Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: a Methodological Perspective

NONRESPONSE BIAS ANALYSES AT THE NATIONAL CENTER FOR EDUCATION STATISTICS

Jonaki Bose[1]

ABSTRACT

In surveys with low response rates, nonresponse bias can be a major concern. While it is not always possible to measure the actual bias due to nonresponse, there are different approaches that help identify potential sources of nonresponse bias. In the National Center for Education Statistics (NCES), surveys with a response rate lower than 70 percent must conduct a nonresponse bias analysis. This paper discusses the different approaches to nonresponse bias analyses using examples from NCES.

KEY WORDS: Nonresponse; Bias; Response Rates

1. INTRODUCTION

Traditionally, not all sampled units respond to a survey. The likelihood for nonresponse is further compounded when there are multiple stages or components of response, e.g., screener interviews, multiple respondents associated with a case, or more than one waves of data collection. For example, the National Center for Education Statistics' (NCES) National Household Education Study (NHES), a random digit dialing (RDD) survey, has a screener interview that has a lower completion rate than any of its other components (Nolin et al., 2000). Similarly, convincing sampled schools to participate is the first stage of the Early Childhood Longitudinal Survey: Kindergarten Class of 1998-99 (ECLS-K), and gaining the cooperation of the schools in the first wave has been harder than gaining cooperation for any other components within the study (Brick, Burke and Lê, 2000). This lack of response from sampled units may contribute to bias in survey estimates.

According NCES standards, if the overall survey response rate (product of the completion rate of the different stages) is less than 70 percent, a nonresponse bias analysis must be conducted to identify potential sources of bias in the estimates due to the high nonresponse.

This paper is based on experiences with household, elementary/secondary and post-secondary surveys at NCES, and examines the process of identifying the needs and characteristics of a survey that influence the types of nonresponse bias analyses that are conducted. It also evaluates the purposes, strengths and weaknesses of different techniques. This effort does not focus on the substantive results that were obtained as a result of such analyses.

2. NONRESPONSE BIAS

2.1 Definition and need for nonresponse bias analyses

Bias is the difference between a survey estimate and the actual population value. In a sample survey it can be considered to be the expected value of this difference based on all possible samples. Nonresponse bias associated with an estimate consists of two components—the amount of nonresponse and the difference in the estimate between the respondents and nonrespondents.

[1] Jonaki Bose, National Center for Education Statistics, 1990 K St. NW, Washington DC 20006, USA

The bias of an estimate can be expressed mathematically to show the relationships between the bias and the two factors discussed above. The bias is given by

$$Bias\ (\hat{y}_r) = p_n\ \{E(\hat{y}_r - \hat{y}_n\ )\}$$

where $\hat{y}_r$ is the estimated characteristic based on the respondents only, p_n is the nonresponse rate, $\hat{y}_n$ is the estimated characteristic based on the nonrespondents only, and E is the expectation operator for averaging over all possible samples (Nolin et al., 2000). Bias can be associated with both unit and item nonresponse.

Thus bias is associated with both low response rates and strong differences in the estimates between respondents and nonrespondents. Any estimate from a study can be subject to bias due to nonresponse across one or more stages. The best way to avoid bias is to improve response rates by using methods such as intensive refusal conversion techniques, incentives, multiple modes of data collection, flexible scheduling, and interviewer training. However, despite best efforts, nonresponse does occur. In such cases, surveys adjust probability-based weights to compensate for nonresponse. However, despite adjusting weights for nonresponse, bias can still persist in estimates.

Evaluation of the bias is not always possible as the true value of the population parameter is unknown. Wherever a true population value is known, the difference between the value computed from the survey data and the true population value can be considered an estimate of the bias related to the survey estimate.

A nonresponse bias analysis is the process that results in the quantification of estimated nonresponse bias, and identification of potential sources of nonresponse bias on estimates. Nonresponse bias analyses allow for the evaluation of survey statistics that are estimated using both base (only reflecting selection probabilities) and nonresponse adjusted weights.

There are different ways in which nonresponse bias analyses are useful. Nonresponse bias analyses serve as an indicator of the quality of the data collected, and help identify potentially biased estimates. Such analyses can help reassure data users, as well as the agency collecting and releasing data, of the quality of the data available. Simultaneously, it warns users of data vulnerable to bias. Such analyses can also be used to evaluate the variables used in nonresponse weighting adjustments. In addition, nonresponse studies can identify sources of potential biases that can be addressed in future data collection waves of a longitudinal study. Longitudinal studies can be particularly vulnerable to nonresponse bias, as bias in the first wave of data collection may persist in future rounds of data collection. Repeated cross-sectional surveys also benefit from such analyses. An analysis of nonresponse was conducted on data from the 1993-94 Schools and Staffing Survey (SASS) (Monaco et al., 1997). As a result, numerous recommendations were generated for future SASS studies. These recommendations were then considered during the 1999-2000 SASS.

2.2 Factors Affecting Approaches to Nonresponse Bias Analysis

There are several different factors that affect which approaches to use in a nonresponse bias analysis. Prior to starting an analysis it is useful to identify these characteristics.

Presence of more than one component: The simplest form of a survey is when there is one instrument for all respondents and there are no screeners (i.e., stages) or multiple components. However, there are surveys where there is a screener interview, and based on responses to the screener interview, respondents may be eligible for another survey. For example, in the NHES, based on responses to a screener, household members may also be asked to respond to an additional survey on topics such as adult education, participation in early childhood programs, and participation in before and after school programs. In other surveys, each case may have more than one associated component. For example, in the Early Childhood Longitudinal Study: Kindergarten Class of 1998-99 (ECLS-K), the primary focus is on the child who is administered a child assessment. In addition, data are collected from the child’s teachers, parents and school administrators. It is important to identify the different components, and who was eligible to complete these components prior to conducting an analysis.

Whether the survey is longitudinal or cross-sectional: The approaches for the first wave in a longitudinal and a cross-sectional survey can be similar. However, the evaluation of bias in subsequent rounds of data collection in a longitudinal survey should also take the first wave into consideration.

Presence of multiple weights: This is tied into the first two points regarding whether the survey has more than one component and whether it is longitudinal in nature. Many surveys have more than one weight, even in the case of cross-sectional studies. Prior to analysis it is useful to consider issues such as which weights are appropriate to use with different approaches, whether it is useful to evaluate estimates based on more than one set of weights, and which populations are included depending on the weights used.

3. METHODS OF ANALYSIS

3.1 Examination of Response Rates

As mentioned earlier, nonresponse bias consists of two components: the extent of nonresponse and the difference between the observed outcomes from respondents and the unobserved outcomes from nonrespondents. In general, the first step that can help determine whether there is need for further evaluation, is the examination of the extent of nonresponse in a survey.

In a single-stage survey there is generally one set of response rates that is of interest when studying the extent of nonresponse. In a survey with more than one stage or components there is more than one type of ‘response’ rate that can be useful. In certain NCES surveys, the terms ‘completion’ and ‘response’ rates have been associated with different concepts (Brick, Burke and Lê, 2000).

Completion rates refer to the percentage of participating units at each stage of sampling and are calculated separately for different components and questionnaires. Response rates refer to the overall percentage of participation in the study and take all stages of sampling into account. For example, in the ECLS-K, the response rate is a product of the school response rates (percent of schools that agreed to participate in the study) and the completion rate of a given component. For example, the child assessment response rate is the product of the school response rate and the child assessment completion rate (percent of children assessed conditioned on participation of their school in the study). Completion rates help identify differences within subgroups at the same level, while response rates describe the broader picture but can confound the sources of bias.

In the 1993-94 SASS, response rates were analyzed in great detail (Monaco et al., 1997). Given that the survey had multiple components, e.g., local education agency (LEA), school, teacher, and student, the related nature of nonresponse was examined. They were interested in knowing about the “jointness of nonresponse”. The response rates were tested for whether they were independent across public and private school administrators, public and private schools, public and private teachers, public and private school libraries, public and private school librarians, and LEAs. For example, the analysis found that private school teachers had a significantly higher rate of response when the school administrator from the teacher’s school responded.

In most NCES surveys response and completion rates are computed both without weights and using weights reflecting only the probability of selection (i.e., base weights). Generally, the evaluation of response and completion rates is done using base weights, which do not include any weighting adjustments.

The evaluation of response rates provides us with a starting point. High response rates not only for the entire sample, but also for subgroups, might indicate that there is no need for further analysis of bias due to nonresponse. As mentioned earlier, at NCES, any overall response rate of less that 70 percent requires a nonresponse bias analysis.

Most of the response rates computed are for entire surveys, stages or components. However, nonresponse related to a survey estimate has two components: unit and item nonresponse. There are surveys in NCES that in addition to studying unit nonresponse, also examined the nature of item nonresponse on the surveys.

In the National Education Longitudinal Study of 1988 (NELS) an extensive analysis of item-level nonresponse was conducted (Spencer et al., 1990). In addition to examining item response rates, the analysis also evaluated item nonresponse rates in terms of factors such as position of the item in the questionnaire, topic and whether the item was contingent on a filter. Items with highest levels of nonresponse were then examined by student characteristics such as gender, race/ethnicity and SES, and cognitive scores. The analysis also examined the average number of items not attempted on cognitive tests based on student characteristics. There are additional methods of evaluating item nonresponse. For example, the relationship between item response rates and date of interview was examined in the Baccalaureate and Beyond Longitudinal Study (Green et al., 1999).

3.2 Comparison of Estimates from Respondents to Population Values

In theory, the optimum way to identify bias in the estimates from a sample of respondents would be to compare the estimates to true population values. For the most part, population values are not available. However, there are, on occasion, sources that may provide population values, either for the entire population or subsets of the population. Some useful sources include sampling frames and administrative records.

In NCES school-based surveys, the Common Core of Data (CCD) and the Private School Survey (PSS) are mostly used as frames for public and private schools respectively. These frames are universe surveys that contain variables such as total school enrollment, instructional level, and percent racial/ethnic minority children in the school. In NCES post-secondary institution-based surveys, the Integrated Post-Secondary Educational System (IPEDS) has been used as a frame and it contains variables such as enrollment, control (i.e., public, private), and highest-level offering.

Institution-based surveys estimates from respondents, at both student and school levels, can be compared to population values in order to identify biases. A confidence interval of the difference in the estimates containing the value zero indicates the absence of bias.

For household based surveys, there is often little information, especially in an RDD survey. RDD surveys are mostly restricted to information on exchange-level and broad geographic characteristics associated with each sampled telephone number. There are other frames used by NCES surveys. The Early Childhood Longitudinal Survey: Birth Cohort mainly uses birth certificates as a frame. Birth certificates, relative to other frames, are unusually rich—they provide not only basic demographic information about the child and parents such as age, sex, and race, but also provide information such as details on maternal health and pre-natal practices.

Weighted estimates can be constructed using either base weights or nonresponse-adjusted weights. Estimates using unadjusted (base) weights are useful for evaluating the bias prior to the nonresponse adjustments. Some statisticians prefer using base weights, as data from the frame itself are used in nonresponse adjustments. However, using nonresponse adjusted weights allows for comparison between ‘final’ estimates and population values.

While the examination of the difference between the estimates and population values provides us with an indicator of bias, this process does not differentiate between sampling bias and nonresponse bias. One could separate the nonresponse bias from the overall bias, by evaluating the bias that would be present in the estimates had the all of the sample units responded. This bias would be due to sampling. Alternately, estimates based on the respondents can be compared to the estimates based on the nonrespondents to get a direct indicator of the nonresponse bias.

3.3 Comparison of Survey Estimates to External Estimates

This approach is one of the most common approaches used by both statisticians and researchers in determining the quality of estimates from a survey. Estimates from a survey are compared to estimates from other sources. Some key questions to ask when performing such comparisons are:

a. Are the actual populations of inference the same?
b. Were the questions and responses worded identically?
c. Were they asked in similar contexts?

d. Did the survey use the same mode of data collection?
e. Were the surveys conducted at the same time?

It is clear from these questions that there are difficulties associated with this common method of evaluating the quality of estimates. This approach does not measure nonresponse bias alone. Some of the differences may be due to measurement differences or true changes over time (Federal Committee on Statistical Methodology, 2001). The measurement differences often supercede any difference due to nonresponse. Also, there may be biases associated with the external estimates. However, large differences may be an indicator of potential problems. Even though this method is not very conclusive, it is one of the most commonly used methods. Since most analysts will at least informally conduct such comparisons, this approach allows an agency to anticipate their concerns prior to the release of data. Generally, in order to make the estimates more comparable, nonresponse adjusted weights are used to make comparisons. Additional adjustments can be made to make estimates more comparable. For example, analysts have estimated survey statistics after subsetting both the survey data and the external data in order to make the populations of inference the same or similar.

For RDD surveys, there is an additional source for comparing certain estimates (Nolin et al., 2000). Certain companies collect data such as household income, presence of household members in various age/sex categories, presences of children, educational attainment of household members, and size of dwelling unit at a telephone number level. The NHES-99 used data from such a company to compare data the survey had collected with data collected by commercial vendors for respondents. The match rate for respondents was about 80 percent, i.e., about 80 percent of the respondents to the NHES had a corresponding record to compare against. However, there were sizeable differences between the survey estimate and data from the commercial vendor that led to concerns about the quality of the data from the vendor. Even though some of these variables were available for both respondents and nonrespondents, due to data quality concerns these variables were not used in the nonresponse adjustment process.

At NCES, there are a few non-NCES surveys that are also used to in the comparative process. Estimates from the Current Population Survey (Brick, Burke and Lê, 2000) and the Survey of Income and Program Participation have been used in the past. In other agencies, comparisons have also been made using data from administrative records. Data collected from individuals can be compared to external sources such as their hospital records and insurance claims. In countries with national registries, there may be opportunities to compare such survey data to data from national registries.

3.4 Linking Respondents to Nonresponse

When data are not available from or on the nonrespondents, one analysis approach is to identify those respondents who are most 'like' the nonrespondents. Depending upon the survey design and the weights associated with the data, there are a few different options available.

3.4.1 Surveys with multiple components and weights

In a survey with multiple components, in addition to unit nonrespondents, there are respondents that may have answered some of the components and not others. This can be considered as an additional level of nonresponse, slightly different from unit and item nonresponse. Examining survey statistics based on the degree of component-level nonresponse helps in the identification of possible nonresponse bias that may be introduced in analyses that use data from more than one component.

For example, in a survey such as the ECLS-K, direct assessment data are collected from the sampled child. In addition data are collected from the child's parents, regular and special education teachers, school administrators, and school records. Due to differential response rates between the different components, multiple weights were created. The choice of a weight for analysis depends on whether the analysis uses data collected at one or more time points, the level of analysis and the source of the data (from one or more components). For example, a weight was created for children with direct assessment data in the first wave of data collection (C-1). A weight was also created for children with all three sources of data—child, parent, and regular teacher in the first wave of data collection (CPT-1). A third weight, a panel weight was created for children with all six sources of data in both the first (three sources) and second waves (another three sources) of data collection (CPT-P). Thus any child with all six sources of data would have an associated C-1, CPT-1 and CPT-P weight, any child with all three sources of wave 1 data would have a C-1 and CPT-1 weight, and any child with a wave 1 child assessment would have a C-1 weight. Thus the children with a CPT-P weight are a subset of the children with a CPT-1 weight who in turn are a subset of children with a C-1 weight. The difference between the three pools of children is additional nonresponse. Thus if the same survey statistic was estimated three times, using the C-1, CPT-1 and CPT-P weights, then any difference in the estimates can be attributed to differences due to nonresponse. These differences would have persisted even after additional nonresponse adjustments to each weight and arguably can be considered as bias introduced due to the additional component-related nonresponse. It is important to note that each of the three weights was adjusted separately for nonresponse, and thus in theory should compensate for the different levels of nonresponse (Brick, Burke and Lê, 2000).

Conducting such an analysis can be very helpful as it does mimic how researchers analyze data. Many analyses use data from more than one source within a survey and so nonresponse within components can potentially contribute bias to the survey statistics. As in most approaches, this evaluation is restricted to unit respondents and provides no information about unit nonrespondents with all components missing. The next approach uses respondents to make inferences about nonrespondents.

3.4.2 Comparing ‘Early’ Respondents to ‘Late’ Respondents

One of the key assumptions in such an approach is that later respondents to a survey are more similar to nonrespondents than are earlier respondents. The Beginning Postsecondary Study modeled the pattern of mean response by date of response (Wine et al., 2000). Respondents were divided into groups of approximately the same size (so that the mean response in each group would have approximately the same precision) based on date of interview and type of institution. Trends for the overall population and for subgroups based on the type of institution were examined. Overall, this approach allowed them to identify that additional (late) respondents would be more likely to have attended less-than-4-year institutions and that they would have been less likely to be enrolled in the spring of 1998. This process was conducted using nonresponse-adjusted weights, but can be done using base weights as well. While it is restricted to actual respondents, it does allow for the extrapolation to the characteristics of nonrespondents.

3.5 Follow-Back Surveys

As mentioned earlier, other than using the frame variables to compare respondents and nonrespondents, none of the approaches actually evaluate whether respondents were different than nonrespondents, and the extent to which the differences introduced bias in different estimates. Follow-back surveys also allow for the quantification of estimated bias. Follow-back surveys are designed to collect at least some key or critical variables either from all or a randomly selected sample of nonrespondents. Intensive nonresponse conversion techniques are used to minimize nonresponse in the sample. The presence of these additional variables on nonrespondents, allows for the further quantification of the actual bias due to nonresponse, especially for key estimates or outcome variables. One drawback is the cost associated with such follow-back surveys. In addition, it is very important to have high response rates for the follow-back studies in order for them to fulfill their purpose. NELS conducted such a follow-back survey (Spencer et al., 1990).

3.6 Comparing Estimates Calculated Using Base and Nonresponse Adjusted Weights

The process of creating nonresponse-adjusted weights includes identifying those characteristics most related to nonresponse. Multivariate analyses are conducted to identify subgroups based on differential response propensities. The assumptions are that within these subgroups the respondents and nonrespondents provide similar responses, and that there are large between-subgroup differences. Cells are defined based on common respondent and nonrespondent characteristics. Within each of these cells, defined generally by several variables, an adjustment factor is applied to the weights for the respondents to compensate for the nonrespondents. The goal of such an adjustment is to eliminate or reduce nonresponse bias. The analysis of response propensity can be done using a categorical search algorithm called Chi-Square Automatic Interaction Detection (CHAID). An entire data set can be divided into cells such that all units within a cell have the same likelihood of responding as determined by the analysis.

One way to evaluate the effect of nonresponse adjustments on different survey estimates is to examine estimates using both the base and nonresponse adjusted weights. If there are large differences, it is possible that the adjustment did indeed reduce the bias in estimates. If there are no differences, it is possible, that the original respondent sample was not very different from the nonrespondents, and so there was not much bias to start with. However, it is also possible that the characteristics that were used to identify the cells were not good predictors of response propensity. Overall, this method is useful in evaluating the effects of nonresponse adjustments on estimates, but does not necessarily inform one about the extent of bias associated with survey estimates.

3.7 Other Methods

In a longitudinal study, once data have been collected in the base year from respondents, nonrespondents to subsequent rounds can be compared to respondents to those rounds using more than just the frame data. This does not of course address the issue of initial unit nonresponse, but the process may provide information on the attrition bias that may be introduced due to the additional nonresponse in future rounds of data collection.

There are other options as well. Statisticians in other agencies have looked at partial completes and break-offs relative to complete interviews, with the assumption that those likely to not complete the interview are more similar to nonrespondents. Similarly, other studies asserted that it is possible that refusal converted respondents or respondents who were more difficult to include in the survey due to initial reluctance are possibly more similar to nonrespondents compared to respondents.

4. EVALUATING THE ESTIMATED BIAS

There are different ways to evaluate bias. The absolute value of a bias does not provide much information on the impact of the bias on estimates. There are a few different ways that have been used in NCES surveys to evaluate the estimated bias.

Determining if the bias is different from zero: If the confidence interval constructed around the bias contains zero then the bias can be considered to be not significant. This technique has been used, for example, when comparing survey statistics against population values obtained from the frame. The bias is considered the difference between the survey statistic and the population value, and examining the confidence interval for a zero helps determine if there is any bias.

Comparing the magnitude of bias to the survey statistics: A simple way to look at the bias is to compare it with the survey statistic. Calculating such a relative bias allows for comparisons across different survey estimates. This does not, however, provide information on the bias relative to the confidence one has on the statistic based on the standard error. However, surveys do calculate a mean ‘relative bias’ value based on the mean of multiple relative bias values.

Comparing the magnitude of the bias to the standard deviation: The estimated bias can also be compared to the standard deviation of the survey statistic. The standard deviation of an estimate is often used to identify substantively important differences.

Comparing the magnitude of the bias to the standard error: Another way of evaluating the estimated bias is relative to the standard error. The mean square error can be expressed as:

$$\text{Mean Square Error} = (\text{Bias})^2 + \text{Variance}$$

Thus if the bias is large relative to the standard error, the bias contributes the most to the mean square error. Often in large samples, the bias will be large relative to the standard error.

CONCLUSIONS

In surveys, it is helpful to have high response rates. High response rates do not guarantee low bias in cases where the respondents and nonrespondents are very different, but lower response rates magnify even greater the effects of the difference between respondents and nonrespondents that contributes to the bias. Once data have been collected for a survey, these analyses help determine data quality, identify vulnerabilities in the data, help improve data collection in future waves for longitudinal studies, and subsequent repetitions of cross-sectional surveys. There are many different approaches available. Approaches for each survey can be customized based on characteristics particular to a survey. Comparing against a frame or using data from a follow-back survey are ways to actually quantify estimated bias due to unit nonresponse. It is also possible to estimate the additional bias introduced by using a subset cases with complete data, especially when there are appropriately calculated weights for these subsets. While it may not be possible to get an exact measure of the bias, nonresponse bias analyses form an integral part of the overall assessment of the quality of data.

REFERENCES

Brick, J. M., J. Burke, and T. Lê (2000), "Analysis of Nonresponse Bias in the Base Year, Early Childhood Longitudinal Study: Kindergarten Class of 1998-99", unpublished report, Washington, DC: National Center For Education Statistics.

Federal Committee on Statistical Methodology (2001), *Measuring and Reporting Sources of Errors in Surveys*, Washington, DC: U.S. Office of Management and Budget (Statistical Working Paper 31), pp. 6-1 – 6-13.

Green, P., S. Myers, C. Veldman, S. Pedlow, and P. R. Knepper (1999), "Nonresponse Bias Analysis", *Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-Up Methodology Report*, pp. 62 – 78, (NCES 1999-159).

Monaco, D., S. Salvucci, F. Zhang, and M. Hu (1997), *An Analysis of Total Nonresponse in the 1993-94 Schools and Staffing Survey*, U.S. Department of Education, National Center for Education Statistics (NCES 98-243).

Nolin, M. J., J. Montaquila, P. Nicchitta, K. Kim, B. Kleiner, J. Lennon, C. Chapman, S. Creighton, and S. Bielick (2000), "A Study of Nonresponse Bias in the NHES: 99", *National Household Education Survey: 1999 Methodology Report*, pp. 122 – 143, (NCES 2000-078).

Spencer, B. D., M. R. Frankel, S. J. Ingels, K. A. Rasinski, R. Tourangeau, and J. A. Owings (1990), "School Nonresponse Analysis", *National Education Longitudinal Study of 1988, Base Year Sample Design Report*, pp. 33 – 48 (NCES 90-463).

Wine, J. S., R.W. Whitmore, R. E. Heuer, M. Biber, D. J. Pratt, and C. D. Carroll (2000), "Measures of Bias", *Beginning Postsecondary Students Longitudinal Study First Follow-up 1996-98 (BPS: 96/98),* pp. 6-41 – 6-64, (NCES 2000-157).