Discussion:
Four Presentations (and a Few Thoughts of My Own) on Some Innovative Contributions of Dr. Ralph Folsom

Phillip S. Kott[1]
[1]7413 Tupelo Drive; Derwood, MD 20855

Abstract
I received slides for all four presentations in time. After saying a bit about each, I will offer musings about Ralph Folsom's contributions and add some of my own.

Key words: Calibration equation, response model, outcome model

The Presentations

As usual, **Rod Little** said a lot of sensible things very clearly. But did he really say that missingness is not at random when we know the population total or share of an auxiliary variable but not the individual values for the nonrespondents? Since he and Don Rubin wrote the book on handling missing data (2019), I cannot claim his definition of missingness not at random is wrong only that my use of the term is a bit different.

Gauri Datta sent me 63 slides from a previous conference. He discussed an important problem (ranking small-domain estimates) that is not usually given the attention it needs. Personally, I prefer the approach of Klein, Wright and Wieczorek to his. It requires fewer assumptions. I thank him for telling me about it.

Akhil Vaish organized this session and said a lot of nice things about Ralph and his contributions to small-area estimation. I will discuss another set of Ralph's contributions that are closer to my heart.

A JSM invited session in 2000 that Ralph and Avi Singh organized changed my life (Folsom and Singh, 2000). Before that, I only knew about linear calibration featured in the so-called GREG estimator. It appears from what I heard in a previous session on nonprobability samples, many are still in the dark about Ralph's contributions in this area.

Ralph developed calibration weighting before Deville and Särndal (1992) and double robustness before Bang and Robins (2005). Sadly, because he was a purely design-based survey statistician he didn't fully realize what he had done. Design-based statisticians tried to keep the models they were (implicitly) assuming hidden. What the model-assisted approach of Carl Särndal (e.g., Särndal et al.,1992) and Ken Brewer (e.g., Brewer, 1995) did was to put those models where everyone could see them.

Finally, **Yulei He** proposed some complex but straightforward algebra for measuring the sensitivity of potential missing confounders when blending a nonprobability with a probability sample. I will discuss the missing elephant in the room: Sensitivity when the key missing confounder is the variable of interest itself.

Background to My Approach
(and a fair amount about Ralph's contributions)

Let A be a probability sample, and B and nonprobability sample. Let y_k be the variable of interest available at least in the nonprobability sample, and $\mathbf{x}_k$ a vector of variables including 1 (or the equivalent) available in both samples.

Let w_k be the weight attached to element k of A. That is, the design weight adjusted, if need be, for unit nonresponse.
Let $p(\mathbf{x}_k{}^T\boldsymbol{\gamma}|\ \mathbf{x}_k)$ be an assumed probability of k's inclusion in B. We often think of the functional form of $p(.)$ as being logistic but other assumptions are possible. Whatever $p(.)$ is assumed to be, it is a function of $\mathbf{x}_k{}^T\boldsymbol{\gamma}$ that holds for any value of $\mathbf{x}_k$. Moreover, the assumption about the functional form of $p(.)$ could be wrong.

Under the assumption about $p(.)$ and some mild conditions, $\mathbf{g}$ satisfying the following *calibration equation* is a consistent estimator for $\boldsymbol{\gamma}$:

$$\sum_B [1/p(\mathbf{x}_k^T\mathbf{g})]\mathbf{x}_k = \sum_A w_k\mathbf{x}_k. \quad (1)$$

This is an obvious extension of how Ralph proposed one should adjust for unit nonresponse in Folsom (1991). Observe that the *weights* $1/p(\mathbf{x}_k^T\mathbf{g})$ in the nonprobability sample calibrate the components of $\mathbf{x}_k$ to their estimated population totals. That is not a terminology Ralph used because his work preceded Deville and Särndal's.

Jae Kim (e.g., in Yang et al., 2020) has pointed out that that final $\mathbf{x}_k$ on both sides of equation (1) can be replaced by any function of $\mathbf{x}_k$. Replacing each $\mathbf{x}_k$ by $p(\mathbf{x}_k{}^T\mathbf{g})\mathbf{x}_k$ produces the maximum-likelihood approach for estimating $\boldsymbol{\gamma}$ by $\mathbf{g}$. I prefer "Ralph's" approach for a reason that will be made clear shortly.

Let $m(\mathbf{x}_k{}^T\boldsymbol{\beta}|\mathbf{x}_k, w_k, p_k)$, where $p_k = p(\mathbf{x}_k{}^T\mathbf{g})$, be an assumed outcome model for the expectation of y_k. (The assumption of its functional form also could be wrong.) Estimate $\boldsymbol{\beta}$ in the assumed model by $\mathbf{b}$ in the usual way. When the outcome model is assumed to hold no matter what w_k or p_k is, weighting is unnecessary.

A doubly robust estimator for the population y-mean given the population size N is

$$\bar{y}_{\mathrm{DR}} = \frac{1}{N}\left\{\sum_B [1/p(\mathbf{x}_k^T\mathbf{g})]y_k + \left[\sum_A w_k m(\boldsymbol{x}_k{}^T\boldsymbol{b}) - \sum_B \frac{m(\boldsymbol{x}_k{}^T\boldsymbol{b})}{p(\boldsymbol{x}_k{}^T\boldsymbol{g})}\right]\right\}.$$

If *either* the inclusion model $p(\mathbf{x}_k{}^T\mathbf{g})$ is correctly specified (e.g., as a logistic function) or the outcome model $m(\mathbf{x}_k{}^T\mathbf{b})$ is correctly specified (e.g., as a linear or logistic function), then the estimator is nearly unbiased in some sense. Observe that if $m(.)$ is linear the term in the squared brackets is 0, and $N^{-1}\sum_B y_k/p(\mathbf{x}_k^T\mathbf{g})$, Ralph's estimator for the population y-mean, is doubly robust. Again, Ralph never made that claim because, as a good design-based survey statistician, he never admitted publicly thinking in terms of an outcome model.

There are three reasons to use weights when estimating a population parameter from a survey sample:

The implicit outcome model could be wrong;

The model could be correct, but the errors could be correlated with the weights; or

Your boss (or the client) tells you to.

Here, we are using weights for either the first or second reasons in $\sum_B y_k/p(\mathbf{x}_k^T\mathbf{g})$ and in $\sum_B m(\mathbf{x}_k^T\mathbf{b})/p(\mathbf{x}_k^T\mathbf{g})$. It is not needed in determining $\mathbf{b}$ unless the third reason comes into play.

Aside 1: That the weights are unnecessary in the estimation of the outcome model for near unbiasedness under the response model is something Rod Little (1983) taught me in a different context (the GREG). I mentioned this once to Rod, and he did not remember that point being made in his article.

Aside 2: Many survey statisticians think that calibration weighting means linear calibration. In our notation, $p(\mathbf{x}_k{}^T\mathbf{g})$ has the unlikely functional form $1/(1+\mathbf{x}_k{}^T\mathbf{g})$ under linear calibration. The weight assigned k in the nonprobability sample is $1/\ p(\mathbf{x}_k{}^T\mathbf{g}), = 1+\ \mathbf{x}_k{}^T\mathbf{g}$, which is linear in $\mathbf{x}_k$. That is why many think calibration and modeling the inclusion probabilities are different methods for removing

the bias of a nonprobability sample. If they were aware of Ralph's work, they would see that when done properly, they can be the same thing.

My Approach

Let $\hat{y}_k = m(\mathbf{x}_k{}^T\mathbf{b})$ and create a new row vector $\tilde{\mathbf{x}}_k{}^T = (\mathbf{x}_k{}^T \; \hat{y}_k)$. Consider the new calibration equation,

$$\sum_B 1/p(\tilde{\mathbf{x}}_k{}^T\mathbf{h})\tilde{\mathbf{x}}_k = \sum_A w_k \tilde{\mathbf{x}}_k, \tag{2}$$

under which one can solve for **h**. Now $\bar{y}_{DR2} = N^{-1}\sum_B 1/p(\tilde{\mathbf{x}}_k{}^T\mathbf{h})y_k$ is a doubly robust estimator of the population *y*-mean. Observe that if *m*(.) is linear and the inclusion model holds, then the estimated coefficient on $\hat{y}_k$ would be 0 (or the equivalent; observe that the components of $\tilde{\mathbf{x}}_k$ are not independent when *m*(.) is linear). When *m*(.) is not linear and the inclusion model holds, the estimated coefficient on $\hat{y}_k$ should not be significantly different from 0.

For the sensitivity analysis promised earlier, we can replace $\tilde{\mathbf{x}}_k$ in $p(\tilde{\mathbf{x}}_k{}^T\mathbf{h})$ of equation (2) by $\check{\mathbf{x}}_k = (\mathbf{x}_k{}^T y_k)$ so that the revised calibration equation becomes

$$\sum_B 1/p(\check{\mathbf{x}}_k{}^T\mathbf{h})\tilde{\mathbf{x}}_k = \sum_A w_k \tilde{\mathbf{x}}_k. \tag{3}$$

Observe that we are still calibrated on $\hat{y}_k$ but inclusion is a function of y_k. Solving for **h** (if possible) in equation (3), which may result in a different solution from the **h** that solves equation (2), produces the estimator $\bar{y}_{DR2U} = N^{-1}\sum_B 1/p(\check{\mathbf{x}}_k{}^T\mathbf{h})y_k$. The difference between this estimator and $\bar{y}_{DR2}$ provides an upper bound on the bias realized when element *k* being in the nonprobability sample is a function of $y_k|\mathbf{x}_k$ (under certain reasonable assumptions) rather than $\hat{y}_k|\mathbf{x}_k$. This is a sensitivity measure because we do not know for certain whether *k* being in the nonprobability sample is a function of $y_k|\mathbf{x}_k$ or $\hat{y}_k|\mathbf{x}_k$. It could very well be a combination of the two. Andridge and Little (2011) proposed a proxy pattern-mixture approach for determining the proper estimator in this situation. That approach strikes me as a fancy/shmancy Bayesian justification for computing something like the average of $\bar{y}_{DR2}$ and $\bar{y}_{DR2U}$.

There are several practical problems with my sensitivity-measure approach:

We need a method for computing the margin of error that includes the sensitivity measure. A colleague at RTI, Lance Couzens, is working on a reasonable method.

Near unbiasedness under the outcome model is lost when missingness is a function of y_k.

In most surveys there are more than one *y*-variable of interest.

The proper or close-to-proper choices for the *p*(.) and *m*(.) functions need to be determined as do the components of $\mathbf{x}_k$. The components may include interaction terms and need not be the same for both functions. Yang *et al.* (2020) provides an intriguing method for making those choices.

Finally, if the probability sample collects y_k values, then a better estimator for the population *y*-mean than either $\bar{y}_{DR2}$ or $\bar{y}_{DR2U}$ (or some combination of the two) would incorporate the estimate from the probability sample, $N^{-1}\sum_A w_k y_k$. I will leave how to do that for another time.

References

Andridge, R.R. and Little, R.J.A. (2011). Proxy pattern-mixture analysis for survey Nonresponse, Journal of Official Statistics, 153-180.

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models, *Biometrics*, 61(4),962-973.

Brewer, K. R. W. (1995). Combining design-based and model-based inference. In *Business Survey Methods*, B.G. Cox (chief ed), Wiley, New York, 589-606.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimation in survey sampling, *Journal of the American Statistical Association*, 87, 376-382.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction, *ASA Proceedings of the Social Statistics Section*, 197-202.

Folsom, R. E. and Singh, A. C. (2000). A generalized exponential model for sampling weight calibration for extreme values, non-response and poststratification, *ASA Proceedings of the Survey Research Methods Section*, 598-603.

Kott, P. S. and Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine, *Survey Research Methods*, 105-111.

Little, R. J. (1983). Estimating a finite population mean from unequal probability samples, *Journal of the American Statistical Association*, 596-604.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data* (3rd edition) Wiley & Sons, Hoboken, NJ.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

SAS (2020). *SAS/STAT® 15.2 User's Guide*, SAS Institute Inc, Cary, NC.

Yang, S., Kim J. K., and Song, R. (2020). Doubly robust inference when combining probability and non-probability samples with high dimensional data, *Journal of the Royal Statistics Society, Series B*, 82(2), 445-464.