# WEIGHTING FOR NONTELEPHONE HOUSEHOLDS IN RDD SURVEYS

**David Ferraro and J. Michael Brick, Westat**
**David Ferraro, Westat, 1650 Research Boulevard, Rockville, Maryland 20850**

**Key Words: Coverage bias, response propensity adjustments, calibration weighting, nonignorable missing data**

## 1. Background

Nearly all random digit dial (RDD) telephone surveys have some level of undercoverage because households without telephones are excluded. The size of the undercoverage bias for a particular estimate depends on the relationship between the estimate and telephone status. In some surveys, such as those in which estimates of the low-income population are the focus, the bias may be sufficient to require either special estimation methods or a dual-frame sample that reduces or eliminates the undercoverage.

In this paper, we investigate alternative methods of adjusting the weights of telephone households in RDD surveys to reduce nontelephone household coverage bias. Adjusting weights may be the most appropriate method for general-purpose surveys that collect data on many variables without resorting to surveying nontelephone households. Other methods that might be attractive for special cases are not considered here because such methods tend to place more emphasis on optimality for a particular estimate than applicability for many variables and domains.

The problem of nontelephone coverage bias may be viewed as attempting to define weighting classes that make the missing data ignorable rather than nonignorable, as is often the case. Many characteristics of interest from a survey (such as poverty status, dropping out of school, etc.) are correlated with having a telephone (the source of the missing data). After discussing two methods that have been proposed earlier, three alternative approaches of weight adjustment are considered.

## 2. Standard RDD Weighting Method

The first and most frequently used method of adjusting weights to compensate for nontelephone coverage bias is standard poststratification or raking adjustments to external control totals. Efforts are often made in RDD surveys to identify control variables that are correlated to having a telephone (race and education are two obvious choices). The poststratification step adjusts the weights of the sample from telephone households to sum to the total of all types of households. The poststratification adjustment is often larger or creates more differentials in the weights than all other types of weight adjustments in RDD surveys. It also tends to reduce the bias due to failing to cover nontelephone households.

The main problem with the standard poststratification adjustment is that it does not reduce the bias enough when the survey estimates are highly correlated with nontelephone coverage. For example, despite exceptional efforts in a survey of high school dropouts, the residual bias after poststratification was still large (Brick, Burke and West, 1992). The inadequacy of the bias reduction may be because the data available for forming control totals are not closely related to telephone coverage.

## 3. Keeter Method

Another weighting approach that has been examined is called the Keeter method (Brick, Waksberg, and Keeter, 1996). In this method, data on interruptions in telephone service are collected in the RDD survey. Since telephone households with interruptions are more ‘similar’ to nontelephone households than to telephone households (Keeter, 1995; Frankel, et al. 1999). The weights of the households with interruptions are adjusted to sum to the total of households with interruptions and households without telephones. The standard poststratification adjustment is then applied to further reduce the bias. This approach is somewhat effective in reducing the coverage bias, but it increases the variance for other statistics that were not very correlated with telephone coverage. This is likely to be a feature for any general purpose weighting adjustment.

While the Keeter approach is promising, problems do arise with that method. First, in the Keeter approach some households are adjusted (those with an interruption) and the others are not adjusted. This places the adjustment in a potential small subset of all households. Second, the theory used to motivate the adjustment assumes that every nontelephone household has some probability of having a telephone. Recent evidence from the National Survey of America’s Family (NSAF) suggests that 70 percent of the nontelephone households did not have a telephone at any time during the previous 12 months. Third, the adjustment is uniform for all households (or some small set of groups within the set of households with interruptions) and this may not be desirable. Fourth, the adjustment is totally dependent on one variable, which may not be measured very reliably and does not allow for any customizing for the goals of the survey. Fifth, the telephone interrupt questions must be included in survey itself and a current and reliable estimate of the number of nontelephone households in the country or area must be available.

Given these concerns we examine alternative methods to adjust the weights for nontelephone coverage. In our research, the 1997 and the 1999 NSAF, which collected data from both telephone and nontelephone households, are used to evaluate the procedures. Data from the 1997 and 1999 NSAF are also used to develop some of the adjustments. For details on the 1997 NSAF see Brick et al. (1999). For details on the 1999 NSAF see Brick et al. (2000). We will also use data from other surveys including the National Health Interview Survey (NHIS) and the Current Population Survey (CPS).

## 4. Alternatives to Consider

The first approach we investigate is an extension of the standard poststratification method (Deville, Särndal, and Sautory, 1993). The main innovation is that the development of the poststratification weighting class will be done explicitly for the purpose of adjusting for households without telephone. The first step is to use *logistic regression* to form weighting classes or adjustment cells for poststratification. The propensity to be a telephone household, say $\hat{p}_i$, is predicted from a regression equation for $i$=1,2,…$n_t$ (the sample of telephone households). Grouping cases with similar propensities, $\hat{p}_i$, adjustment cells are formed that are homogeneous with respect to telephone status. The weights are poststratified to known control totals that contain counts for both telephone and nontelephone households. The predictor variables include items such as telephone interruption, poverty status, food stamps,[1] etc. This method is related to what Little (1986) calls response propensity stratification.

The control file we use for the NSAF from 1997 is the 1997 March CPS supplement. We ran the logistic model with CPS data to compute the parameter estimates and propensities. We then applied the CPS parameter estimates to the NSAF data to compute the propensities of the sample data. In our research, we found that households above 200 percent of poverty have a nearly zero propensity to be a nontelephone household and are very difficult to model. Therefore, we created one adjustment cell with households above 200 percent of poverty without additional modeling. We then use logistic regression on the households below 200 percent of poverty to form three additional adjustment cells based on the propensity to be a nontelphone household. The three adjustment cells contain households with low, medium and high propensities. This partitioning greatly improved the predictive power of the models and the estimates. The final weighting classes were added as an additional dimension to the raking procedure of the regular NSAF survey weights.

A limitation of this method is that the survey and the external control file must both contain all of the variables that are used to form the predicted propensities and the variables should be measured consistently. For example, since telephone interruption status is not included in most files used for external control totals (including CPS), this variable could not be included in this modified poststratification approach. The models developed in this manner could be used in other RDD surveys with the same set of predictor variables, provided the other surveys included all the auxiliary items. This issue is not a major concern with the NSAF because the 1997 and 1999 surveys are available and could be used in modeling with or without the telephone interrupt variable. However for other applications it may be limiting.

Little (1986) suggests that creating classes or cells based on the predicted propensities may be preferable to weighting directly by the inverse of the predicted propensity because it places less reliance on the model specification. The propensities are used only to order the sample to create cells, rather than to supply factors to be used directly in the weighting. The weighting class method should improve the variance of the estimates because it eliminates large adjustment factors that could arise if the propensities were used directly.

The second approach we examine is a closely related member of the same set of calibration estimation methods (Lundström and Särndal, 1999). Within this class we examine the use of raking to deal with both the coverage bias and the effect of the adjustments on the variance of the estimates. The procedure is to calibrate the weights using raking to agree with estimates for key variables. More specifically, we calibrate the weights so that the new weights, $w_i^*$, satisfy the equations, $\sum_{i=1}^{n_t} w_i^* y_{ji} = \hat{y}_j$, where $\hat{y}_j$ is the estimated total for variable $j$ from the survey with both telephone and nontelephone households. In our case this reduces to raking the original weights to the $\hat{y}_j$ (this is

[1] Note that we intentionally exclude variables that will be used in the poststratification step that follows the coverage adjustment in the hope that this will allow us to use more variables for the coverage adjustment step.

sample-based raking if the control total number are estimates with sampling errors).

For our research, we used telephone interruption from the 1997 NHIS as an additional dimension in the raking described for the modified poststratification. We computed percentage of households with and without telephone interruption from the NHIS and formed two cells for this additional dimension so that the total conformed to the total for the other raking dimensions.

This second approach is somewhat more flexible than poststratification for two reasons. First, the auxiliary variables do not have to be obtained from the same external control file as is required with poststratification. Thus, the telephone interruption data from the NHIS can be used in addition to the CPS controls with this approach. Second, with raking it is easy to add or delete dimensions if the auxiliary variables are not available from either the survey or the control file without having to do a great deal of additional analysis. This simplicity may make this approach more portable across telephone surveys.

A third approach studied is adding a step to the weighting process to explicitly compensate for nontelephone coverage using an assumed nonignorable model (Little and Rubin, 1987). The model is constructed from the same type of logistic regression analysis described earlier using data files with both telephone and nontelphone households. We developed a relative simple model for the nonignorability. It assumes that within classes the proportion in nontelephone households is a constant. To implement it we computed the ratio $r_i = \dfrac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n_t} w_i}$ for each adjustment class $i$, where $n$ is the number of all persons and $n_t$ is the number of persons in nontelephone households. These classes and ratios constitute the nonignorable model. The ratios can be applied to telephone surveys to adjust for coverage assuming the variables in the model are collected in the survey.

We used the 1999 NSAF to compute the ratio $r_t$ using the same logistic regression and adjustment cells as described for the modified poststratification method. We applied these factors to the 1997 NSAF data as an additional weighting step before the raking procedure of the regular NSAF survey weights.

The advantage of this method is that the model is developed once from a data set with the variables and both telephone and nontelephone households and then is applied to a telephone survey. Current external control files with the auxiliary data files are not needed. The final step is the standard poststratification or raking to the known counts of all persons that is typical of telephone surveys.

## 5. Results

The evaluation of these approaches compares the bias and variance estimates computed using each of the methods. Using the 1997 NSAF data, we produce the following sets of adjusted weights:

1. Regular survey weights including the telephone and nontelephone samples;
2. Standard poststratification of the telephone sample;
3. Keeter adjusted;
4. Modified poststratification of the telephone sample;
5. Calibration (raking) of the telephone sample; and
6. Nonignorable model adjusted telephone sample.

The analysis examines a variety of estimates including estimated totals and proportions for the full sample and domains of interest. The characteristics chosen were mainly those that we thought would have a significant bias if nontelephone households were excluded. The analysis of the first three methods is given in Brick, Flores-Cervantes, Wang, and Hankins, (1999). We compare the results from that study with those found from the 1997 NSAF data using all six methods.

The "standard" estimate is derived from the regular survey weights from the 1997 NSAF with both telephone and nontelephone cases. The evaluation procedure assumes that the standard estimate is unbiased, which in reality is not the case, but we use it as such for comparative purposes for the five alternative estimates. The bias can then be defined as difference between the alternative estimate and the standard estimate. We use two measures of comparison. The first measure is the percent bias given by the ratio of two estimates

$$\text{Percent Bias} = \frac{\text{bias estimate}}{\text{standard estimate}}$$

The second measure is the MSE ratio given by

$$\text{MSE Ratio} = \frac{\begin{pmatrix}\text{unbiased}\\ \text{estimate}\\ \text{of the bias}\end{pmatrix}^2 + \begin{matrix}\text{variance of}\\ \text{alternative}\\ \text{estimate}\end{matrix}}{\text{variance of standard estimate}}$$

Table 1 shows the median of the absolute value of the percent bias and the MSE ratio for each method. Clearly, all the alternative methods reduce the bias and the MSE ratio as compared to the standard poststratification. With a percent bias of 1.44 and a MSE ratio of 1.04 it appears that the modified poststratification method produces the estimates that are closest to the standard estimates.

Table 1. Comparison median values of absolute percent bias and MSE ratios

| | Standard PS | Ketter | Modified PS | Calibration | Non-ignorable |
|---|---|---|---|---|---|
| Percent bias | 4.50 | 2.25 | 1.44 | 1.87 | 2.41 |
| MSE Ratio | 4.05 | 1.30 | 1.04 | 1.10 | 1.38 |

The percent bias is shown graphically in Figure 1 for further comparison of the alternative methods. While there are a few estimates for each method that have large percent biases, most have a small bias. It is also important to consider that the percent bias can be large for rare characteristics, even though the bias itself is relatively small.

The MSE ratios for the alternative methods are shown graphically in Figure 2. Again, some ratios are large but the majority are less than 2. This is especially true for the modified poststratification method. The MSE is dominated by the bias for a number of these estimates when the bias is large. The variances of some of the alternative estimates are actually less than the standard estimates. This is in part due to the NSAF sample design and not necessarily due to the method.

Since the standard estimate is not the actual population value, making comparisons to it may penalize the alternative methods. For example, the standard estimate for an specific statistic may be larger or smaller than the population value. Figure 3 shows the 90 percent confidence intervals around the bias for each estimate for the modified poststratification method. The confidence intervals are similar for the other methods. The confidence intervals for the bias cross zero for the majority of the variables. This highlights the fact that it is important to rely on more than one variable for the evaluation. In this case the bias is not precisely estimated and it may cause us to be overly critical of alternative methods.

We were somewhat surprised that the calibration method did not produce better results. We believe it is at least partly due to the NHIS telephone interruption variable having a lower percentage of cases having an interruption than NSAF. This forces these cases to have a raking factor near 1.0 for this dimension. However, households with telephone interruption also tend to have a high propensity to be a nontelephone household, so most are in the high propensity cell. This means that in the raking adjustment the households with telephone interruption in the NSAF and have a high propensity not to have a telephone have a factor near one. This is not desirable since these same households in the modified poststratification method have the largest factors and are the most representative of the nontelephone population.

## 6. Conclusion

All the alternative methods compared are better than the standard poststratification method for low income estimates. The alternative methods presented in this paper are more attractive to the Keeter method because they do not require dependence on one variable, spread the weights over a broader number of respondents, and do not require reliable estimate of number nontelephone households. The modified poststratification method appears to have a slight advantage over the other methods, but this needs further study.

Further advances maybe possible by improving the modeling. There are also indications that the creation of the adjustment cells may be nearly as important as the modeling. Further research is planned on these topics.

## 7. References


Brick, J.M., Burke, J., and West, W. (1992). *Telephone undercoverage bias of 14- to 21-year-olds and 3- to 5-year-olds* (U.S. Department of Education Technical Report No. 2. NCES 92-101). Washington, DC: U.S. Department of Education.

Brick, J.M., Waksberg, J., and Keeter, S. (1996). Using data on interruptions in telephone service as coverage adjustments, *Survey Methodology*, 22(2), 185-197.

Brick, J.M., Shapiro, G., Flores-Cervantes, I., Ferraro, D., Strickler, T. (1999). *1997 NSAF snapshot survey weights* (Report No. 3). Washington, DC: Urban Institute.

Brick, J.M., Flores-Cervantes, I., Wang, K., and Hankins, T. (1999). Evaluation of the use of data on interruptions in telephone service. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 376-381.

Brick, J.M., Broene, P., Ferraro, D., Hankins, T., and Strickler, T. (2000). *1999 NSAF snapshot survey weights* (Report No. 3). Washington, DC: Urban Institute.

Deville, C., Särndal, C. and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

Frankel, M.R., Srinath, K.P., Battaglia, M.P., Hoaglin, D.C., Wright, R.A., and Smith, P.J. (1999). Reducing nontelephone bias in RDD surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 934-939.

Keeter, S. (1995). Estimating noncoverage bias from a phone survey. *Public Opinion Quarterly*, 59, 196-217.

Lundström, S. and Särndal, C-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, New York: John Wiley & Sons.

**Acknowledgement**

The authors gratefully acknowledge Paul Zador (Westat) for his insight with the logistic regression modeling.

Figure 1. Percent Absolute Bias for Alternative Methods

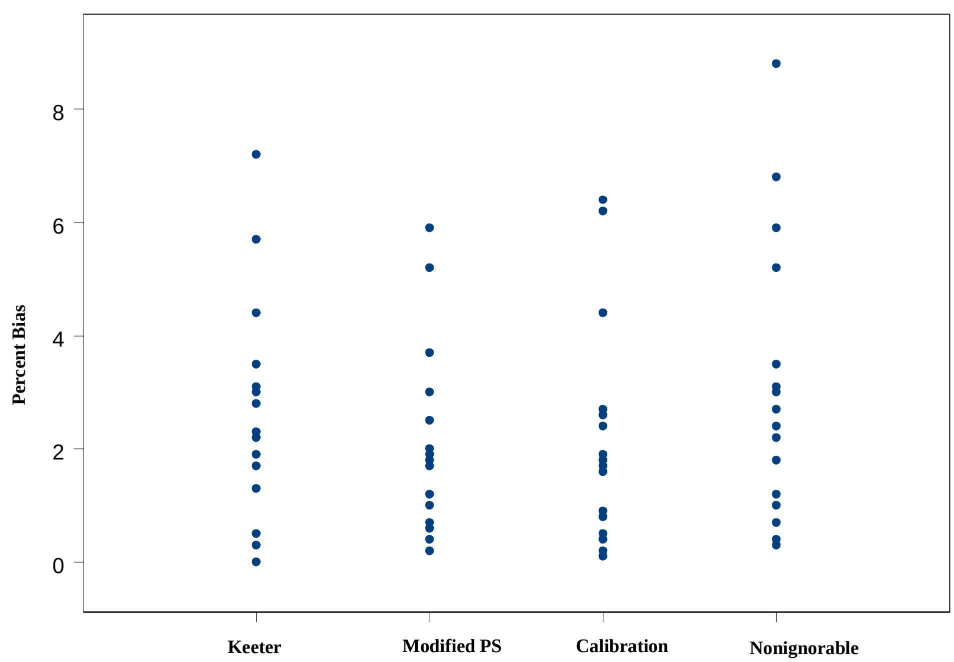

Figure 2. MSE Ratios for Alternative Methods

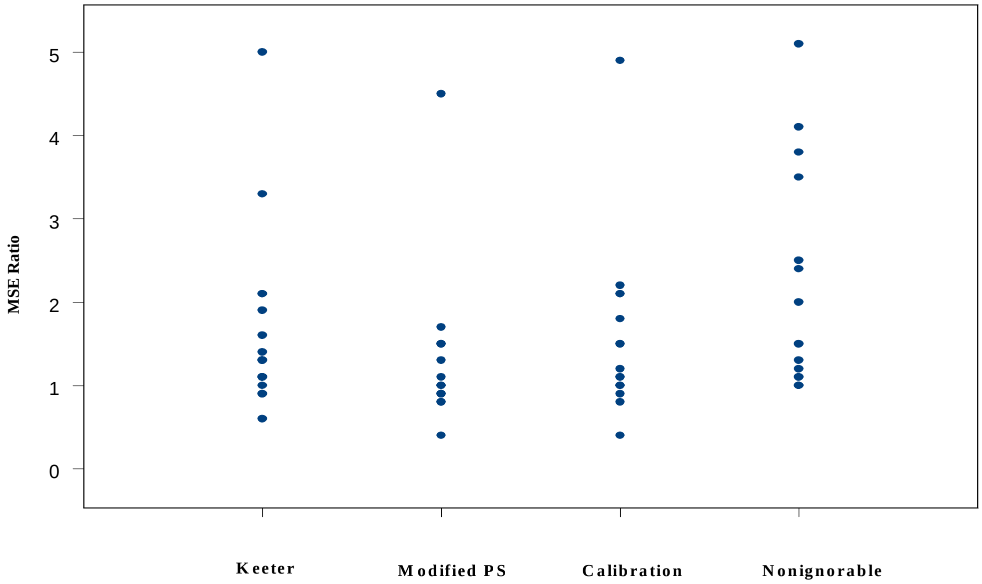


Figure 3. Confidence Intervals Around the Bias for the Modified Poststratification Method