
Investigating Nonresponse in Federal Surveys

Carolyn F. Shettle, National Science Foundation

Patricia M. Guenther, Agricultural Research Service

Daniel Kasprzyk, National Center for Education Statistics

Maria Elena Gonzalez, Office of Management and Budget

During the late 1970s the Statistical Policy Office of the Office of Management and Budget organized an interagency committee of Federal statisticians to review areas of statistical and survey methodology affecting the quality of Federal data. This committee, the Federal Committee on Statistical Methodology (FCSM), consists of individuals selected for their interest and expertise in survey methods and their interest in improving Federal statistical data (Gonzalez, 1994). Since its origin, FCSM has studied a number of methodological topics and reported the results in a series of working papers published by the Office of Management and Budget. Although FCSM has studied many topics related to nonsampling errors, it had not directly addressed the issue of nonresponse in Federal surveys prior to 1991.

While survey researchers have always considered nonresponse to be an indicator of the quality of the survey data, their interest in this topic has grown in the last two decades. The Panel on Incomplete Data, established by the Committee on National Statistics (CNSTAT) in 1977, produced three volumes of articles focused on incomplete data in sample surveys (Madow et al., 1983). The Council of American Survey Research Organizations (CASRO) reviewed response rate definitions with the intent of establishing uniformity of definitions across surveys (CASRO, 1982). Steeh (1981) reviewed trends in the response rates in academic surveys, indicating a decline in their response rates over time. During the last ten years, Federal statistical agencies have become increasingly concerned about their ability to maintain high response rates within the constraints imposed by a tight budget climate.

In response to the growing interest in understanding nonresponse in Federal surveys, FCSM organized a Subcommittee with an initial charge to begin an effort to better understand unit nonresponse in sur-

veys, including the levels of nonresponse and measures used to compute nonresponse rates. The proposed approach was to conduct a broad-based review of the level of unit nonresponse rates, currently and over-time in Federal surveys. This paper provides an overview of the Subcommittee's work. It explains the problems we encountered during sample design and data collection, our major findings, and our recommendations for the future. More detailed findings concerning the demographic and establishment surveys studied by the Subcommittee are contained in the papers by Johnson et al. (1994) and Osmin et al. (1994).

■ Problems Encountered during Sample Design and Data Collection

Sample Design

After verifying a central source of information on nonresponse in Federal surveys did not exist, we designed our own data collection. Recognizing the difficulty of designing and implementing a "survey of surveys," we planned a systematic collection of information on a set of Federal surveys. Our goals were to collect response rate information for the period 1981-1991, to learn how Federal agencies measure and document the components of response rates and to determine if any survey design features might affect response rates, as previous literature had suggested.

We could not find an easily accessible sampling frame that offered good coverage of Federally sponsored surveys. Consequently, we selected a purposive sample of major Federal surveys that varied on characteristics we believed to be important, based on our knowledge of the nonresponse literature. We included many of the major ongoing Federal surveys. The sample included 26 demographic surveys and 21 establishment surveys.

Questionnaire Development

Several major differences between demographic surveys and establishment surveys affect the measurement of response rates. First, the difference in methods used to collect information from individuals or households and those used to collect information from organizations results in different needs for monitoring data collection. For example, in surveys of individuals, the nonresponse category called "not at home" is more important than it is in surveys of organizations.

Secondly, the types of statistics commonly reported are different. Demographic surveys typically focus on the estimated number of persons with specific characteristics of interest, while establishment surveys usually focus on totals. For example, a demographic survey might examine the extent to which individuals eat meals away from home, while an establishment survey might examine the total dollar volume of restaurant sales. This difference affects the relative importance of unweighted versus weighted measures of nonresponse. To understand the behavior of individuals, it is as important to obtain information from households in which all meals are consumed at home, as it is to obtain information from households consuming most meals away from home. Thus, weighting is not needed to distinguish between "important" and "unimportant" respondents. However, weighting is desirable in highly stratified demographic surveys to compensate for unequal sample selection probabilities.

In an establishment survey, with a goal of estimating total restaurant sales, a failure to include a restaurant with a high sales volume would be much more serious than a failure to include a small establishment. Thus, for establishment surveys, it is highly desirable to use an appropriately weighted response rate, even in the absence of disproportionate sampling rates.

The third and most obviously important difference is that many establishment surveys are mandated by law, while almost all demographic surveys

are voluntary. This difference affects the motivation of sample members to respond and results in very different refusal conversion procedures. The strong impact of a mandatory appeal in demographic surveys has been demonstrated by Dillman et al. (1994) in their work with pretests of the 2000 Census.

Because of these differences and because of the differences in the backgrounds and interests of the Subcommittee members, we formed two subgroups -- one for demographic surveys and one for establishment surveys. The two subgroups attempted to coordinate their efforts as much as feasible.

Each subgroup developed its own questionnaire, because we believed that it was not possible to construct a single instrument that would meet the needs of both subgroups; however, there was considerable overlap between the two instruments. Both questionnaires asked for a description of how agencies calculated response rates, what components of response rates they captured during data collection monitoring, and what types of post-survey adjustments for unit nonresponse they performed. Johnson et al. (1994) and Osmin et al. (1994) discuss the objectives and detailed contents of the two questionnaires. We pretested the questionnaires within the agencies represented on the Subcommittee and then used the pretest information to revise the questionnaires.

Data Collection

Subcommittee members volunteered to be "shepherds" for the selected surveys. The shepherd for a survey contacted the survey operations staff, identified an appropriate respondent for the questionnaire, alerted the respondent to the due dates, and did whatever was necessary to assist in the completion of the questionnaire.

Due to the persistence of the shepherds, we encountered no unit nonresponse to our survey. However, we received many responses well beyond the requested due date. It was also not unusual for the

shepherd to have to consult several individuals to obtain the information requested. These individuals included, for example, survey managers at the data collection agencies, project managers at the sponsoring agencies, and staff survey statisticians.

We incurred item nonresponse for a variety of reasons. The questionnaire presented great difficulty for some respondents. The degree of difficulty was largely dependent upon how systematically documented and, thus, readily available, the response rate information was and how closely the documented information resembled the information requested. Further, the lack of standard nonresponse terminology affected the quality of the information we were able to gather. In several cases we were unable to reproduce the response rates originally provided by the respondents from the counts they provided; callbacks were necessary for clarification. The response rates we used in our analyses were the ones computed by analysts on the Subcommittee and were not necessarily those provided on the questionnaires. We also eliminated one case from the analytic dataset because it used administrative records and, thus, had a 100% response rate.

Naturally, since the questions on each of the two instruments developed were different, it was not possible to conduct the same analyses with both sets of resulting data. Synthesizing the results of the two subgroups proved to be difficult.

■ Findings

Response Rate Trends

The first task we addressed in our analysis was the description of the trends in response rates. There was a perception among the members of the Federal statistical community, including FCSM, that response rates were declining. We were accordingly prepared to measure the severity of the problem and investigate correlates of it. However, as discussed by Johnson et al. (1994) and Osmin et al. (1994), our results indicated that the response rates for the selected surveys had not, in fact, declined during the time period covered (from 1981 to 1991).

Since the sample of surveys was small and purposive, we cannot generalize our finding of no decline in response rates with confidence. However, given the pervasiveness of the assumption that response rates in Federal surveys have been declining, it is interesting to speculate about the possible reasons for our unanticipated finding:

- ❑ The time period for the comparisons was fairly short. While we attempted to obtain information for a ten-year period, most respondents provided information for a shorter period of time. It is possible that we were unable to detect a decline because of the lack of historical records containing the information needed to track response rates over time.
- ❑ If there were surveys discontinued during the ten year period because of declining response rates, they would not have been available for selection into the study. If this were true, it would have leant an upward bias to the trend data for the remaining surveys.
- ❑ In selecting surveys from their own agencies, Subcommittee members may have disproportionately selected surveys they wished to "showcase."
- ❑ The sensitivity of the Federal statistical community to the problem of nonresponse bias and concern with declining respondent cooperation may have led to increased efforts at obtaining respondent cooperation.

Unfortunately, we did not collect the information needed to distinguish among these alternatives. Most importantly, we cannot determine whether survey sponsors have maintained response rates by increasing expenditures. We failed to collect this information, because we did not think comparable cost information across surveys would be available.

Documentation of Response Rates

There was not a high degree of consistency among the surveys we reviewed in how they measured and reported nonresponse rates.

Most demographic surveys calculated response rate by dividing the number of in-scope respondents by the number of in-scope respondents plus the number of nonrespondents. In a couple of cases, however, the denominator contained the number in the total sample. In one case, an estimate of ineligibles among nonrespondents was made and subtracted from the denominator.

In surveys in which there were multiple sampling stages (e.g., selecting individuals within a household or faculty within a school), there was inconsistency in how the two response rates were combined. For example, some but not all of the surveys multiplied the percentage of responding households by the percentage of individuals responding within households to obtain a final response rate for the survey.

While most demographic surveys only calculated unweighted response rates, a very few also calculated weighted response rates.

The types of response rates encountered in the establishment surveys varied even more widely than in the demographic surveys. In addition to using the same basic calculation typically used in the demographic surveys, some establishment surveys reported coverage-type response rates, i.e., they measured the percentage of a key variable reported by respondents (e.g., the proportion of total production of steel in tons). There were also establishment surveys that included in the numerator of the response rate nonresponding establishments for which they could impute information from alternate data sources.

In addition to variations in how they calculated response rates, surveys varied considerably in the number and types of response and nonresponse components they tracked. The following categories were among the most frequently used categories: "refusals," "temporarily absent," "no one home," "ineligible," "language barrier," "out of business," "out of scope," "postmaster return."

Many of the differences in the response rate calculations and component documentation have their origins in basic differences among the surveys. There are, for example, good reasons that coverage rates

are much more common in establishment surveys than in surveys of individuals. Similarly, some categories are only relevant for certain types of data collection methodologies. For example, "postmaster returns" would be irrelevant in a telephone survey.

However, when we use the single term "response rate" to describe quite different rates, we impede research related to response rates and make it confusing for data users to understand our work. Similarly, to the extent that different nonresponse components have different impacts on nonresponse bias, both research and user understanding are hindered when we use different taxonomies for documenting the nonresponse components.

Post-survey Adjustment Techniques

All the surveys in the study attempted to reduce the impact of nonresponse through post-survey adjustment techniques. However, the techniques used varied considerably among the surveys. Among the methods employed were ratio adjustment or weighting up techniques, post-stratification, raking, regression, and imputation.

■ Recommendations

The Subcommittee made four major recommendations.

Recommendation 1: Survey staffs should compute response rates in a uniform fashion over time and document response rate components on each edition of a survey.

The Subcommittee chose not to recommend that every survey use the same response rate computations. Other groups have recommended such uniformity (see CASRO, 1982). In our view, every definition of response rate components offers some useful information. Some response rate definitions inform the designers about the success of the survey operations; others focus on different causes of nonresponse. One can distinguish between measures that are useful as management tools and measures that data users need to assess the quality of the sur-

vey data. What data users and survey researchers need are clear definitions of the response rate components used.

Recommendation 2: Survey staffs for repeated surveys should monitor response rate components (e.g., refusals, not-at-homes, out-of-scopes, address not locatable, postmaster returns) over time, in conjunction with documentation of cost and design changes.

We believe that response rate components are useful tools to monitor changes in the quality of survey statistics. Response rates should be easily accessible and timely. By themselves, they are not error measures; however, for repeated surveys, changes in response rate components may signal the need for supplementary study of nonresponse error properties. Such changes can alert the survey designers to changes in the "survey-taking climate" that affect data collection, point to changes in the administrative controls over response rates that may need adjustment, and help to measure the effects of any design changes made.

For ongoing surveys, graphs of time series of response rate components, juxtaposed with costs for each collection cycle and indicators of design changes introduced in that cycle, can be valuable management tools. Survey managers need good tools to diagnose the causes of cost changes in data collection activities. Falling response rates, especially those associated with cases requiring much effort prior to the ultimate nonresponse, magnify cost pressures on surveys.

Recommendation 3: Agencies that sponsor surveys should be empowered to report the response rates of these surveys. The sponsoring agency should explain how response rates are computed for each survey it sponsors. Response rates for any one survey should be reported using the same measures over time, so that users may compare the response rates. Response rate components, including actual counts, should also be published in survey reports.

The agencies that sponsor surveys should compute and explain in their survey publications the response rates for each of the surveys they sponsor. Surveys that are sponsored over time should report the same measure of response for all data collection periods, so that users can compare these measures over time.

An agency may need to report response rates for various surveys in different ways, depending on the type of survey design. The method used to compute the response rates should be described in the publications issued.

The results of Recommendations 1 and 2 should be shared routinely with the users of survey data, along with discussions of the relevance of response rates in evaluating the quality of the survey. An analysis of the characteristics of the nonrespondents should be implemented routinely as part of each cycle of data collection.

Recommendation 4: Some research on nonresponse can have real payoffs. It should be encouraged by survey administrators as a way to improve the effectiveness of data collection operations.

We believe that areas of research most likely to yield payoffs include:

- Studies of the relative costs of final efforts to raise response rates, through persuasion, repeated callbacks, and other measures. When these costs are compared to number of cases added to the respondent pool, the relative cost per case can be computed. Studies of the effects of these final cases can be made in an effort to assess the cost effectiveness in terms of mean square error of the final efforts. Examples of recent studies that use such approaches include studies by Kalsbeek et al. (1994) and by Mitchell, Mooney, and Shettle (1994).
- Studies of the measurement error properties of information provided by the reluctant re-

spondent cases, relative to the nonresponse bias in statistics that would omit them from computations. This would address a key question in survey design: When data collectors exert great effort to persuade the reluctant to respond, is one type of error -- nonresponse -- merely exchanged for another type -- reporting error? Perhaps, those persuaded to respond provide less accurate data.

- Studies on what variables should be collected to improve post-survey adjustment for unit nonresponse (see Madow et al., 1983, Recommendation 10(2)). When observable or inferred characteristics of nonrespondent units are related to the survey variables and to the likelihood of participation, then using measures of these characteristics in post-survey adjustment models can be a cost effective method of reducing overall mean square errors. Such an approach is used by the U.S. Department of Agriculture. The selection of appropriate variables has been described by Kott and Guenther (Guenther and Tippett, 1993, Chapters 2 and 3) and the weighting procedure by Fuller, Loughin, and Baker, of Iowa State University (Fuller et al., 1994).

■ References

- Council of American Survey Research Organizations (CASRO) (1982), *On the Definitions of Response Rates*, Port Jefferson, N.Y.
- Dillman, Don A.; Clark, Jon R; and Treat, James B. (1994), "Influence of 13 Design Factors on Completion Rates to Decennial Census Questionnaires," *Proceedings of the Annual Research Conference*, U.S. Bureau of the Census.
- Fuller, Wayne A; Loughin, Marie M; and Baker, Harold D. (1994), "Regression Weighting in the Presence of Nonresponse with Application to the 1987-1988 Nationwide Food Consumption Survey," *Survey Methodology*, 20, 1,75-95.
- Gonzalez, Maria Elena (1994), "Improving Data Awareness in the United States Statistical Agencies," *The American Statistician*, 48, 1.
- Guenther, Patricia M., and Tippett, Katherine S. (eds.) (1993), *Evaluation of Nonresponse in the Nationwide Food Consumption Survey, 1987-88*, U.S. Department of Agriculture, Human Nutrition Information Service.
- Johnson, Ayah E.; Botman, Steven L.; and Basiostis, Peter (1994), "Nonresponse in Federal Demographic Surveys: 1981-1991," *Proceedings of the Section on Survey Research Methods, American Statistical Association* (in this volume).
- Kalsbeek, William D.; Botman, Steven L.; Massey, James T.; and Liu, Pao-Wen (1994), "Cost-Efficiency and the Number of Allowable Call Attempts in the National Health Interview Survey," *Journal of Official Statistics*, Statistics Sweden, 10, 2, 133-152.
- Madow, W. G.; Nisselson, H.; Olkin, I.; and Rubin, D. B. (eds.) (1983), *Incomplete Data in Sample Surveys*, Vols. 1-3. New York: Academic Press.
- Mitchell, Susan; Mooney, Geraldine; and Shettle, Carolyn (1994), "The Impact of Using Priority Mail in a Mixed Mode Survey," *Proceedings of the Section on Government Statistics, American Statistical Association*.
- Osmint, Jeffrey B.; McMahon, Paul B.; and Ware-Martin, Antoinette (1994), "Response in Federally Sponsored Establishment Surveys," *Proceedings of the Section on Survey Research Methods, American Statistical Association* (in this volume).
- Steeh, C. G., (1981), "Trends in Nonresponse Rates, 1952-1979," *Public Opinion Quarterly*, 45; 40-57. ■