WEIGHTING ISSUES IN AN RDD PANEL SURVEY

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

Key Words: Composite weighting, rotating sample

1. Introduction

The National Survey of America's Families (NSAF) is part of a multi-year study to assess the New Federalism by tracking ongoing social policy reforms and relating policy changes to the status and well-being of children and adults. The major objective of the study is to assess the effects of the devolution of responsibility for major social programs such as Aid to Families with Dependent Children from the federal to the state level. The NSAF is collecting information on the economic, health, and social dimensions of well being of children, non-aged adults, and their families in 13 states that will be intensively studied as part of the project, and in the balance of the nation to permit national estimates. The 13 states were selected to vary in terms of their size and geographic location, the dominant political party and key baseline indicators of well being and fiscal capacity. A sample of the balance of the nation is included so those national estimates can also be produced. Low-income families are oversampled because the policy changes of interest are anticipated to affect them most. The initial round of the NSAF took place in 1997, and a follow-up round was done in 1999. This study is being directed by the Urban Institute and Child Trends and is funded by a consortium of foundations, led by the Annie E. Casey Foundation. Westat is responsible for data collection and related activities.

NSAF is a dual frame survey with both RDD and area components. This paper describes the sample design and decision-making process for the RDD component of the 1999 NSAF. The goal of the design of Round 2 was to produce reliable estimates of both current conditions at Round 2 and change between Round 1 and Round 2 (from 1997 to 1999). After considering the options of an independent or a partially overlapping Round 2 sample, we chose the partially overlapping sample. That is, a subsample of the Round 1 numbers was in the Round 2 sample, along with new numbers not in existence at the time of Round 1. The correlation from overlapping the sample increases the precision of the change estimates. We studied two potential disadvantages to retaining some of the Round 1 sample - the possibility of lowering the response rate in Round 2, and the introduction time-in-sample bias. We estimated these would have little negative effect. Therefore, we decided to retain about 45 percent of the Round 1 sample. Furthermore, we decided to sample the numbers at differential rates that would maximize overlap. More details of the sample design, estimation procedures, and response rates for the 1999 NSAF are available in methodology reports on the Urban Institute web site (newfederalism.urban.org/nsaf/index.htm).

We divided the Round 2 sample into five sampling strata. The subsample of reused Round 1 numbers was stratified into four strata based on their Round 1 disposition (CO-complete, NW-not working/not residential, RB-nonresponse, and NA-never answered). Each stratum was subsampled using differential subsampling rates for the Round 2 sample. The fifth stratum consists of the new RDD phone numbers newly selected for the Round 2 sample. This paper discusses issues in computing the weights for the various strata.

Baseweights are computed differently depending on from which set of banks the phone number came. Both the Round 1 and Round 2 samples were selected from two sets of banks. For Round 1, the first set of banks were those that existed both in Round 1 and Round 2, giving these numbers two chances of selection. This is referred to as the new, old frame and the continuing recycled Round 1 sample. The second set of banks is those that existed only in Round 1 frame and were not in the Round 2 frame. Thus, these numbers did not have a second chance of selection. This is referred to as the non-continuing recycled Round 1 sample. The Round 2 frame can also be thought of as being composed of two sets of numbers. The new, new Round 2 frame was used to sample telephone numbers from telephone banks that existed only in Round 2. These numbers had one chance of selection. The new, old Round 2 frame was used to sample telephone numbers from telephone banks that existed both in the Round 1 and Round 2 frames. These numbers had two chances of being sampled.

Since the continuing recycled Round 1 sample and the new, old sample are selected from the same population, we can improve the precision of the resulting estimates by computing composite weights for them. The paper describes the composite weighting method. We compute an initial baseweight and a composite weight for each stratum. We examine two methods of compositing, with and without a design effect. We also evaluate the loss in efficiency associated with using a compositing factor that is not optimal. This is important because only one factor can be used and each variable could have its own optimal factor. We also examine the loss that might be encountered if a time-in-sample bias exists in the recycled Round 1 sample.

2. Sample Design

This section describes the sample design for the RDD component of the 1999 NSAF, explains the rationale for decisions about the design, and compares expectations at the time of the design to outcomes of the survey.

The goal of the design of Round 2 was to produce reliable estimates of both current level in 1999 and change between Round 1 and Round 2 (from 1997 to 1999). Estimates of current level and change for low-income households were especially targeted. For the RDD component of the Round 2 survey, one option was to select the telephone numbers independent of the Round 1 sample, while a second option was to retain all or some of the Round 1 sampled telephone numbers (plus a supplement for telephone numbers established after 1997).

Selecting an independent sample would provide reliable cross-sectional estimates for 1999, but there would be no correlation between years to reduce the variance of change estimates. This was the main disadvantage of independent sample design. Retaining some or all of the sampled telephone numbers from Round 1 directly addresses this problem, but has its own disadvantages. The two principal issues related to retaining some of the sampled numbers from Round 1 are the potential for lowering the response rate in Round 2 and for introducing time-in-sample bias.

The gain in precision due to correlation from overlapping the sample of telephone numbers and the potential for additional nonresponse and time-in-sample bias were studied during the design phase. Based on correlations observed in other surveys, we assumed that a complete overlap of telephone numbers would lead to an average correlation of 0.36 between rounds (assuming the 100 percent overlap produced a 60 percent overlap of households and the average correlation would be 0.6). At this level, an independent sample would have to be over 2.5 times larger to produce equally precise change estimates. Thus, the overlap had a powerful appeal.

With respect to nonresponse bias, we used data from other longitudinal surveys to project the loss due to asking the same households to respond to the NSAF questionnaire again. The studies suggested that asking the household members to respond a second time might cause a 3 to 7 percent loss in the response rates. This results in about a 3 percent overall loss in response rates for the survey, assuming a 100 percent overlap. We concluded that additional nonresponse bias due to a 3 percent reduction in response rates would not be substantial for the vast majority of estimates.

Interviewing a sample more than once also introduces the possibility of time-in-sample bias (Brooks and Bailar, 1978), that has been studied in several rotating panel surveys. The results of most of these studies, especially for the SIPP (McCormick et al., 1992) which has similar content, showed very little evidence of time-in-sample bias. We concluded that, while the potential for some time-in-sample bias existed, substantively it would have a very small or negligible effect on the estimates.

Based on these evaluations, telephone numbers sampled in Round 1 were retained for Round 2. However, only about 45 percent of the sampled telephone numbers were retained to protect against the potential for nonresponse bias, and to a lesser extent, time-in-sample bias. More detail on these decisions is given by Judkins et al. (forthcoming).

Rather than simply subsampling all telephone numbers at the same rate, the possibility of stratifying the numbers based on their Round 1 status was examined. Five strata were created based on the Round 1 household status: CO-completed, NW–not working, or not residential, NA–never answered, RB-refused, NEW–not sampled in Round 1. The last stratum consisted of telephone numbers that could have been but were not sampled in Round 1 (called the new, old stratum) and telephone numbers newly established after Round 1(the new, new stratum). A cost-variance model was developed and used to determine the optimal subsampling rates in each stratum. The optimal subsampling rates showed that differential subsampling by strata only slightly increased the effective sample size over using a constant rate. Nevertheless, the differential rate design was adopted because it resulted in a much larger overlap of completed interviews and this was deemed to be valuable for estimating change. The last column of Table 1 shows the rate of retaining the Round 1 sample of telephone numbers for Round 2. For example, only 0.222 of the 25,344 telephone numbers that were classified as NA in Round 1 were subsampled for Round 2.

The key design parameters that had to be estimated based on experience of the project staff were the Round 2 residency and response rates. Table 2 shows that in general the estimates used in planning were very close to the observed rates. The difference in response rates at the household screener between the re-used and newly interviewed telephone numbers was about 2 percent, so the effect on nonresponse due to the retention of the numbers from Round 1 was less than 1 percent. This was less than the maximum projected and supports the conclusion that the additional nonresponse bias due to retaining telephone numbers is likely to be small. For a complete set of response rate tables, see Brick et al. (forthcoming).

Table 1. Sample sizes by round and subsampling retention rates, by stratum

Stratum	Round 1 sample size	Round 2 sample size	Subsample retention rate
CO-completed	177,450	118,300	0.667
NW-not working	231,267	77,089	0.333
NA-never answered	25,344	5,632	0.222
RB-refusal	49,199	16,400	0.333
Old Total	483,260	217,421	0.450
NEW-new sample	--	166,229	--

Table 2. Expected and observed residency and screener response rates, by stratum

Stratum	Expected residency rate	Observed residency rate	Expected response rate	Observed response rate
CO-completed	.75	.80	.89	.84
NW-not working	.25	.18	.77	.73
NA-never answered	.15	.14	.77	.59
RB-refusal	.75	.72	.33	.45
NEW-new sample	.48	.42	.77	.78

As a result of the sampling operation, telephone numbers that were in existence in both 1997 and 1999 could be selected at either time. Telephone numbers could not be selected twice because the Round 1 sample was deleted from the Round 2 frame of numbers. The next section describes how composite weights were computed to combine the observations from the samples of telephone numbers sampled from Round 1 and Round 2, given that they had two chances of being sampled.

3. Weights

This section describes the computation of the composite weight for each stratum. As described in the previous section, telephone numbers were selected from the Round 1 or the new stratum. There is an additional complication to this because conceptually the old sample was selected from two sets of banks. The first set of banks were those that existed both in Round 1 and Round 2, so these numbers had a second chance of being selected when the sample was selected in the new, old frame. This will be called the continuing recycled Round 1 sample. The second set of banks was those that existed only in the Round 1 frame and were not in the Round 2 frame. Thus, these numbers did not have a second chance of selection. They will be called the non-continuing recycled Round 1 sample.

Since the non-continuing recycled Round 1 sample was selected from banks that don't exist in the Round 2 frame, we assume that very few are working numbers. Those that do exist are an imperfection in the frame, as the weighted total will sum too more than universe size (in terms of the frame of telephone numbers available at Round 2).

3.1 Baseweights

The baseweights were computed based on the sample's probability of selection, which varied depending on the frame from which it was selected. This section describes each of the possibilities.

The continuing and recycled Round 1 sample had two chances of selection. Their banks existed in both the Round 1 and Round 2 frames. The baseweight is the inverse of the probability of selection in Round 1 times the inverse of the subsampling rate of the Round 1 numbers by stratum in Round 2. The composite weight revises the weight of the continuing recycled Round 1 sample by considering the new, old sample and is discussed in Section 3.2. Note that the noncontinuing recycled Round 1 sample had only once chance of being sampled and their baseweight is simple and no composite is possible.

The numbers in the new, old frame had two chances of selection, in Round 1 (but not selected) and in Round 2. Their banks existed in both the Round 1 and Round 2 frames. The baseweight of the new, old sample is the inverse of the probability of selection in Round 2. The composite weight of the new, old sample is discussed in Section 3.2.

The numbers in the new, new frame had only one chance of selection. Their banks only existed in the Round 2 frame and not in the Round 1 frame. The baseweight of the new, new sample is the inverse of the probability of selection in Round 2.

For a more detailed explanation of how the baseweights were computed, see Brick et al. (forthcoming).

3.2 Composite Weights

Since the continuing recycled Round 1 sample and the new, old sample are selected from the same population, we can improve the precision of the resulting estimates by computing composite weights for them. The composite weight is of the form

$$CW_i = (1-\alpha)BW_i \text{ if } i \in \text{ the continuing recycled Round 1 sample,}$$
$$CW_i = (\alpha)BW_i \text{ if } i \in \text{ the new, old sample.}$$

where BW_i is the baseweight as described above.

Two methods of compositing were studied. The first method computes compositing factors based on the expected number of completed cases from each sample. Those factors are then applied to the baseweights using the formula above. The second method incorporates the expected design effect from each sample before computing the factors. The factors are applied as in the first method. A description of the methods is given below.

Composite Weights Without a Design Effect

In this method the composite factor is computed based on the expected number of completed screeners from the two samples. Note that the base and composite weights were developed and used for several purposes while data collection was being conducted so expected rates were used instead of actual rates. These could be replaced by the final observed rates but such small difference invariably will have negligible effects on the precision of the estimates. For the continuing recycled Round 1 cases, we apply the expected residential and screener response rates from our design assumptions by stratum to produce an expected number of completed screeners.

The total number of expected completed screeners from the continuing recycled Round 1 cases summed across the four strata is S_{old}.

For the new, old sample, we also computed the expected number of completed screeners from this set. To do this we assume the expected completion rate (the product of the residential and screener response rate) for these numbers in Round 2 will be the same as observed in Round 1 and apply this rate to the new, old sample.

The expected completed screeners from the new, old sample are then $S_{new, old}$.

A compositing factor is $\alpha = S_{new, old} / (S_{new, old} + S_{old})$.

Using this compositing factor, the composite weight for the continuing recycled Round 1 sample is $CW = (1-\alpha)BW$. For the new, old sample, the composite weight is $CW = (\alpha)BW$. This assumes estimates from the continuing recycled Round 1 sample have the same expectations as estimates from the new, old sample and both are unbiased.

Composite Weights With a Design Effect

A slight improvement in the compositing can be introduced by incorporating the expected design effect into the computation of the composite weights. To do this, first we compute the expected design effect due to differential subsampling of the Round 1 cases. This was done using the formula given by Kish (1992) and is denoted as DE. Since the differential subsampling is only done in the recycled Round 1 cases, this is applied to those cases only. The new factor to be used in the composite for these cases is $S_{old}^* = S_{old} / DEFF$.

The new composite factor is then $\alpha^* = S_{new, old} / (S_{new, old} + S_{old}^*)$. As before, the composite weights are computed by applying the new factors to the baseweights. That is $CW^* = (1-\alpha^*)BW$ for the continuing recycled Round 1 cases, and $CW^* = (\alpha^*)BW$ for the new, old sample.

In the balance of the nation, the budget was increased between Round 1 and Round 2, so we added additional sample. The budget was decreased in all the other sites. As a result, we computed a different α^* for the balance of the nation (~0.5) and the other sites (~0.3).

The optimal value of the composite factor is α^*, under the assumptions stated. These are the factors used to construct the weights for Round 2.

The proportionate increase in variance from using the composite factor without the design effect, α, rather than the optimal α_0 as presented by Chu et al. (1999) is

$$\text{Loss} = 1 + \frac{(\alpha - \alpha_0)^2}{\alpha_0(1-\alpha_0)}.$$

The loss can be seen graphically in Figure 1. The loss is small whenever α is close to α_0. The loss is greatest at the endpoints when α is much bigger or smaller than α_0.

4. Time in Sample Bias

As noted earlier, there was a concern that recontacting some interviewed households from Round 1 might affect their responses. We call this time-in-sample bias. When one of the estimates involved in the composite is biased then the loss is a function of mean square error rather than variance (it is a proportionate loss in terms of mean square error not variance). The expression given above is also appropriate in this case, as can be easily shown. The only difference is that the loss is now the proportionate loss in mean square error.

Suppose we computed and used α_u, the optimal factor assuming both estimates are unbiased. Now, assume that the estimate from continuing recycled Round 1 sample is biased, due to the time-in-sample effect. It is possible to relate the optimal factor assuming both estimates are unbiased to the optimal allowing for the time-in-sample bias of the old sample. In this case,

$$\alpha_0 = \frac{\alpha_u(\gamma+1)}{1+\gamma\alpha_u},$$

where

$$\gamma = \frac{\text{Bias}^2(Z)}{Var(Z)},$$

The loss can then be re-written as

$$\text{Loss} = 1 + \frac{\alpha_u \gamma^2 (1-\alpha_u)}{\gamma+1},$$

Initial investigations suggest that for NSAF that γ is in the .25 to .50 range with 1.0 being the upper bound. If we let $\alpha = \alpha_u = 0.5$, then the loss due to ignoring the bias is between 1.01 and 1.04 with an upper bound of 1.13. If we let $\alpha = \alpha_u = 0.3$, then the loss is still between 1.01 and 1.04 with an upper bound slightly less at 1.11. Thus, the loss arising from the use of the composite factor that does not allow for time-in-sample bias for the NSAF appears to be very small.

5. Conclusion

The estimation problem encountered in the 1999 NSAF involved utilizing the partial overlap of the sample from the earlier survey to produce current (1999) estimates that have high precision. The method of composite weighting is a natural way of incorporating the information from both sample selections into the survey weights for the 1999 estimates. We examine composite weighting factors assuming the surveys have equal precision and then improve upon those by including the design effect in the overlap sample that is due differential sampling of the previous sample.

The composite weighting factor is established once for all the weighting process and is applied to all estimates. In fact, for this survey the factor was assigned prior to completion of the survey for other reasons. Since no one factor is optimal for all estimates, we examine the consequences of using composite factors that were not optimal. A simple expression of the loss in precision shows that if the factor is in the general vicinity of the optimum, then the loss is small. However, use of arbitrary composite factors that deviate substantially from the optimal factor does cause large losses in precision. Given the information available to develop the factors in the NSAF, we are fairly confident that the factor used is in the neighborhood of the optimal for nearly all estimates. Of course, for small subgroups some loss due to random variation is inevitable.

A related issue is how the potential time-in-sample bias in the estimates from the overlap sample affects the composite factors and the mean square error of the estimates. By assuming that the retained sampled might result in biased estimates we show the optimal factor is a function of the bias. Furthermore, we derive an expression for the proportional loss in the mean square error that results from using a factor that is not optimal. When the bias of the retained sample estimate is small relative to its standard error, then the loss in mean square error is small if the factor is close to the optimal factor assuming both estimates are unbiased. However, substantial losses can occur if the bias is relatively large or the factor is not close to the optimal factor.

6. References

Brick, J.M., Broene, P., Ferraro, D., Hankins, T., and Strickler, T. (forthcoming). *No. 3: 1999 NSAF Sample Estimation Survey Weights*, The Urban Institute, Washington, DC.

Brick, J.M., Broene, P., Cantor, D., Ferraro, D., Hankins, T., Rauch, C., and Strickler, T. (forthcoming). *No. 7: 1999 NSAF Response Rates and Methods Evaluation*, The Urban Institute, Washington, DC.

Brooks, C., and Bailar, B. (1978). *Statistical Policy Working Paper 3, An Error Profile: Employment as Measured in the Current Population Survey*. U. S. Department of Commerce, Office of Federal Statistical Policy and Standards.

Chu, A., Brick, J.M., and Kalton, G.K. (1999). "Weights for Combining Surveys Across Time or Space." Paper presented at the International Statistical Institute Meetings in Helsinki, Finland.

Judkins, D., Brick, J.M., Broene, P., Ferraro, D., and Strickler, T. (forthcoming). *No. 2: 1999 NSAF Sample Design*, The Urban Institute, Washington, DC.

Kish, L. (1992). Weighting for Unequal Pi. *Journal of Official Statistics*, **8**, pp. 183-200.

McCormick, M.K., Butler, D.M., and Singh, R.P. (1992). Investigating Time in Sample Effect for the Survey of Income and Program Participation. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 554-559.

Figure 1