Identifying and Reducing Nonresponse Bias throughout the Survey Process

Thomas Krenzke, Wendy Van de Kerckhove, and Leyla Mohadjer
Westat

Keywords: Weighting, Data Collection, Assessments

1. Introduction

Nonresponse bias can be a major contributing source of total survey error. Too often, this problem is dealt with only in the post-collection approach, through adjustment in the weighting process or through a post-survey evaluation of the potential for bias. The goal of this paper is to suggest the need to place awareness of nonresponse bias throughout the total survey process, that is, to make it part of the initial planning, and, subsequently, to identify, develop, and implement procedures and approaches throughout the survey to minimize nonresponse bias, as well as to plan an evaluation of its potential effects on the data. Specifically, we highlight how to *Plan*, for and how to *Identify* - *Reduce* - and *Evaluate* nonresponse bias.

There are two components of total survey error: variable error (measured through the calculation of variances) and bias. Total survey error is usually measured through the mean square error (MSE). The variance is the first term in the following equation for the MSE of a survey estimate:

$$\text{MSE} = \text{Variance} + \text{Bias}^2 \qquad (1)$$

Bias, the second term in the above equation, contains all sources of error other than variable error. A major component of bias is from nonresponse, that is, the bias due to the failure of some selected persons in the sample to respond to the survey. An estimate for nonresponse bias, assuming that nonresponse is the only source of bias, is expressed in Cochran (1977) as

$$Bias(\bar{y}_R) = (1 - W_R)(\bar{Y}_R - \bar{Y}_N), \qquad (2)$$

where W_R is the response rate and $\bar{Y}_R$ and $\bar{Y}_N$ are the mean values of the survey items estimated among the respondents and nonrespondents, respectively. Thus, the estimates from any survey are subject to bias when some selected persons fail to participate in the survey and when nonrespondents are found to be different from respondents. Nonresponse bias can be substantial when two conditions hold: (1) The response rate is relatively low, and (2) the difference between the characteristics of respondents and nonrespondents is relatively large. Response rates are widely used as a measure of survey quality. But response rates alone are not a good indicator of nonresponse bias. You must also consider the difference between $\bar{Y}_R$ and $\bar{Y}_N$.

Understanding the underlying missing data relationships will help in making decisions about how to reduce nonresponse bias. One common relationship is when missing data are related to known auxiliary information (*X*) such as education attainment, gender, and race. In this case, nonresponse bias can be reduced through weighting procedures using auxiliary variables to form weighting cells.

Another common missing data relationship is when missing data are related to the survey outcome (*Y*). This is called nonignorable nonresponse and is the most difficult to treat (see Stasney (1986), Huisman, et al (1998), Raab, et al (1999) and Cohen and Duffy (2002) for suggested approaches to deal with nonignorable nonresponse). An example of nonignorable nonresponse would result from nonresponse due to language barriers in an English literacy assessment. In such a case, the missing data should not be ignored because the person's English literacy score was most likely below the average respondent's score. Nonresponse bias due to nonignorable nonresponse can be reduced to a certain extent through weighting if auxiliary variables and survey outcomes are correlated; however, nonresponse bias will still remain and cannot be ignored. Other bias reduction procedures besides weighting are necessary to reduce the effects from nonignorable nonresponse.

In this paper, we emphasize a need for an active awareness of nonresponse bias from the start of the survey process, beginning with *plans* to achieve high response rates (Section 2). Next, we provide ways to *identify* the potential for nonresponse bias before and during the data collection period (Section 3). Several approaches to *reduce* nonresponse bias during data collection through focused followup attempts, and after data collection through weighting and estimation techniques are discussed (Section 4). Also in Section 4, an illustration is provided to gauge the magnitude of bias, variance, and MSE reduction from focused followup procedures in data collection, using the 2003 Adult Literacy and Lifeskills (ALL) data. After data collection, there is a need to *evaluate* the remaining high bias domains through some standard evaluation procedures (Section 5).

2. *Plan* to Achieve High Response Rates

Early in the planning process of a survey, and well before data collection, a plan to achieve high response rates should be developed and reviewed. Biemer and Lyberg (2003) devote a chapter on ways to reduce nonresponse bias, including ways to increase response rates before and during data collection. Communicating the plan is essential since the plan will affect different survey staff (operations, field, statisticians, systems). Here are some activities that can be put into action to achieve high response rates.

1. Plan to schedule field activities with respect to special features of the survey. For example, when sampling inmates, the prisons have their own schedules so plans should be flexible in order to work around prison schedules.
2. Test your procedures and your forms to eliminate problems of wording, length, or approaches.
3. Plan community outreach in order to alert the community to data collection in their area. This will help their awareness, interest, and trust as the interviewers approach the selected persons for the interview.
4. Plan to notify the respondents in advance, if possible.
5. Plan to use incentives, that is, pay those who cooperate with the survey, if possible. Within the past twenty years, the use of incentives has been one of the best ways to raise response rates.
6. Plan a 'friendly' questionnaire, that is, make multiple choice answers, clear wording, etc. Mail surveys with difficult skip patterns may increase nonresponse. Other issues affecting response rates are reporting burden and the length of the questionnaire.
7. Limit interviewer effects on nonresponse bias by investing in interviewer training. There is much in the literature on offering incentives; Martin, et al (2001) provides an example as it applies to the Survey of Income Program Participation.
8. Provide interviewers with language capabilities, especially bilingual (e.g., Spanish and English speaking) interviewers in specific areas.
9. Specify the followup approaches. Another way to increase response is to design followup procedures in accordance with traditional current best methods.
10. For domains, specify focused followup approaches in high nonresponse domains. Implement approaches designed to reach nonrespondents and obtain their cooperation. This is discussed further in Sections 3 and 4. Plans can be developed before data collection to use auxiliary data in monitoring data collection and to identify problematic domains in terms of MSE.

3. *Identify* the Potential for Nonresponse Bias Before and During Data Collection

3.1 Before Data Collection

There are some tasks that can be completed prior to data collection that will facilitate the attempts to reduce nonresponse bias throughout the survey process. First, nonresponse patterns from similar surveys can be studied. This will give insights into why people decided not to participate in their study, which may lead to identifying outcome-related reasons for nonresponse. It may also tell you what auxiliary variables are related to the survey outcome, which may lead to making assumptions about nonresponse bias during the field period. For example, in certain low response domains, prior knowledge from similar studies allows survey managers to prioritize field operations depending on the domains' estimated impact on the survey outcome statistics.

Second, reasons for nonresponse that may be related to the survey outcome can be identified. The reason for nonresponse is sometimes as good as the survey data. In the situation mentioned previously, the person did not participate in an English language assessment because he does not speak English. Knowing that a person cannot speak English is about as good as knowing that person's English literacy level and could be used in estimation procedures. Once the reasons are identified, then disposition codes can be created so that the reasons are captured during the data collection attempts. In addition, noninterview reports are completed by interviewers in order to provide other good information about reasons for nonresponse. Interviewers can collect information about the nonrespondents that could be useful for weighting the data or for evaluating the impact of nonresponse bias. Sometimes questionnaire items can be formulated to reduce nonresponse bias for key outcome items. For example, Juster and Smith (1997) used followup brackets on income and assets to reduce the nonignorable nonresponse effects on household economic items.

Third, identify data sources that contain auxiliary data for all sample units that are correlated with survey outcomes. Consider a household sample survey, which collects data in three main stages: the Screener, Background Questionnaire (BQ), and assessment. A rich data source for Screener respondents and nonrespondents is the U.S. Census Summary Files (SFs). The SFs contain tract and block group-level data on income, poverty, household size, education, race/ethnicity, age, gender, and more. This data can be gathered prior to data collection. The Screener data are also a good data source for BQ respondents and nonrespondents and may include data on age, race and gender, for example. A wealth of BQ data may be available for respondents and nonrespondents to the assessment. Once the auxiliary variables are collected, they can be put to good use during data collection.

3.2 During Data Collection

Data collection efforts should be distributed proportionally to the difficulty of reaching sampled units. It is important to closely monitor response rates, move quickly to low response areas, and aggressively counter observed problems.

A useful approach after all traditional data collection approaches have been exhausted, is to focus available resources on follow up attempts in low response domains. Using the auxiliary variables, identify subgroups (pockets) of potential nonresponse bias through the use of segmentation modeling.

Segmentation modeling can be used to split the sample into cells, maximizing the difference in response rates among cells. In Figure 1, the variables *X1* and *X2* are used to split the sample into five cells, two of which are highlighted to indicate low response rates. By focusing followup attempts in the highlighted cells, the respondents in the overall sample will look more like the nonrespondents in terms of the auxiliary variables. If the auxiliary variables are not used in followup attempts, like the box labeled ‘sample’, then more of the same (in terms of auxiliary variables) will be collected. Segmentation modeling and focused followup attempts address missing data related to auxiliary variables, however, weight adjustments can have a similar impact on nonresponse bias. To really have an impact on survey outcomes, it is best to address the missing data related to survey outcomes in the field.

Once the pockets of high nonresponse rates are identified, a decision needs to be made as to whether or not to extend resources into these areas. The size of the domain will factor into the decision, since small domains in the segmentation analysis will not contribute significant bias to the overall survey estimate. It helps to define a minimum cell size that is considered important when doing the segmentation. It is also useful to bring in prior knowledge at this point to determine the importance of the domain with respect to its impact on survey outcomes. Once a decision is made, emphasis needs to be placed on getting cooperation from those that are different from initial respondents such that the difference between $\bar{Y}_R$ and $\bar{Y}_N$ is reduced. The next section addresses this challenging task.

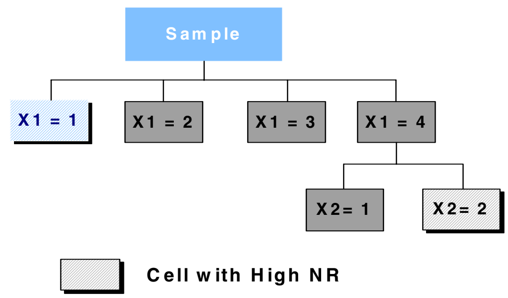

Figure 1. Segmentation modeling illustration

4. *Reduce* Nonresponse Bias During and After Data Collection

4.1 During Data Collection

Within pockets identified as having the most potential to reduce nonresponse bias, it is necessary to implement bias reducing data collection procedures to address nonignorable nonresponse that exists. In doing this approach, we are addressing both components of nonresponse bias: the response rate, and the difference between $\bar{Y}_R$ and $\bar{Y}_N$. Also, we realize that two missing data relationships exist (missing at random and nonignorable nonresponse) and that we can use auxiliary variables and some prior knowledge about survey outcomes during the data collection to identify problematic areas.

Some effective bias reducing procedures were implemented in the 2003 ALL study, conducted by Westat for the National Center for Education Statistics. The ALL sample consisted of an area probability sample of households, where 3,420 persons between the ages of 16 and 65 were tested in English literacy.

Toward the end of the field period, NCES allowed the ALL field period to be extended one month so that focused followup procedures could be continued. We emphasize that the procedures are focused in order to avoid spending large amounts of resources on unfocused followup attempts, only to find that the followup respondents are very similar to the initial respondents, and thus accomplishing only minor reduction in bias.

Some of the most effective focused followup strategies in ALL were called 'blitzes'. That is, the best interviewers were identified and trained in focused followup techniques in areas with high nonresponse, such as New York City. The interviewers were paid incentives for the blitzes. It is believed that the high quality interviewers were better able to convert the initial refusals and reach the tougher cases. The added attention and special strategies relating to gated communities, which were conducted to reach sampled dwelling units behind locked gates also yielded improved results. Another effective approach was to provide mixed modes for notification of followups (flyers, phone calls, letters) since a mode that is good for one, may not be effective for another. The upgraded field operations at the end of data collection also included collecting neighbor information to identify ineligible DUs and using hard-copy screeners. Groves and Couper (1998) provide discussions on interviewer effects, use of mixed modes, a brief discussion on focused followup attempts, and other data collection strategies for household surveys.

We illustrate the effects of the ALL focused followup attempts by comparing the distributions and literacy scores of late respondents with those who had responded during the earlier period during data collection, similar to other analyses such as those conducted by Cohen and Duffy (2002), and Krenzke and Griffin (1997). For this analysis, late respondents were defined to be the last 10 percent of respondents based on the BQ completion date, mostly interviewed during the period of focused followups and upgraded field operations aimed at improving response rates. Other respondents (those who responded during the first 90 percent of the data collection time period) are referred to as early respondents.

Table 1 compares the demographic distributions of the early and late respondents with respect to region. The next step compared mean literacy assessment scores of early and late respondents. Overall, the mean document and quantitative scores of the last 10 percent of respondents were not significantly different from those of the first 90 percent of respondents (although they were still almost 5% different). However, the mean score of the last 10 percent of respondents was approximately 10 percent lower than that of the early respondents in the West region, statistically significant at the $\alpha = 0.10$ level.

Table 1. Differences in mean literacy scores between early and late respondents: 2003 ALL

	Distri-bution of first 90%	Distri-bution of last 10%	Relative difference in mean scores	P-value
Overall	100%	100%	-0.046	0.211
Region				
Northeast	14.8	21.3	-0.011	0.626
Midwest	23.9	15.5	-0.022	0.513
South	33.6	46.8	-0.032	0.694
West	27.7	16.4	-0.102	0.064

The results indicate a small, but potentially important nonresponse bias in literacy scores if the focused followups had not been successful in obtaining responses from late respondents (that is, if data about late respondents were missing). Also, some small level of nonresponse bias may be present to the extent that nonrespondents are similar to late respondents. These numbers were computed prior to weight adjustments.

Focused followup attempts can be done within pockets, resulting from the segmentation analysis, with the most potential for improving the quality of the estimates. We recommend running the analysis several times throughout the field period in order to help plan the focused attempts. The segmentation analysis should be based on as much rich data from the data sources as possible. One point to consider is to combine data collection stages of nonresponse, since it is the total combined impact of nonresponse that is important.

4.2 After Data Collection: Weighting

It is important to reduce the nonresponse bias after data collection is completed to the extent possible. Generally this is accomplished through an adjustment in the weights assigned to respondents. Weight adjustment cells are typically formed for different subgroups of the sample based on auxiliary variables. Prior to forming cells, segmentation modeling can be performed using auxiliary variables and the final response status. The outcome is a data set containing data from respondents whose information has been adjusted to compensate for the data missing for nonrespondents. The underlying assumption for these adjustments is that respondents are fully representative of nonrespondents based on models using auxiliary variables. Since respondents cannot fully represent nonrespondents, some unknown bias remains, even after introducing weighting adjustments for nonresponse or undercoverage. Weighting adjustments will reduce the effects of missing data related to survey outcomes to the extent that auxiliary variables and survey outcomes are correlated. As with the ALL study, some surveys need nonresponse adjustments for multiple stages of data collection. The impact of weight adjustments on bias should be monitored throughout the weighting steps.

4.3 Illustration of Effects on MSE from Focused Followup Attempts

The following illustration, using the 3,420 cases with outcome-literacy data from the 2003 ALL study, demonstrates the statistical effects of a focused followup approach. The impact of using focused followups on standard errors will be less for the clustered ALL sample than for a simple random sample. With regard to bias, the impact of using an area sample for this illustration is not so clear, as it depends upon the missing data relationships. The cases were divided into 24 cells based on correlated variables (X) related to literacy score (Y) with a correlation $\rho(X, Y) = 0.35$. The cells were used in the focused followup attempts as well as weighting adjustments.

The first step was to select 100 initial sets of respondents at a response rate of 75 percent. A nonignorable nonresponse relationship was employed by assuming that the nonresponse indicator (NR) was related to their literacy score. The correlation factor (p) between respondents and nonrespondents and literacy score varied between 0.4 and 0.2.

After selecting the initial set of respondents, the three cells with the highest rates of nonresponse were identified for targeting, which increased the response rate in these three cells by 10 percent using a "get anyone you can get" approach; the rest of the cells remained unchanged. Weighting adjustments were then applied to the initial and final sets of respondents using the 24 cells.

To evaluate the effects of the focused followup, the following were calculated for the initial and final sets of respondents:

- Standard error, as the average of jackknife estimates;
- Bias relative to the standard error of the mean; and
- Mean square error, MSE = Variance + Bias^2.

The results are shown in a series of plots. Each plot contains the above statistics for the overall sample (black dots) and the low response rate cells (light dots) over the 100 iterations. The x-axis of the plot shows the statistic for the initial set of respondents, and the y-axis is for the final set of respondents. The line shows where the two are equal, or in other words, where targeting the low response rate cells had no effect. Points falling below the line indicate a beneficial effect of targeting.

From Figures 2 and 3, targeting appears to have little effect on the standard error, both for the overall sample and the low response rate cells. Figure 4 shows the relative bias for the situation in which response status is highly correlated with the outcome. In this situation, the relative bias is reduced after targeting, particularly for the three targeted cells. However, there is less effect on relative bias when the correlation between response status and the outcome is reduced to 0.2 (Figure 5).

These differences carry through when combining bias and standard error to get MSE. As shown in Figure 6, when there is a strong relationship between response status and the outcome variable, the focused followup was beneficial in reducing MSE, both for the overall sample and for the targeted cells. Increasing the response rate in the three low response rate cells by 10 percent, increased the overall response rate by approximately 1 percent, which was associated with an average decrease in MSE of about 8 percent for the overall sample, and 42 percent for the targeted cells.

However, when the relationship between response status and the outcome variable was not as strong (Figure 7), the benefits were not as great. In this case, the MSE decreased by 7 percent on average for the overall sample, and 22 percent for the targeted cells. In some iterations, the MSE actually increased. Table 2 summarizes the results of the focused followup illustration.

Table 2. Percent change in response rate and MSE before and after targeting

$p(Y, NR)$	Subgroup	Response rate	MSE		
			Min	Mean	Max
0.4	Overall	1.1	-13.7	-8.1	-3.0
	Low RR cells	9.9	-60.8	-42.2	-20.6
0.2	Overall	1.2	-21.1	-6.9	4.6
	Low RR cells	9.9	-55.0	-21.9	34.2

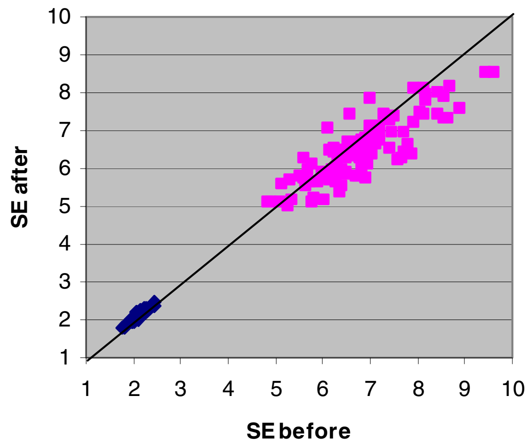

Figure 2. Standard error before and after targeting, $\rho(Y, NR) = 0.4$

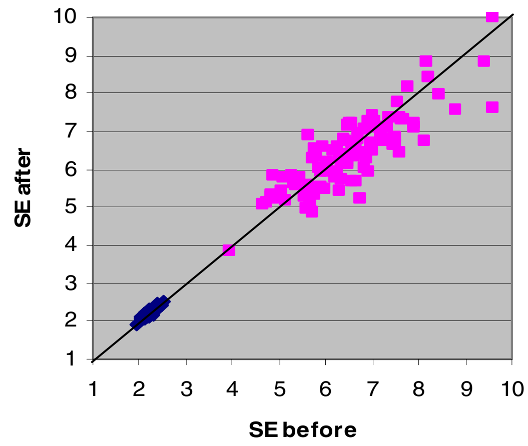

Figure 3. Standard error before and after targeting, $\rho(Y, NR) = 0.2$

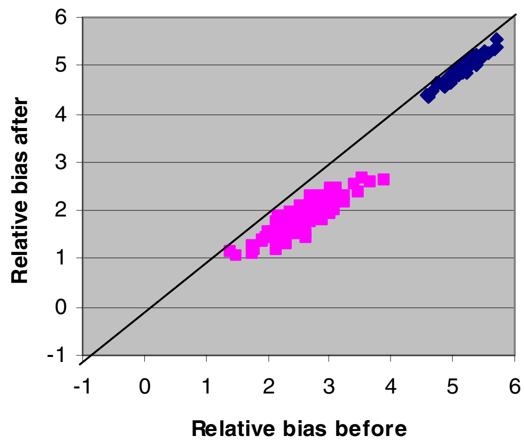

Figure 4. Relative bias before and after targeting, $\rho(Y, NR) = 0.4$

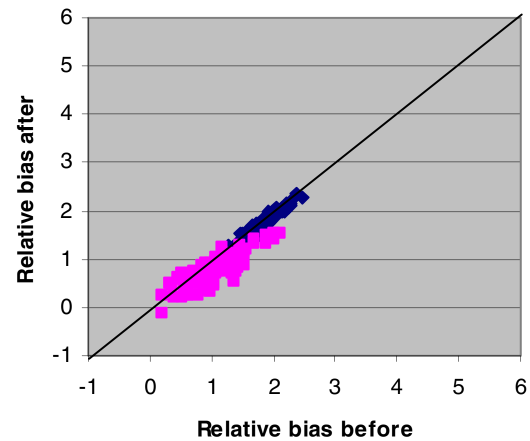

Figure 5. Relative bias before and after targeting, $\rho(Y,NR) = 0.2$

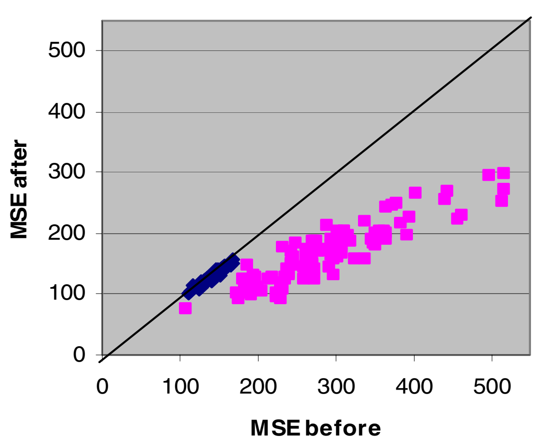

Figure 6. MSE before and after targeting, $\rho(Y, NR) = 0.4$

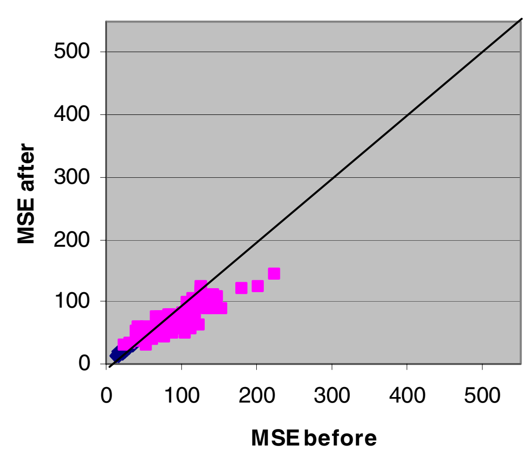

Figure 7. MSE before and after targeting, $\rho(Y, NR) = 0.2$

4.4 Handling Outcome-related Reasons for Nonparticipation

After data collection and if the reasons for nonresponse are known, you may be able to distinguish between outcome-related (e.g., language barriers in a literacy assessment, physical disabilities in a health-related study), and non outcome-related nonresponse. Typical weighting adjustments are not appropriate, since these nonrespondents should not be represented by the respondents. Being able to identify outcome-related reasons for nonparticipation was very useful in the ALL weighting process. In ALL, the outcome-related nonrespondent dwelling units at the screener stage were represented by outcome-related nonrespondent persons at the BQ stage. The outcome-related BQ nonrespondents were then given final weights since they were part of the population of interest. Throughout the weighting process, respondents were not allowed to represent outcome-related nonresponse.

Another way of using outcome-related nonresponse in the estimation stage is by making assumptions from prior knowledge about their outcome, if they had responded. This can be done through the use of estimation models and imputation models. In the ALL survey, literacy-related nonrespondents were imputed with lower than average scores. Where uncertainty exists, a reasonable approach is to exclude such cases and, as needed, to redefine the target population.

5. *Evaluate* the Potential for Nonresponse Bias

After weighting procedures are completed, the last step in monitoring nonresponse bias within the study is to evaluate the remaining potential for nonresponse bias. There are many good examples of a formal nonresponse bias analysis, including the work by Dixon (2002) on the consumer expenditure survey. Most government agencies provide guidelines. For instance, the National Center for Education Statistics offers Statistical Standard 4-1:

"Any survey stage of data collection with a unit or item response rate less than 85 percent must be evaluated for the potential magnitude of nonresponse bias before the data may be released."

As new OMB guidelines are developed, be aware of changes to agency standards and guidelines. Typically, for each stage that does not meet the response rate standard, compare estimated percentages among the respondents to that of the total eligible sample in order to identify any potential bias due to nonresponse. The analysis can use sample base weights and auxiliary variables known for both respondents and nonrespondents that are believed to be related to survey outcomes. A multivariate analysis of the relationship between the nonresponse indicator and survey outcomes, using segmentation analysis or logistic regression, can also be performed to identify the areas with the greatest potential for bias before weighting adjustments. To assess the effects of weighting adjustments on bias, estimates can be compared for auxiliary data before and after each weighting adjustment.

A profile of the sample before and after the weight adjustment for nonresponse is merely the first step in understanding their potential effects on the statistics. This is because weight adjustments are designed to reduce the effects arising from differences in auxiliary data between the respondents and nonrespondents. The weighting process cannot take all differences between respondents and nonrespondents into account. The biases will thus arise from the residual differences.

In the evaluation of nonresponse bias, one should measure the total impact of nonresponse across data collection stages. In addition, measuring the magnitude and significance of directly outcome-related nonresponse (such as language barriers) can prove to be very informative as to the extent of nonignorable nonresponse. Using information from response status codes or noninterview forms, the outcome-related nonresponse relative to non-related nonresponse can be evaluated. Using the Nonresponse Followup Study to the 2003 National Assessment of Adult Literacy (NAAL), the distributions among literacy-related nonresponse were compared to distributions among non-related nonresponse to gain insights into differential prevalence. In addition, one may probe further by evaluating outcome-related nonresponse with regard to important analytic domains. An example is the English as a Second Language Nonparticipants post-hoc analysis of the 2003 NAAL (Van de Kerckhove, et al., 2003). In this analysis, language barrier cases were analyzed by the census, screener, and BQ data to learn more about the characteristics of language barrier cases.

We note that these kinds of analyses can only present rough approximations of the effects of nonresponse biases. Extrapolating from differences in socioeconomic characteristics between respondents and nonrespondents to survey data differences can only be done by assuming that the survey data patterns for nonrespondents are similar to those of respondents, within each socioeconomic category. It is unlikely that this assumption holds exactly, but nonetheless, the evaluation is useful in that it points to where potential biases may exist.

6. Summary

It is important to emphasize awareness of nonresponse bias throughout the survey process. That is, to *Plan*, and then to *Identify – Reduce – Evaluate* nonresponse bias. Segmentation modeling is highlighted as a useful tool during and after the field period.

We also conclude that bias-reducing data collection efforts can be very effective if a strong nonignorable nonresponse mechanism exists AND you can get cooperation from those that are different from initial respondents.

Consideration must be given as to how resources are to be used to improve the quality of the survey estimates. Of course there are always cost-quality tradeoffs. An alternative to focused followup attempts is a more thorough, but costly, followup survey of nonrespondents, such as those described in Triplett, et al (2002) and Huisman et al (1998). How are resources to be used in the most effective manner to increase the quality of the outcome statistics? How far does an agency want to go? Groves (1989) serves as a good reference. Cost models are provided, mostly for followup phone calls. Groves also provides a figure that shows the MSE reduction as a result of followup phone calls.

For the future, research is required to determine the costs in order to determine if focused followup attempts are really cost effective and to explore scenarios of focused followup attempts other than what are provided in our illustration.

Acknowledgement

The authors want to acknowledge Daniel Levine for his thorough review and helpful comments.

References

Biemer, P. and Lyberg, P. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc.

Cochran, W.G. (1977). *Sampling Techniques*. 3rd edition. John Wiley & Sons, New York.

Cohen, G., and Duffy, J.C. (2002), "Are nonrespondents to health surveys less healthy than respondents?", *Journal of Official Statistics*, **18 (1)**, 13-23.

Dixon, J. (2002), "Nonresponse bias in the Consumer Expenditure Survey", *ASA Proceedings of the Joint Statistical Meetings*, 797-802.

Groves, R. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons.

Groves, R. and Couper, M. (1998). *Nonresponse in Household Interview Surveys*. John Wiley & Sons.

Huisman, M., Krol, B., and van Sonderen, E. (1998), "Handling missing data by re-approaching non-respondents", *Quality and Quantity*, **32**, 77-91.

Juster, F. Thomas, and Smith, James P. (1997), "Improving the quality of economic data: Lessons from the HRS and AHEAD", *Journal of the American Statistical Association*, **92**, 1268-1278.

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

Krenzke, T., and Griffin, D. (1997). Who was counted last in the 1990 census? *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Martin, E., Abreu, D., and Winters, F. (2001), "Money and motive: Effects of incentives on panel attrition in the survey of income and program participation", *Journal of Official Statistics*, **17 (2)**, 267-284.

Raab, G.M., and Donnelly, C.A. (1999), "Information on sexual behavior when some data are missing", *Applied Statistics*, **48**, 117-133.

Stasny, E.A. (1986), "Estimating gross flows using panel data with nonresponse: An example from the Canadian Labor Force Survey", *Journal of the American Statistical Association*, **81**, 42-47.

Triplett, T., Wang, K., Safir, A., Steinbach, R., and Pratt, S. (2002), "Using a short followup survey to compare respondents and nonrespondents", *ASA Proceedings of the Joint Statistical Meetings*, 3496-3501.

Van de Kerckhove, W., Krenzke, T., Mohadjer, L., Li, L., Annis, T. (2005). National Assessment of Adult Literacy: ESL Nonparticipant Study Task 3: Analysis Report. Draft report prepared for the National Center for Education Statistics, July 20, 2005.