Journal of Official Statistics, Vol. 30, No. 3, 2014, pp. 521–532, http://dx.doi.org/10.2478/JOS-2014-0032

Developing Calibration Weights and Standard-Error Estimates for a Survey of Drug-Related Emergency-Department Visits

Phillip S. Kott[1] *and C. Daniel Day*[2]

This article describes a two-step calibration-weighting scheme for a stratified simple random sample of hospital emergency departments. The first step adjusts for unit nonresponse. The second increases the statistical efficiency of most estimators of interest. Both use a measure of emergency-department size and other useful auxiliary variables contained in the sampling frame. Although many survey variables are roughly a linear function of the measure of size, response is better modeled as a function of the log of that measure. Consequently the log of size is a calibration variable in the nonresponse-adjustment step, while the measure of size itself is a calibration variable in the second calibration step. Nonlinear calibration procedures are employed in both steps. We show with 2010 DAWN data that estimating variances as if a one-step calibration weighting routine had been used when there were in fact two steps can, after appropriately adjusting the finite-population correct in some sense, produce standard-error estimates that tend to be slightly conservative.

Key words: Frame variable; response model; prediction model; general exponential model; finite population correction.

1. Introduction

The Drug Abuse Warning Network or DAWN (Substance Abuse and Mental Health Services Administration 2012) was a national stratified random sample of US hospitals used to estimate annual drug-related emergency-department visits and related statistics. This article describes a calibration-weighting strategy for the DAWN that was never implemented because the survey was discontinued after 2012. Nevertheless, we feel this strategy and our contemplated approach to variance/mean squared error estimation contained some innovative features worth sharing.

The DAWN sample was drawn from a list frame provided by the American Hospital Association (AHA). The frame was stratified by location, size, and ownership type (public *vs.* private). Hospitals were oversampled within 13 metropolitan areas, for which domain estimates were published when respondent sample sizes were deemed large enough.

[1] RTI International, 6110 Executive Blvd., Rockville, MD 20852, U.S.A. Email: pkott@rti.org
[2] Substance Abuse and Mental Health Services Administration, 1 Choke Cherry Road, Rockville MD 20857, U.S.A. Email: charles.day@samhsa.hhs.gov
Acknowledgments: The views expressed in this article do not necessary reflect those of the Substance Abuse and Mental Health Services Administration. The authors express their gratitude to an associate editor and several referees whose comments improved the presentation considerably.

In the estimation strategy used operationally for DAWN, the weight for a respondent began with the hospital's design weight. A nonresponse adjustment factor was applied to each weight to account for those hospitals that were sampled but did not participate in the DAWN survey. This was followed by a sample balancing – often called a "poststratification" adjustment – to improve the efficiency (reduce the variances) of most of the resulting nearly (i.e., asymptotically) unbiased estimates. Both steps employed simple weighting-class adjustments requiring *ad hoc* collapsing schemes when there were too few respondents in a class or the class adjustment factor was deemed too large.

In this article, we will describe alternative approaches to these two adjustments. For simplicity, we will ignore the subsampling of visits and visit-level nonresponse adjustments that took place within some DAWN hospitals.

The new nonresponse adjustment factors use a calibration-weighting routine that implicitly models the probability that a hospital responds to (participates in) the DAWN survey. It does this by assuming hospital response is a function of its characteristics, such as its size, measured by annual emergency-department (ED) visits on the AHA frame, ownership (public or private), region, and the population density of the county in which it is located. If the response model is correctly specified, as we assume it is, then employing this calibration-weighting routine produces nearly unbiased estimates of DAWN totals.

The new sample-balancing adjustment factors are produced using a version of nearly pseudo-optimal calibration (Kott 2011) that forces each final weight to no less than 1. Sample balancing exploits the fact that the variables measured by the DAWN survey, such as annual drug-related ED visits, are functions of characteristics known for all hospitals on the AHA frame. Calibrating the respondents' weights so that the estimated totals of (some of) those characteristics computed from the respondent sample exactly equal corresponding frame (AHA) totals tends to increase the efficiency of estimated DAWN totals, which remain nearly unbiased.

Evaluation of the nonresponse pattern in DAWN data from 2010 lead us to treating the hospitals from the 13 metropolitan areas as one subpopulation and the remaining hospitals as a separate subpopulation. For brevity's sake we restrict our attention in this article to nonresponse modeling and weight adjustments for the former subpopulation. Similar methods can be used for the subpopulation of remaining hospitals. The impact of finite-population correction on variance estimation is much less of an issue in that subpopulation.

Although the DAWN published domain estimates for many of the 13 metropolitan areas, we investigated domain estimates within the four US census regions instead. This kept the respondent sample sizes within domains more respectable given that much of the theory underpinning calibration weighting is asymptotic.

Since many DAWN hospitals were sampled with certainty (before nonresponse), we restrict our attention in this article to linearization-based variance estimators of nearly unbiased estimated totals that require finite population correction. Most software designed to estimate variances using linearization-based methods only capture the *increase* in variance from the respondent sample size being smaller than the before-nonresponse sample size and from the final weights being more variable than the original weights. We will describe linearization-based methods that also capture the *decrease* in variance resulting from hitting calibration targets as well as from finite population correction.

The software package SUDAAN 11® (RTI 2012) can produce linearization-based measures that estimate variances appropriately when there is a single step of calibration weighting, but not (easily) when there are multiple calibration steps. We will discuss a simplified variance estimator for the DAWN given our two-step calibration scheme that can be implemented in SUDAAN 11. The resulting estimated variances tend to be slightly conservative when applied to DAWN data from 2010.

Calibration weighting for the DAWN is discussed in Section 2. Section 3 addresses variance estimation after calibration weighting. Section 4 contrasts alternative variance estimators using DAWN data, while Section 5 offers some concluding remarks.

2. Calibration Weighting for the DAWN

2.1. Nonresponse Adjustment

Let d_k be the design weight for a sampled DAWN hospital k. For our purposes, this was the population size of the stratum (say h) containing k divided by its sample size (N_h/n_h). The strata within a metropolitan area were determined by size class (up to three within an area) and ownership type.

Following Folsom (1991), our nonresponse-adjusted weight for a DAWN respondent k has the form:

$$a_k = d_k\left[1 + \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\right], \tag{1}$$

where $\mathbf{x}_k$ is a vector of the respondent's characteristics to be described shortly, and $\mathbf{g}$ is determined using Newton's method (successive linear approximation) so that the calibration equation

$$\sum_R a_j\mathbf{x}_j = \sum_R d_j\left[1 + \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\right]\mathbf{x}_j = \sum_S d_j\mathbf{x}_j \tag{2}$$

holds where R is the respondent sample and S the sample before nonresponse.

The value

$$p_k = p(\mathbf{g}^T\mathbf{x}_k) = 1/\left[1 + \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\right]$$

implicitly estimates the probability that k is a respondent given its characteristics in vector $\mathbf{x}_k$.

Although $p(\mathbf{g}^T\mathbf{x}_k)$ is a logistic function of $\mathbf{g}^T\mathbf{x}_k$, this method in not the same as finding $\mathbf{g}$ using either maximum likelihood (i.e., so that $\sum_S\left\{\left[1 + \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\right]^{-1} - I_j\right\}\mathbf{x}_j = \mathbf{0}$, where $I_j = 1$ if $j \in R$ and 0 otherwise) or quasi-maximum (i.e., so that $\sum_S d_j\left\{\left[1 + \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\right]^{-1} - I_j\right\}\mathbf{x}_j = \mathbf{0}$). Kim and Riddles (2012) show why the calibration approach in Equation (2) can lead to estimated totals with smaller variances than maximum-likelihood-based alternatives.

Preliminary analyses of 2010 DAWN data strongly suggested that the probability of response was better modeled as the log of the AHA emergency-department visits than as a direct function of ED visits. This is a more sensible result than it may appear to be. It means that a one percent increase in the size measure lead to an r percent increase in the odds of response, all other things being equal.

After extensive model searching, we ultimately assumed unit response to be a logistic model of an $\mathbf{x}_k$ vector containing the log of the number of AHA emergency-department visits, which we denote $\log(q_k)$, dummy variables for each of the 13 metropolitan areas, $d_{1k}, \ldots, d_{13k}$, an indicator for a public (as opposed to private) hospital, d_{Pk}, an interaction term between the public indicators and one of the area dummies $d_{Pk}\, d_{13k}$, and the log of the population density within the ZIP code containing the hospital (from US Census Bureau 2012) with imputation of missing values when needed, t_k. Note that q_k must always be positive, which it was, so that $\log(q_k)$ can be defined.

Although we assume we know the correct form of the model governing the response probabilities for each hospital, $\rho_k = p(\boldsymbol{\gamma}^T \mathbf{x}_k) = 1/[1 + \exp(\boldsymbol{\gamma}^T \mathbf{x}_k)]$, we can only estimate the parameter $\boldsymbol{\gamma}$ with $\mathbf{g}$ in Equation (2). We further assume that whether or not a hospital k responds given $\mathbf{x}_k$ is independent of whether another hospital responds.

2.2. Sample Balancing

Like most government surveys, the DAWN produces a number of estimates. It is possible that a weight adjustment will decrease the variances of some estimates while increasing those of others. Nevertheless, we chose to focus our sample-balancing efforts on reducing the variance of a single estimate: the total number of drug-related emergency-department visits. This can be viewed as the "flagship" variable of the DAWN survey. Not only is it important in its own right, but it is related to many of the DAWN survey variables.

Using the nonresponse-adjusted weights from the previous step (the a_k), ignoring strata (and thus the need to collapse strata with only a single responding hospital) but otherwise using a routine sensitive to the sampling design, we fit linear models of drug-related emergency-department visits, y_k, using covariates available on the AHA frame.

The model we liked best effectively modeled not y_k but y_k/q_k as a function of four census-region dummies, $u_{1k}, \ldots, u_{4k}$, $\log(q_k)$, and $u_{1k}d_{Pk}$ through $u_{4k}d_{Pk}$. Observe that y_k/q_k is the ratio of the number of drug-related emergency-department visits to a proxy of all emergency-department visits (using a previous year's data). The final model fit y_k as a linear function of $q_k u_{1k}, \ldots, q_k u_{4k}$, $q_k \log(q_k)$, and $q_k u_{1k} d_{Pk}$ through $q_k u_{4k} d_{Pk}$.

Following the advice in Kott (2011), we set final calibration weights at

$$w_k = a_k \frac{\ell_k(u_k - 1) + u_k(1 - \ell_k)\exp(B_k[a_k - 1]\mathbf{h}^T \mathbf{z}_k)}{(u_k - 1) + (1 - \ell_k)\exp(B_k[a_k - 1]\mathbf{h}^T \mathbf{z}_k)}, \tag{3}$$

where $\mathbf{z}_k = (q_k u_{1k}, \ldots, q_k u_{4k}, q_k \log(q_k), q_k u_{1k} d_{Pk}, \ldots, q_k u_{4k} d_{Pk})^T$, $B_k = (u_k - \ell_k)/[(1 - \ell_k)(u_k - 1)]$, $\ell_k = 1/a_k$, and $\mathbf{h}$ is found so that the calibration equation, $\sum_R w_j \mathbf{z}_j = \sum_U \mathbf{z}_j$, holds.

The fraction on the right-hand side of Equation (3) is a particular version of the general exponential model of Folsom and Singh (2000):

$$f(\mathbf{h}^T \boldsymbol{\delta}_k; u_k, c_k, \ell_k) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp(A_k \mathbf{h}^T \boldsymbol{\delta}_k)}{(u_k - c_k) + (c_k - \ell_k)\exp(A_k \mathbf{h}^T \boldsymbol{\delta}_k)}. \tag{4}$$

This version is centered at 1 (all c_k are 1) with all $A_k = B_k$. With some work, one can see that the right-hand side of Equation (4) is nearly equal to $1 + \mathbf{h}^T \boldsymbol{\delta}_k$ when $\mathbf{h}^T \boldsymbol{\delta}_k$ is small, which it should be assuming we have already appropriately adjusted for nonresponse (and there are no frame coverage issues). By setting $\ell_k = 1/a_k$, no weight can be less than 1. Finally, letting $\boldsymbol{\delta}_k = [a_k - 1]\mathbf{z}_k$ will tend to produce more efficient estimates than the conventional setting $\boldsymbol{\delta}_k = \mathbf{z}_k$.

If no restriction is put on the upper size of the weight adjustment in Equation (3), that is, if all $u_k = \infty$, then

$$w_k = 1 + (a_k - 1)\exp(B_k[a_k - 1]\mathbf{h}^T\mathbf{z}_k).$$

The third census region has only 32 respondents. Without restricting the u_k some of those have relatively large $w_k q_k$ values. This suggested to us setting u_k in this region to $.105Q/q_k$, where Q was the sum or the q_j in the region. This restricts the size of $w_k q_k = a_k u_k q_k$ to 10.5% of Q. We chose 10.5% because a restriction to 10% was not possible without the calibration equations failing to hold.

3. Variance Estimation

Both the weight-adjustment functions, whether a_k/d_k in Equation (1) or w_k/a_k in Equation (3), are versions of Folsom and Singh's general exponential model:

$$f(\phi; u_k, c_k, \ell_k) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp(A_k\phi)}{(u_k - c_k) + (c_k - \ell_k)\exp(A_k\phi)}$$

where $A_k = (u_k - \ell_k)/[(c_k - \ell_k)(u_k - c_k)]$. For variance estimation under a correctly specified response model, one needs the derivative of $f(.)$ with respect to ϕ, which is

$$f'(\phi; u_k, c_k, \ell_k) = \frac{(u_k - f_{1k})(f_{1k} - \ell_k)}{(u_k - c_k)(c_k - \ell_k)} \tag{5}$$

where $f_{1k} = f(\phi; u_k, c_k, \ell_k)$.

3.1. *One Calibration-Weighting Step*

If we only calibrated for nonresponse, a good estimator for the variance of $t_{y,a} = \Sigma_R a_k y_k$, assuming the response model is correctly specified, would be

$$v(t_{y,a}) = \sum_{h=1}^{H}\sum_{k\in S_h}\left(1 - \frac{n_h}{N_h}\right)\left(\frac{n_h}{n_h - 1}\right) \times \left[\left(\theta d_k \mathbf{x}_k^T\mathbf{b}_1 + a_k e_{1k}\right) - \frac{1}{n_h}\sum_{j\in S_h}\left(\theta d_j \mathbf{x}_j^T\mathbf{b}_1 + a_j e_{1j}\right)\right]^2 + \sum_{k\in R} d_k\left(f_{1k}^2 - f_{1k}\right)e_{1k}^2, \tag{6}$$

where $a_k = 0$ when hospital k is not in the set of responding hospitals R, S_h denotes a stratum ($h = 1, \ldots, H$) containing n_h sampled hospitals and N_h total hospitals, n is the total number of sampled hospitals (in our case, 367), $f(\mathbf{g}^T\mathbf{x}_k; \infty,2,1) = f_{1k} = a_k/d_k = 1 + \exp(\mathbf{g}^T\mathbf{x}_k)$ is the weight-adjustment factor, $f'(\mathbf{g}^T\mathbf{x}_k; \infty,2,1) = \exp(\mathbf{g}^T\mathbf{x}_k)$,

$$\begin{aligned}\mathbf{b}_1 &= \left[\sum_R d_k f'(\mathbf{g}^T\mathbf{x}_k;\infty,2,1)\mathbf{x}_k\mathbf{x}_k^T\right]^{-1}\sum_R d_k f'(\mathbf{g}^T\mathbf{x}_k;\infty,2,1)\mathbf{x}_k y_k \\ &= \left[\sum_R d_k \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\mathbf{x}_k\mathbf{x}_k^T\right]^{-1}\sum_R d_k \exp\left(\mathbf{g}^T\mathbf{x}_k\right)\mathbf{x}_k y_k,\end{aligned} \tag{7}$$

$$e_{1k} = y_k - \mathbf{x}_k^T\mathbf{b}_1, \text{ and } \theta = 1.$$

Table 1 displays the sample and respondent sizes for our 2010 DAWN data within strata. The certainty strata from across the metropolitan areas have been combined.

See, for example, Kott and Liao (2012) for a fuller explanation of why Equation (6) provides a nearly unbiased estimator for the variance of $t_{y,a}$ when unit response is a logistic function of $\mathbf{x}_k$. The argument there parallels an earlier one in Kott (2006) where instead of the respondent sample being calibrated to the full sample as in Equation (2), the respondent (or full) sample was calibrated to the population using $\sum_R a_j\mathbf{x}_j = \sum_U \mathbf{x}_j$. Equation (6) was proposed in Kott (2006) with $\theta = 0$. The article shows that by injecting $f'(\mathbf{g}^T\mathbf{x}_k; \infty,2,1)$ into $\mathbf{b}_1$, one is able to avoid accounting for the p_k only being estimates of the hospital response probabilities.

Were a simple random sample drawn *with* replacement within the H strata or if the sampling fraction (n_h/N_h) in each stratum were small enough to ignore, a good variance estimator would be

$$v_{WR}(t_{y,a}) = \sum_{h=1}^{H}\sum_{k\in S_h}\left(\frac{n_h}{n_h-1}\right)\left[\left(\theta d_k\mathbf{x}_k^T\mathbf{b}_1 + a_k e_{1k}\right) - \frac{1}{n_h}\sum_{j\in S_h}\left(\theta d_j\mathbf{x}_j^T\mathbf{b}_1 + a_j e_{1j}\right)\right]^2 \tag{8}$$

The added variance due to nonresponse is contained within what looks like a naïve single-phase variance estimator in Equation (8). The added variability due to the response/nonresponse phase comes from the $a_k = d_k f_{1k} I_k = d_k I_k/p_k$, where I_k is the response indicator for hospital k, and p_k remains the hospital's implicitly estimated probability of response. Since the I_k are independent across hospitals, the naïve single-phase variance estimator fully captures the added variance due to nonresponse (for which $\sum_R d_k^2\left(f_{1k}^2 - f_{1k}\right)e_{1k}^2$ would be a good estimator).

3.2. Two Calibration-Weighting Steps

Kott and Liao (2012) also provide a nearly unbiased variance estimator for $t_{y,w} = \Sigma_R w_k y_k$ when unit response is a logistic function of $\mathbf{x}_k$:

$$\begin{aligned} v(t_{y,a}) = &\sum_{h=1}^{H}\sum_{k\in S_h}\left(1-\frac{n_h}{N_h}\right)\left(\frac{n_h}{n_h-1}\right) \\ &\times\left[\left(d_k\mathbf{x}_k^T\tilde{\mathbf{b}}_1 + a_k f_{2k}\tilde{e}_{1k}\right) - \frac{1}{n_h}\sum_{j\in S_h}\left(d_j\mathbf{x}_j^T\tilde{\mathbf{b}}_1 + a_j f_{2j}\tilde{e}_{1j}\right)\right]^2 \\ &+ \sum_{k\in R} d_k\left(\left[f_{1k}f_{2k}\tilde{e}_{1k}\right]^2 - f_{1k}f_{2k}\tilde{e}_{1k}^2\right),\end{aligned} \tag{9}$$

where

$$\tilde{\mathbf{b}}_1 = \left[\sum_R d_k \exp\left(\mathbf{g}^T \mathbf{x}_k\right) \mathbf{x}_k \mathbf{x}_k^T\right]^{-1} \sum_R d_k \exp\left(\mathbf{g}^T \mathbf{x}_k\right) \mathbf{x}_k f_{2k} e_{2k},$$

$$f_{2k} = f([a_k - 1]\mathbf{h}^T \mathbf{z}_k; u_k, 1, 1/a_k),$$

$$e_{2k} = y_k - \mathbf{z}_k^T \mathbf{b}_2,$$

$$\tilde{e}_{1k} = e_{2k} - \mathbf{x}_k^T \tilde{\mathbf{b}}_1,$$

$$\mathbf{b}_2 = \left(\sum_R a_j f'([a_j - 1]\mathbf{h}^T \mathbf{z}_j; u_j, 1, 1/a_j)[a_j - 1]\mathbf{z}_j \mathbf{z}_j^T\right)^{-1}$$

$$\sum_R a_j f'([a_j - 1]\mathbf{h}^T \mathbf{z}_j; u_j, 1, 1/a_j)[a_j - 1]\mathbf{z}_j y_j,$$

and $f'(.)$ is defined using Equation (5). To a large extent, Equation (9) is Equation (6) but with y_k replaced by $f_{2k}\, e_{2k}$ causing $\tilde{\mathbf{b}}_1$and $\tilde{e}_{1k}$ to replace $\mathbf{b}_1$ and e_{1k}. Recall that f_{2k} is very close to 1 under the assumption that we modeled the nonresponse correctly.

Observe that if $\tilde{\mathbf{b}}_1 = \mathbf{0}$, we would have the simplified expression:

$$v(t_{y,a;S}) = \sum_{h=1}^{H} \sum_{k \in S_h} \left(1 - \frac{n_h}{N_h}\right)\left(\frac{n_h}{n_h - 1}\right)\left[a_k f_{2k} e_{2k} - \frac{1}{n_h} \sum_{j \in S_h} a_j f_{2j} e_{2j}\right]^2$$

$$+ \sum_{k \in R} a_k \left(f_{1k}[f_{2k} e_{2k}]^2 - f_{2k} e_{2k}^2\right).$$

This is almost the variance estimator one would get by ignoring the first calibration step and pretending the a_k were the design weights:

$$v(t_{y,a;S'}) = \sum_{h=1}^{H} \sum_{k \in S_h} \left(1 - \frac{n_h}{N_h}\right)\left(\frac{n_h}{n_h - 1}\right)$$

$$\times \left[a_k f_{2k} e_{2k} - \frac{1}{n_h} \sum_{j \in S_h} a_j f_{2j} e_{2j}\right]^2 + \sum_{k \in R} a_k \left([f_{2k} e_{2k}]^2 - f_{2k} e_{2k}^2\right).$$

The difference is the f_{1k}, which appears in $v(t_{y,a;S})$ but not in $v(t_{y,a;S'})$ and makes the former larger than the latter except when all the sampling fractions are ignorably small or there is no nonresponse.

Now suppose instead we assume a linear prediction model consistent with treating $\mathbf{b}_1$ as $\mathbf{0}$. In particular,

$$y_k | \mathbf{z}_k, \mathbf{x}_k = \mathbf{z}_k^T \boldsymbol{\beta}_2 + \varepsilon_{2k}, \tag{10}$$

where the ε_{2k} was uncorrelated random variables each with a mean of zero and a variance of κq_k for some unknown k, whether or not the hospital was sampled or responded when sampled.

It is not hard to see that the model variance of $t_{y,w}$ as an estimator for $\Sigma_U y_k$ given the respondent sample is $\sum_R (w_k^2 - w_k)\kappa q_k$. Similarly, the variance estimator in Equation (6) will have nearly the same prediction-model expectation if the N_h is replaced by

$$N_h{}^* = n_h \frac{\sum_{R_h} a_k{}^2 f_{2k}{}^2 q_k}{\sum_{R_h} a_k f_{2k}{}^2 q_k} \tag{11}$$

when the respondent sample in stratum h is not empty (otherwise, set $N_h{}^*$ to, say, 1000). Since the variance estimator is nearly unbiased given any respondent sample, it is also nearly unbiased on average across all respondent samples, that is, under the combination of the assumed response and prediction models and the original sampling mechanism. Note that when all the stratum sample fractions are ignorably small, this variance estimator coincides with $v(t_{y,a;S})$ (but not generally otherwise).

4. An Application

In this section, we compare variance estimators computed after:

1. Calibrating only for nonresponse pretending the sample was drawn with replacement;
2. Calibrating only for nonresponse;
3. Calibrating for both nonresponse and sample balance but pretending the sample was drawn with replacement;
4. Calibrating for both nonresponse and sample balance;
5. Calibrating for both nonresponse and sample balance but pretending the sample was drawn with replacement and using the simplified version of variance estimation described in the subsection 3.2;
6. Calibrating for both nonresponse and sample balance using the simplified version of variance estimation described in the subsection 3.2.

Since the estimated totals are different when we only calibrate for nonresponse, we compare estimated coefficients of variation (*cv*s) rather than estimated variances. Henceforth, we will abbreviate an estimated coefficient of variation as cv. The fourth variance-estimation method above produced nearly unbiased estimates of the variances for the following six estimated totals we investigated at the US and census-region levels:

all drug-related hospital visits,

alcohol-related visits,

illicit-drug-but-not-alcohol-related visits,

psychotherapeutics-related visits,

stimulant-related visits, and

drug-related visits ending in death

computed within each census region and across the four regions.

We computed some variance estimates pretending the sample was drawn with replacement since that is how many variances are estimated in practice, either because

Table 1. Population, sample, and respondent sizes in subpopulation 1 (13 "metro" areas)

Stratum	Population Size	Sample Size	Respondent Size	N_h^* (Equation (11))
Certainties	254	254	123	683.74
Probability Strata				
East				
Metro Area 1				
Stratum 1	10	8	6	14.34
Stratum 2	10	7	4	15.26
Stratum 3	10	3	2	13.51
Metro Area 2				
Stratum 1	4	2	1	3.96
Stratum 2	8	6	5	11.78
Stratum 3	14	9	4	22.33
South				
Metro Area 3				
Stratum 1	6	5	4	5.65
Stratum 2	44	28	15	108.92
Metro Area 4				
Stratum 1	18	5	4	22.67
Midwest				
Metro Area 6				
Stratum 1	6	3	1	43.34
Stratum 2	7	3	1	31.02
Metro Area 7				
Stratum 1	5	3	0	1000.00
Stratum 2	7	3	2	15.67
Metro Area 8				
Stratum 1	19	4	2	74.72
West				
Metro Area 9				
Stratum 1	6	5	3	8.97
Metro Area 10				
Stratum 1	10	9	3	19.49
Metro Area 11				
Stratum 1	4	3	0	1000.00
Metro Area 13				
Stratum 1	4	3	3	4.09
Stratum 2	5	4	4	5.21
Total	451	367	187	

Metro Areas 5 and 12 have no probability strata (all certainties)

stratum sampling fractions are very small, as they are *not* here, or because the assumption makes variance estimation both easy and conservative. It also lets us see what damage, if any, resulted from our prediction-model-based treatment of finite-population correction.

Both pretending samples were drawn with replacement (WR) and treating them as drawn without replacement (WOR), the relative increase in the *cv*'s from only calibrating for nonresponse are displayed in the first two columns of Table 2. We looked at relative differences in the *cv*s because the different weights from using one or two calibration-weighting steps lead to different estimated totals. We measured relative differences by taking the log of the ratio of the *cv*s being compared (e.g., $\log(cv_A/cv_B)$) because that measure is symmetric.

It is easy to see there is considerable *cv* reduction in most, but not all, cases from the sample balancing in the second calibration-weighting step. The *cv* of the estimates of the

Table 2. Relative increase in estimated coefficients of variation (cv) due to adjusting only for nonresponse or using the simplified variance estimator

Estimator	Adjusting only for nonresponse		Simplified variance estimator	
	WR $\log(cv_1/cv_3)$	WOR $\log(cv_2/cv_4)$	WR $\log(cv_5/cv_3)$	WOR $\log(cv_6/cv_4)$
All regions				
Drug-related visits	48.73	45.21	1.27	7.04
Alcohol-related visits	22.26	17.89	0.59	4.96
Illicit-drug-related visits	19.82	13.77	0.78	7.30
Psychotherapeutics-related visits	21.57	16.57	0.70	6.45
Stimulant-related visits	38.49	34.92	2.31	8.90
Resulted in death	4.93	−8.81	−0.13	7.60
East				
Drug-related visits	76.78	83.14	2.57	8.72
Alcohol-related visits	37.12	35.96	1.41	5.25
Illicit-drug-related visits	48.44	47.85	1.50	10.31
Psychotherapeutics-related visits	44.71	48.04	3.34	5.90
Stimulant-related visits	50.88	56.26	3.05	10.06
Resulted in death	11.79	17.21	0.24	2.33
South				
Drug-related visits	82.93	87.53	0.31	2.46
Alcohol-related visits	26.79	26.00	0.61	2.76
Illicit-drug-related visits	18.46	16.29	0.78	1.61
Psychotherapeutics-related visits	61.01	62.57	−0.05	1.33
Stimulant-related visits	78.39	83.29	0.56	2.97
Resulted in death	23.05	21.03	0.46	0.98
Midwest				
Drug-related visits	118.44	102.45	−0.58	21.37
Alcohol-related visits	106.02	76.14	−0.79	19.74
Illicit-drug-related visits	96.51	70.14	1.28	24.55
Psychotherapeutics-related visits	44.18	29.80	−0.31	17.53
Stimulant-related visits	98.91	84.06	−0.55	20.09
Resulted in death	−14.25	−16.70	0.54	15.32
West				
Drug-related visits	66.50	49.16	0.44	0.02
Alcohol-related visits	49.07	37.72	0.24	10.91
Illicit-drug-related visits	52.15	45.74	−0.05	22.66
Psychotherapeutics-related visits	47.78	36.27	0.41	0.17
Stimulant-related visits	56.66	43.43	0.62	−0.06
Resulted in death	9.52	6.16	−0.43	−2.77
Mean	47.93	42.30	0.70	8.22
Min	−14.25	−16.70	−0.79	−2.77
Max	118.44	102.45	3.34	24.55

number of deaths from drug-related visits both across the US and in the Midwest are larger after sample balancing. All other *cv*s are smaller, over 40% smaller on average.

Columns 3 and 4 show that using the simplified variance estimator described in the last subsection (Equation (6) with the N_h replaced by the N_h* in Table 1) increases the *cv*s more often than not. When it is not conservative, the simplified method is never more than 3% lower than its nearly unbiased counterpart in the 30 *cv*s we computed. The results tend to be more conservative and much more variable when the without-replacement version of the variance estimator is used, and we employ Equation (11) to counteract what would otherwise be an over-correction for the large sampling fractions in most strata. Replacing q_k in Equation (11) by q_k^2 would make the simplified *cv*s a bit less conservative (not shown). The average upward bias would drop to 4.67%, with a minimum of −7.01% and a maximum of 21.13%.

5. Some Concluding Remarks

We have shown how to produce calibration weights for the 2010 DAWN respondent sample of hospitals in two steps – the first to remove the bias from unit nonresponse assuming that we modeled response correctly as a logistic function of covariates, and the second to provide sample balance and thereby increase the statistical efficiency of most estimated totals. We have also shown how to compute nearly unbiased measures of the standard errors of DAWN-estimated totals, providing a simplified version that, although not nearly unbiased, appears to be mostly conservative and is easily computed using SUDAAN 11.

The reason why the simplified version tends to be conservative is that it replaces a respondent-sample derived estimate for a parameter ($\mathbf{b}_1$) by $\mathbf{0}$. To the extent that there are efficiency gains to be made from the nonresponse calibration-weighting step *in addition* to those made in the sample-balancing step − and there may not be any (we are effectively regressing a residual, e_{2k} on $\mathbf{x}_k$, in the nonresponse-adjustment step) − this simplification will tend to underestimate the true standard error of the two-step calibration.

Since we were able to compute a nearly unbiased measure of the standard errors of two-step-calibrated estimates, an obvious question is why bother introducing a simplified version of the computation? The obvious reason is that statisticians will not be able to mimic what we have done for variance estimation without great effort. Moreover, this effort grows for estimated ratios, like the fraction of drug-related hospital visits involving alcohol.

Some may wonder why we did not perform the calibration-weighting steps in the reverse order: sample balancing first, followed by nonresponse adjustment. That clearly could be done, but we will not follow up on it here. Something to consider before reversing the calibration steps, however, is that upper bounds on the final weights cannot be set in the nonresponse-adjustment step unless one is willing to change the form of the response model being fit. This runs the risk of introducing nonresponse bias. No such risk exists when setting upper bounds in the sample-balancing step.

6. References

Folsom, R.E. 1991. "Exponential and Logistic Weight Adjustments for Sampling and Nonresponse Error Reduction." In *Proceedings of the American Statistical Association, Social Statistics Section*, 197–202.

Folsom, R.E. and A.C. Singh. 2000. "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification." In *Proceedings of the American Statistical Association, Survey Research Methods Section*, 598–603. Available at: https://www.amstat.org/sections/srms/Proceedings/papers/2000_099.pdf (accessed July 1, 2014).

Kim, J.K. and M. Riddles. 2012. "Some Theory for Propensity Scoring Adjustment Estimator." *Survey Methodology* 38: 157–165.

Kott, P.S. 2011. "A Nearly Pseudo-optimal Method for Keeping Calibration Weights from Falling Below Unity in the Absence of Nonresponse or Frame Errors." *Pakistan Journal of Statistics* 27: 391–396.

Kott, P.S. 2006. "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors." *Survey Methodology* 32: 133–142.

Kott, P.S. and D. Liao. 2012. "One Step or Two? Calibration Weighting from a Complete List Frame with Nonresponse." Under review by *Survey Methodology* (presented at the Symposium on the Analysis of Survey Data and Small Area Estimation, in honour of the 75th Birthday of Professor J. N. K. Rao).

RTI International 2012. *SUDAAN Language Manual, Release 11.0*. Research Triangle Park, NC: RTI International.

Substance Abuse and Mental Health Services Administration 2012. *Drug Abuse Warning Network (DAWN)*. Available at: http://www.samhsa.gov/data/DAWN.aspx (accessed July 1, 2014).

US Census Bureau 2012. *ZIP Code™ Tabulation Areas (ZCTAs™)*. Available at: https://www.census.gov/geo/reference/zctas.html (accessed July 1, 2014).

Received November 2012
Revised February 2014
Accepted May 2014