our experience, it is sufficient to have three sets of bounds: (l_1, c, u_1), (l_2, c, u_2) , and (l_3, c, u_3) for high extremes, non-extremes, and low extremes, respectively. The center c is set to 1.0 for post-stratification adjustment and to a number greater than 1.0 (e.g., the overall inverse response propensity) for the nonresponse adjustment. The scaling factors η_A and η_B $(=1-\eta_A)$ appearing in the adjustment factors can be interpreted as parameters that reflect different relative effective sample sizes for the two samples as well as effect of dependence of samples.

Estimates of the λ-parameters in the adjustment factors (3.6) are obtained by simultaneously solving the following equations.

$$\sum_{s_A} \boldsymbol{x}_{kA}\, d_{kA}\, a_{kA} = T_{\boldsymbol{x}}$$

$$\sum_{s_B} \boldsymbol{x}_{kB}\, d_{kB}\, a_{kB} = T_{\boldsymbol{x}}$$

$$\sum_{s_A} z_{kA} d_{kA} a_{kA} - \sum_{s_B} z_{kB} d_{kB} a_{kB} = 0 \qquad (3.7)$$

where, $T_{\boldsymbol{x}}$ is the vector of usual post-stratification totals for $\boldsymbol{x}$. RTI's GEM macro uses Newton-Raphson to solve these equations.

The adjustment factors (3.6) can be obtained directly by minimizing a distance function using Lagrange multipliers under the calibration constraints (3.5) given below.

$$\Delta(w,d) = \eta_A \Delta_A(w_A, d_A) + \eta_B \Delta_B(w_B, d_B),$$

$$\Delta_A(w_A, d_A) =$$

$$\sum_{s_A} \frac{d_{kA}}{A_{kA}} \left[(a_{kA} - l_{kA}) \log \frac{a_{kA} - l_{kA}}{c_{kA} - l_{kA}} + (u_{kA} - a_{kA}) \log \frac{u_{kA} - a_{kA}}{u_{kA} - c_{kA}} \right] \qquad (3.8)$$

and Δ_B is similarly defined.

While GEM is a generalization of Deville and Sarndal's (1992) Logit method to allow for unit-specific bounds, the dual-frame GEM is a further extension of GEM to encompass zero controls defined by (3.7).

3.3 DFC Steps. The proposed DFC method can be summarized in the following steps:

1. Define two samples of selected units: Sample A (s_A* = all units selected for the main survey), and Sample B (s_B* = Phase 1 respondents plus all units selected for the follow-up). Also, set values for the η-parameters as well as bounds for the nonresponse and post-stratification adjustment factors.
2. Perform a Hajek-ratio adjustment to the design weights for each sample so that each set of weights adds up to the specified population totals.
3. Perform a nonresponse adjustment via GEM to the weights in Step 2 for the set of respondents s_A in s_A* and s_B in s_B*; the adjustments factors are defined as $a_{kA,nr}$ and $a_{kB,nr}$, respectively. Note that the sum of the adjusted weights for both sets

of respondents equals the total population count specified in Step 2 due to the Hajek adjustment.

4. Implement GEM post-stratification on the concatenated sample $s_{A\|B}$ with the usual demographic controls and the new calibration controls of zero (3.7); the adjustments factors are defined as $a_{kA,ps}$ and $a_{kB,ps}$ for the respective samples.
5. Choose η_A such that the objective function defined by the variance or trace of the covariance matrix (see next section) is minimized. Calculate the final calibrated weights for Sample A (w_{kA}) and Sample B (w_{kB}). Use the formula in 3.2a (or generalized variance minimization) to calculate the parameter ζ_A and hence, w_k* as defined earlier (3.4).

4. Variance Estimation

We will consider the cases of simple and complex designs separately.

4.1 Simple Designs with Nonrandom Adjustment Factors. If the design is a (stratified) SRS at both phases and if the GEM calibration adjustment factors ($a_{kA,nr}$ and $a_{kA,ps}$ for Sample A and $a_{kB,nr}$ and $a_{kB,ps}$ for Sample B) are treated as nonrandom, then we can use the estimate developed by Rao (1973), see e.g., Lohr (1999). In this paper, we first review the usual estimator for double sampling for stratification.

The following formula for variance estimator assumes no nonresponse in the follow-up survey ($s_B^* = s_B$) and could be modified via Taylor linearization in the case of nonresponse. The estimator is written as

$$\hat{T}_{y(B)} = \frac{N}{n_A^*}\sum_{s_A} y_k + \frac{N}{n_A^*}\frac{1}{\nu}\sum_{s_M} y_k \qquad (4.1)$$

where, ν is defined by $n_B = n_A + \nu\left(n_A^* - n_A\right)$, η_A* is the size of s_A*,and s_M is the size of the follow-up subsample.

Then, ignoring the finite population correction, we have

$$\hat{V}\left(\hat{T}_{y(B)}\right) = N^2\left[\frac{n_A - 1}{n_A^* - 1}\frac{s^2_{y(A)}}{n_A^*} + \frac{n_B - 1}{n_A^* - 1}\frac{s^2_{y(M)}}{\nu n_A^*}\right] + N^2\left[\frac{1}{n_A^* - 1}\left\{\frac{n_A}{n_A^*}\left(\bar{y}_A - \bar{y}\right)^2 + \frac{n_M}{n_A^*}\left(\bar{y}_M - \bar{y}\right)^2\right\}\right] \qquad (4.2)$$

where,

$$\bar{y}_A = \frac{1}{n_A}\sum_{s_A} y_k,\ \bar{y}_M = \frac{1}{n_M}\sum_{s_M} y_k,\ n_M = \nu\left(n_A^* - n_A\right)$$

and
$$\bar{y} = \frac{n_A}{n_A^*}\bar{y}_A + \frac{n_M}{n_A^*}\bar{y}_M$$

We can express $\hat{T}_{y(A\|B)}$ in the above form (4.1), and then use the formula (4.2) for variance estimation. Here, the adjustment factors $a_{k,nr}$ and $a_{k,nr}$ are subsumed in y_k since they are treated as nonrandom.

4.2 Simple Designs with Random Adjustment Factors. This is the more realistic case. To properly account for random adjustment factors, we can use Taylor-linearization (see e.g., Singh and Folsom, 2000; Binder, 1996; Binder, et al., 2000) to obtain an approximately unbiased estimator of variance. We will assume that the parameters η_A, η_B, ζ_A and ζ_B are given *a priori*, and therefore can be treated as nonrandom. This is a reasonable assumption if these parameters are estimated from historical data. Now for simplicity in illustration, we will consider only the post-stratification adjustment factors as random. For $\hat{T}_{z(A\|B)}$, the linearized estimator containing the Taylor deviations as residuals is given below. It can be done in two ways, one based on sample A and the other based sample B. We have

$$\hat{T}_{z(A\|B)} = \sum_{s_A} z_{kA}w_{kA} = \sum_{s_A} z_{kA}d_{kA}a_{kA,nr}a_{kA,ps} \approx \sum_{s_{A\|B}} \Delta_k + const \equiv \hat{T}^{linA}_{z(A\|B)} + const \qquad (4.3)$$

where,

$$\Delta_k = \delta_{kA}w_{kA}z_{kA} - A_k' H^{-1}\sum_{s_A} w_{kA}B_k z_{kA}$$
$$A_k' = (\delta_{kA}w_{kA}x_{kA}',\quad \delta_{kB}w_{kB}x_{kB}',\quad \delta_{kA}w_{kA}z_{kA}' - \delta_{kB}w_{kB}z_{kB}')$$
$$B_k' = (\delta_{kA}a^{-1}_{kA,ps}\phi_{kA,ps}x_{kA}',\quad \delta_{kB}a^{-1}_{kB,ps}\phi_{kB,ps}x_{kB}',\quad \delta_{kA}a^{-1}_{kA,ps}\phi_{kA,ps}z_{kA}' - \delta_{kB}a^{-1}_{kB,ps}\phi_{kB,ps}z_{kB}') \qquad (4.4a)$$

Additionally, δ_{kA} is one if the unit is in s_A, and zero otherwise; δ_{kB} is similarly defined. The variable $\varphi_{kA,ps}$ is defined as

$$\phi_{kA,ps} = \eta_A^{-1}m_{kA}\frac{(u_{kA,ps} - a_{kA,ps})(a_{kA,ps} - l_{kA,ps})}{(u_{kA,ps} - c_{kA,ps})(c_{kA,ps} - l_{kA,ps})} \qquad (4.4b)$$

and the matrix H is $\sum_{s_{A\|B}} x^*_{k,ps}x'^*_{k,ps}d_k a_{k,nr}\phi_{k,ps}$ where $x'_{kA,ps} = (\delta_{kA}x_{kA}', \delta_{kB}x_{kB}', \delta_{kA}z_{kA}' - \delta_{kB}z_{kB}')$.

Alternatively and equivalently, we can write the linearized estimator in terms of sample B as

$$\hat{T}_{z(A\|B)} = \sum_{s_B} z_{kB} w_{kB} = \sum_{s_B} z_{kB} d_{kB} a_{kB,nr} a_{kB,ps}$$
$$\approx \sum_{s_{A\|B}} \Delta_k + const \equiv \hat{T}^{linB}_{z(A\|B)} + const \qquad (4.5)$$

For an arbitrary y-variable, the linearized expression for $\hat{T}_{y(A\|B)}$ for variance estimation purposes would involve ζ_A as the two samples may not give identical calibrated estimates. It can easily be obtained as

$$\hat{T}_{y(A\|B)} \approx \zeta_A \hat{T}^{linA}_{y(A\|B)} + (1-\zeta_A)\hat{T}^{linB}_{y(A\|B)} \qquad (4.6)$$

which can be re-expressed as a sum of two terms, one involving sample A and the other sample B. If y happens to be one of the z-variables, then we would get the same variance estimate as (4.3) or (4.5) because

$$\zeta_A^2\, Var(\hat{T}^{linA}_{z(A\|B)}) + (1-\zeta_A)^2\, Var(\hat{T}^{linB}_{z(A\|B)})$$
$$+ 2\zeta_A(1-\zeta_A)\, Cov(\hat{T}^{linA}_{z(A\|B)}, \hat{T}^{linB}_{z(A\|B)})$$
$$= [\zeta_A^2 + (1-\zeta_A)^2 + 2\zeta_A(1-\zeta_A)]\, Var(\hat{T}_{z(A\|B)})$$
$$= Var(\hat{T}_{z(A\|B)}) \qquad (4.7)$$

Once the linearized version of the estimator is available, an estimate of the variance can be obtained from the formula (4.2) by replacing the y-variable by the appropriate residuals. Similar formulas can be developed to account for variation due to both nonresponse and post-stratification adjustments. Note that the usual raking-ratio method of weight calibration is a special case of GEM, and the simple post-stratification consisting of ratio adjustments is a special case of raking ratio. Thus the residuals given above represent a generalization of residuals used for linearizing ratio estimates.

We remark that replication methods, such as Jackknife, can be used as an alternative to Taylor linearization (Kott & Stukel, 1997; Fuller, 1998; Kim and Sitter, 2003).

4.3 Complex Designs. For single-phase, multi-stage designs, simple variance estimate formulas are available if the PSUs are treated as being drawn with replacement, and unbiased estimates of the population total from each PSU are available. Under this assumption, simple or stratified SRS-type formulas can be used for variance estimation. This is possible because for usual multi-stage designs, sampling at the second and higher stages satisfy the assumptions of invariance (i.e., higher stage selection probabilities do not depend on the outcome of the first stage), and independence (i.e., selection of SSUs is independent from PSU to PSU), see e.g., Sarndal, Swensson, and Wretman (1992). While these are sufficient conditions, they can be relaxed. The main requirements are that (1) conditional on the first-stage sample, higher stage selections are independent across PSUs, and (2) the estimate from each PSU is conditionally unbiased for the PSU total.

Now, for two-phase sampling, typically the assumption of independence of the conditional SSU selection within PSUs is violated. However, as in our case of double sampling for stratification, if the stratification respects PSU boundaries, and second-phase sampling is designed to be nested within PSUs (i.e., the design treats PSUs as substrata), then the simplified single-phase variance estimation method would be applicable under the usual assumption of with replacement PSU selection. Note that in the case of surveys with nonresponse follow-up, the follow-up sample units are selected independently within each PSU. If the first-phase sampling does not involve PSUs, then it may be reasonable to construct pseudo-PSUs for the sake of simplified variance estimation. To account for the random calibration adjustment factors, the simplified formula based on the with-replacement PSU assumption can be applied to the linearized estimator of Sections 4.1 and 4.2.

5. 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 over 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 (using the AAPOR RR3 definition) 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 respondents) 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' 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.

We achieved a 55.5 percent overall response rate (AAPOR RR3 definition) among the 1,000 mail nonrespondents selected for telephone follow-up. 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. A total of 5,709 eligible sample members responded to either the mail survey or the telephone follow-up. The overall weighted response rate (a.k.a. the *effective response rate)* among eligible sample members for the combined mail survey and telephone follow-up was 70.5 percent with a 95 percent confidence interval of +/- 3.3 percent. The weighted response rate can be thought of as an estimate of a population parameter. That is, the parameter is the response rate that would be achieved if everyone on the sampling frame had been selected for the survey.

We applied the DFC methodology detailed in the previous sections to compute the GWHS analysis weights where $s_A{}^*$ = 10,301 veterans initially selected for the survey; and $s_B{}^*$ = 5,182 mail respondents plus 1,000 mail nonrespondents selected for follow-up. Note that $s_B{}^*$ is a proper subset of $s_A{}^*$. The design weights are such that $\Sigma_{sA^*}\ d_{kA} = \Sigma_{sB^*}\ d_{kB} = 685{,}074$ veterans on the sampling frame. The two over-lapping samples of respondents are s_A containing 5,182 respondents to the mail survey; and s_B containing the 5,182 respondents to the mail survey plus 527 follow-up respondents. Note that s_A is a proper subset of s_B.

To estimate the variances of survey outcomes in a design-consistent fashion, we created 294 variance replicates (a.k.a. random groups) that enabled us to combine data obtained from the mail survey and the telephone follow-up. Within each of 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. The primary advantage of random groups is that standard survey software packages (e.g., SUDAAN®) can be used to analyze the data. In fact, the variance estimates that we obtained for outcomes for mail respondents (i.e., excluding the follow-up) using random groups are only slightly conservative compared to the usual variances obtained for (single-phase) stratified designs. For two-phase variance estimation, we calculated 294 sets of replicate weights for use with the 'delete one' Jackknife method of variance estimation (Lohr 1999, p 298). We constructed each set of replicate weights by serially deleting one replicate from the sample and then adjusting the DFC weights to account for the deleted replicate.

The post-stratification control totals T_x corresponded to the following 17 counts:

- First-stage strata (8): Gender x Component x CCEP evaluation
- Branch of Service (4): Army, Navy and Coast Guard, Marine Corps, Air Force

- Race/ethnicity (3): White, Black, Other
- Military rank group (2): Officer, Enlisted.

Next, we used GEM separately for each sample to calculate post-stratification adjustment factors $a_{kA,ps}$ and $a_{kB,ps}$ that were applied to the nonresponse-adjusted design weights to force them to sum to the 17 control totals. It may be useful to perform this post-stratification separately for each sample to reduce coverage bias before doing the final DFC for improving efficiency. These weights are used as input weights to the DFC procedure. The following ten survey outcomes comprise the z-vector of zero controls:

- CDC Multi-Symptom Illness (2 levels)
- Post-traumatic Stress Disorder (2 levels)
- Chronic Fatigue Indicator (2 levels)
- SF 36 Impairment Score (continuous)
- Hopkins Symptom Depression (continuous)
- Chalder 13-Item Fatigue Score (2 levels)
- Partner has Discomfort during Sex (3 levels)
- Current Smoking status (3 levels)
- Current Drinking status (5 levels)
- Current Marital Status (2 levels)

To obtain each of the 294 DFC replicate weights, we used GEM to calculate DFC adjustment factors $a_{kA,DFC}$ and $a_{kB,DFC}$ that were applied to the nonresponse-adjusted and post-stratified design weights so that the differences between the ten key outcomes using w_{KA} and w_{kB} were zero while maintaining the 17 control totals. Using a grid search, we determined that a scaling constant of η_A = 0.80 minimized the sum of the variances of the key outcome variables. We used the scaling constant to calculate the DFC weights w_{KA} and w_{kB} and then a combining factor of ζ_A = 0.82 to calculate w^*_k, the set of final DFC weights for the full respondent sample s_B.

5.1 Unequal Weighting Effects. The combination of a one-in-five sub-sampling rate for the telephone follow-up and a 55 percent response rate to the follow-up resulted in analysis weights for follow-up respondents that were approximately ten times as large as those for mail respondents. As a result, the reduction in bias obtained from the follow-up was adversely affected by the increase in sampling variance that resulted from the increased variability in the sampling weights of the combined sample.

In **Table 1**, we show that before DFC the effective sample size actually decreases significantly from 1,672 to 535 when the follow-up respondents are included in the analysis. In other words, the variances associated with estimates based on the overall sample are larger than those based only on the mail portion of the survey. After DFC, the effective sample sizes of the overall sample exceed those of the mail survey for every major reporting domain.

5.2 Effects of Dual-Frame Calibration (DFC) on Survey Estimates. In **Table 2**, we present survey estimates and corresponding sampling errors before and after DFC calibration. Two sets of survey outcomes are presented. The first set includes the ten key outcomes that comprise the z-vector of zero controls. After DFC, the difference between these estimates for these variables using w_{KA} and w_{kB} is zero. The second set of 'other' outcomes illustrates the effects of the DFC procedure on outcomes that are not explicitly part of the calibration procedure (y variables). For these outcomes, the DFC estimator is the composite of the w_{KA} and w_{kB} estimates using the combining factor of ζ_A = 0.82.

6. Concluding Remarks

The method proposed in this paper provides more efficient estimates when efficiency is addressed at the estimation stage. It was emphasized that this is not substitute for efficiency considerations at the design stage using optimal allocation for double sampling for stratification. Different designs in the main and follow-up samples were allowed in the proposed DFC framework. The GEM calibration method with range restrictions and built-in controls for extreme values provided a convenient tool to produce a final set of calibrated weights for each sample. For those variables not collected in the follow-up survey, estimates are constructed based on only sample A weights. It was observed that simplified variance estimates as in single phase designs can be obtained by nesting the second phase within first phase (pseudo) PSUs.

In this paper, we did not address the issue of possible bias difference in estimates due to different survey modes used in the initial and follow-up surveys as in our application to the GWHS. Since the calibration process makes estimates for a set of key study variables (z) equal for the two samples by using the zero controls, the proposed DFC method eliminates the difference in bias from the two estimates for the selected z-variables. This does not imply that the final composite estimator is free from bias but that it represents a compromise between the biases of the two estimates.

Acknowledgments: The first author's research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada held at Carleton University, Ottawa under an adjunct research professorship. The authors thank Drs. Lori Ebert and Elizabeth Federman of RTI for their help in applying this methodology. The data used in this article is based on a survey conducted by RTI International for the U.S. Army Medical Research and Material Command. Thanks are also due to Pinliang Chen of RTI for developing a DFC version of the GEM macro for weight calibration.

References

Binder, D.A. (1996) "Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach." *Survey Methodology*, 22, pp. 17-22.

Binder, D.A., Babyak, C., Brodeur, M., Hidiroglou, M., and Jocelyn, W. (2000) Variance estimation for two-phase stratified sampling. *The Canadian Journal of Statistics*, Vol. 28, No. 3.

Cochran, W.G. (1977) *Sampling Techniques, Second Edition.* John Wiley & Sons, New York.

Deville, J.-C., and Sarndal, C.-E. (1992). Calibration estimation in survey sampling. JASA, 87, 376-382

Fay, R.E. (1991) A design-based perspective on missing data variance. In Proceedings of the Annual Research Conference, US Bureau of the Census, 429-440

Folsom, R.E. Jr. 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.

Fuller, W.A. (1998) Replication variance estimation for two-phase samples, Statistica Sinica, 8, 1153-1164

Hajek, J. (1971). Comments on the paper 'An essay on the logical foundations of survey sampling, Part One', in *Foundations of Statistical Inference,* eds. V.P. Godambe and D.A. Sprott. Holt, Rinehart, and Winston of Canada. Ltd, Toronto.

Hansen, M. H., and Hurvitz, W.N. (1946). The problem of nonresponse in surveys. Journal of the American Statistical Association, 41, 517-529

Kim, J.K., and Sitter, R.R. (2003). Efficient replication variance estimation for two-phase sampling, Statistica Sinica, 13, 641-653

Kish, L. (1965) *Survey Sampling*, John Wiley & Sons, New York.

Kott, P.S., and D.M. Stukel (1997) "Can the Jackknife Be Used With a Two-Phase Sample." *Survey Methodology*, Vol. 23, No. 2, pp. 81-89.

Lohr, S.L. (1999) *Sampling: Design and Analysis*, Duxbury Press, New York.

Potter, F. (1990) "A Study of Procedures to Identify and Trim Extreme Sampling Weights." *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 225-230.

Rao, J.N.K. (1973) On double sampling for stratification and analytical surveys. Biometrika, 60:125-133, 669.

Sarndal, C.E., Swensson, and Wretman. (1992). *Model-assisted survey sampling*. Springer-Verlag.

Shao, J., and Steel, P. (1999) Variance estimation for survey data with composite imputation and Non-negligible sampling fractions. JASA, 94, 254-265

Singh, A.C. and Sarndal, C.-E. (2003) Optimal Regression, Generalized Regression, and Modified Regression, manuscript under preparation.

Singh, A.C. and Folsom, R.E. Jr. (2000) Bias Corrected Estimating Function approach for calibration adjusted variance estimation. Proc. Amer. Stat. Assoc, Surv. Res. Meth. Sec., 610-615

Singh, A.C. and S. Wu (2003). An extension of generalized regression estimator for dual-frame surveys. Proceedings of the *American Statistical Association, Section on Survey Research Methods*, in print.

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

Table 1. Comparison of Unequal Weighting Effects (UWE) and Effective Sample Sizes (Eff. n) Before and After Dual-Frame Calibration (DFC)

Domain	Respondents: Mail Only	Respondents: Mail & Follow-Up	Before DFC, Mail-Only: UWE	Before DFC, Mail-Only: Eff. n	Before DFC, Mail & FUs: UWE	Before DFC, Mail & FUs: Eff. n	After DFC, Mail-Only: UWE	After DFC, Mail-Only: Eff. n	After DFC, Mail & FU: UWE	After DFC, Mail & FU: Eff. n
Overall	5,182	5,709	3.10	1,672	10.67	535	3.11	1,666	3.09	1,850
Active Duty	3,214	3,566	2.62	1,227	9.01	396	2.63	1,223	2.62	1,362
Reserves	1,968	2,143	2.32	848	7.50	286	2.33	845	2.33	919
Males	3,382	3,735	2.31	1,464	7.98	468	2.32	1,458	2.31	1,620
Females	1,800	1,974	1.76	1,021	6.51	303	1.77	1,014	1.79	1,100
Active Duty, Males	2,100	2,339	1.91	1,101	6.58	355	1.91	1,097	1.91	1,223
Active Duty, Females	1,114	1,227	1.72	649	6.60	186	1.73	645	1.75	700
Reserve, Males	1,282	1,396	1.93	666	6.26	223	1.93	664	1.94	721
Reserve, Females	686	747	1.79	383	5.81	129	1.80	381	1.82	411

Table 2. Effects of the DFC Procedure on Survey Outcomes

Survey Outcomes	Before DFC, Sample A: Mean	Before DFC, Sample A: SE	Before DFC, Sample B: Mean	Before DFC, Sample B: SE	Diff	After DFC, Sample A w_{kA}: Mean	After DFC, Sample A w_{kA}: SE	After DFC, Sample B w_{kB}: Mean	After DFC, Sample B w_{kB}: SE	DFC w^*_k: Mean	DFC w^*_k: SE
Key Outcomes (z-variables):											
% w/Multi-symptom Illness	68	1.16	63.6	1.91	4.5 §	67.3	1.05	67.3	1.05	67.3	1.05
% w/PTSD Indicator	7.6	0.63	8.5	1.17	-0.9	7.8	0.59	7.8	0.59	7.8	0.59
Chronic Fatigue Indicator	10.3	0.64	11.7	1.49	-1.3	10.5	0.65	10.5	0.65	10.5	0.65
Chalder Fatigue Score	60.5	1.26	57.9	2.05	2.6	60.1	1.16	60.1	1.16	60.1	1.16
% w/Sexual Discomfort	10.1	0.77	9.5	1.44	0.5	10.1	0.75	10.1	0.75	10.1	0.75
% Current Smoker	24.3	1.16	28.6	1.93	-4.3 §	25.0	1.05	25.0	1.05	25.0	1.05
% Drinking (Mult X/Week)	31.3	1.20	32.1	2.05	-0.8	31.4	1.10	31.4	1.10	31.4	1.10
% Married/Cohabiting	76.4	0.96	73.8	1.94	2.6 §	76.1	0.92	76.1	0.92	76.1	0.92
Depression Subscale	1.7	0.02	1.7	0.03	0.0	1.7	0.01	1.7	0.01	1.7	0.01
Physical Impairment Score	7.0	0.04	7.0	0.07	0.1	7.0	0.04	7.0	0.04	7.0	0.04
Other Outcomes:											
% w/General Health =Excel	8.8	0.63	11.7	1.53	-3.0	8.8	0.60	11.3	1.50	9.2	0.61
% w/General Health =VGood	29.4	1.14	27.1	1.74	2.3	29.5	1.11	26.8	1.77	29.0	1.05
% w/General Health =Good	38.5	1.21	39.2	2.31	-0.7	38.3	1.20	40.7	2.18	38.7	1.18
% w/General Health =Fair	20.0	1.13	18.4	1.69	1.5	20.0	1.09	17.9	1.50	19.7	1.05
% w/General Health =Poor	3.3	0.42	3.5	1.07	-0.2	3.4	0.44	3.3	0.96	3.4	0.42
% w/Reduced Time at Work	16.5	0.92	18.6	1.70	-2.1	16.7	0.96	17.2	1.19	16.8	0.93
% w/Limit Type of Activities	22.9	1.14	24.5	2.09	-1.6 §	23.1	1.11	23.0	1.37	23.1	1.10
% w/Loss of Interest in ADLs	57.7	1.27	61.5	1.93	-3.8 §	57.7	1.19	61.3	1.73	58.4	1.18

§ Difference significant at the 0.05 level.