Response in Federally Sponsored Establishment Surveys

Jeffrey B. Osmint, U. S. Bureau of Mines
Paul McMahon, Internal Revenue Service
Antoinette Ware Martin, Energy Information Administration *

A continuing concern over the level of nonresponse in Federal surveys led the U.S. Office of Management and Budget (OMB) Federal Committee on Statistical Methodology to commission the Subcommittee on Nonresponse in 1991 (Steeh, 1981; deHeer and Israels, 1992). The purpose of the Subcommittee was to study issues regarding unit nonresponse. The scope of the Subcommittee was limited to unit nonresponse, because there are significant problems in collecting item response data and in interpreting any such data that might be available.

The Subcommittee's investigation centered on a survey of a number of U.S. government statistical offices and selected surveys which they have conducted (Bailar and Lanphier, 1978). At an early date, the Subcommittee's data collection and analysis efforts were split according to the type of survey: demographic or establishment. The differences between demographic and establishment surveys are many and dictate a separate treatment of the subject matter. This paper presents the results of the Subcommittee's investigation for establishment surveys.

■ Background

OMB's *Standard Industrial Classification (SIC) Manual* defines an establishment as "an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed." For the purposes of this paper, we also include surveys of corporations, partnerships, schools, charities, farms, mines, hospitals, manufacturers, financial institutions, and government agencies.

An establishment, for our purposes, is an operation or organization that has a limited societal purpose. That purpose may be to create a profit, promote some policy, support some social goal or heal the sick, but it is, in any case, fairly explicit.

As sampling units, establishments have several features that distinguish them from their demographic counterparts. Demographic units, either families or individuals, can be associated with one geographical location at some specific time, though that association can be quite difficult to document. This specific association may not exist for establishment units, for they may have several sites or none at all.

Establishments have a number of financially and legally mandated bookkeeping requirements (General Accounting Office, 1993). Such requirements are quite limited in the case of demographic units. Further, an establishment's records are subject to regulatory review and detailed specification, as in tax matters. In larger establishments, which would have the more complex dealings, this leads to a centralized office that uses trained employees for record maintenance. These larger organizations are also more likely to engage in strategic planning, thus making them both suppliers and users of survey data.

Another distinctive trait of establishment sample units is the large range of values for monetary and other quantitative fields. Quite modest businesses, for example, will have receipts valued in the millions of dollars, an exceptionally large amount for a demographic unit. All populations, both demographic and establishment, have some observations in the tails of the distributions. But what is striking about establishment populations is that the upper tail is greatly extended by a handful of units. These units are usually well known to the survey designers. In a demographic population it is rare that a handful of sampling units could dominate national estimates, but this is the usual situation for establishments.

*The opinions expressed in this paper are those of the authors and do not necessarily represent those of their agencies.

The smaller population size, distributional characteristics, and the problems with location affect the sample design considerations. They mitigate the usefulness of geographic sampling frames and, at the very least, some list frame is required. The more supplementary data on that list, the better.

The distribution and the dominating effect of the few large entities leads to sample designs that depend on stratification (hence, the need for the supplementary data on the list frame), with probabilities of selection that are related to size. In most cases, not only are there selection rates of 100 percent, but there are also lists of "critical cases" that receive special attention throughout the survey process. Survey resource constraints also lead to the use of model based sample designs in the establishment area (Sarndal et al.,1992).

The sample design may complicate the issue of telling the story of the effectiveness of the survey's execution, however, since the widespread use of different sampling probabilities for the various strata leads to weights of great variability. This raises an important issue in the appropriate measure of nonresponse. If unweighted nonresponse rates are used, then smaller entities carry as much weight as larger operations. Since the real issue is providing a measure of how reliable the estimates from a given sample are, the usual equation -- dividing the number of sample responses (including perhaps the out of scope and ineligibles) by the number originally selected -- can easily have no bearing on the quality of the estimates. For this reason, weighted counts are usually used to evaluate the coverage. Comparison of survey estimates to benchmark estimates (with no adjustment for nonresponse) is also an appropriate measure of quality.

■ Nonresponse

Before presenting the results from the Subcommittee's study, we discuss the various types and the general measures of nonresponse in establishment surveys and the typical methods used to adjust for nonresponse.

Item nonresponse refers to a responding unit's failure to complete individual items on the questionnaire. Unit nonresponse is the failure to obtain any information from a sample unit. For some surveys, when key survey questions or a percentage of the questions are not answered, or the reported data are unusable, the response is considered a unit nonresponse. As stated in the introduction, unit nonresponse in establishment surveys is the focus of this paper.

The practice of using lists to formulate frames in establishment surveys is often the root of nonresponse. The problem in the use of a list is that the list is frequently inaccurate. It may contain units that are indeterminately out of business or out of scope. Also, the address associated with a unit may be incorrect, resulting in a noncontact for the unit. Fortunately, this type of noncontact is usually most frequent among the smaller operations. The use of control lists of major establishments tends to reduce the effects on national estimates of totals, since these units are included in the sample. Another consideration is the fact that establishment surveys are frequently mailed, making it difficult to determine if a nonresponse is a refusal.

■ Response Rates

The response rate is often used as an initial guide to the quality of the survey data. High response rates are perceived, rightly or wrongly, as an indicator that the results are valid.

The absence of a standardized response rate has created a wide diversity in definitions among and within survey organizations (U. S. Department of Commerce, 1977; Council of American Survey Research Organizations (CASRO), 1982). The general form of a response rate (R) is: R=(number of completed questionnaires)/(number of eligible units). Most surveys use this basic formula, but define the components differently according to internal and external uses and survey objectives.

■ Response Rate Components

A review of the literature on nonresponse reveals various usages of the terms "completed questionnaires" and "eligible unit" in establishment surveys. A few common definitions for the "completed questionaire" are: returned questionnaire (includes partially completed questionnaires for some surveys), usable unit, unit responding by publication deadline, and reporting unit for which a measurement is obtained. In some surveys, certainty units for which data are imputed or derived from a secondary source are treated as units that completed questionnaires.

An eligible unit is a unit of the sampling frame that is a member of the target population, neither out-of-scope nor out-of-business (except, for example, in surveys of bankruptcies). These units can be difficult to identify, because of the volatile nature of some frames. A respondent that is no longer in scope or in operation or merges with another entity may become eligible again at a later date. For this reason respondents are often not removed from the frame, but the respondent status is updated.

In practice, eligible units are sometimes defined as: units that are mailed a questionnaire (sometimes includes postal returns and incorrect addresses), units that receive a questionnaire (sometimes includes incorrect addresses), all units in the sampling frame except those that could not be contacted, and all units (addresses) in the sampling frame except those confirmed vacant.

Another widely-used indicator is the weighted response rate. A weighted response rate (WR) is based on a quantity of primary interest for the survey and is defined to be: WR = (total weighted quantity for responding establishments)/(total estimated quantity for all eligible establishments). The denominator in a weighted response rate may be obtained from data collected for a previous reporting period or from outside sources, including administrative records.

Weighted response rates are useful because response rates in establishment surveys usually diminish with decreasing establishment size. The nonresponse of smaller units and/or the use of incomplete or insufficient list frames that tend to omit smaller units may lead to a relatively low unweighted rate but a high weighted rate.

■ Adjusting for Nonresponse

Typical methods used to compensate for unit nonresponse in establishment surveys by order of predominance are: imputation (including regression and hot deck techniques and the use of administrative records), adjustment of weights at the processing or estimation stage, ratio adjustment, raking, post-stratification, and substitution.

The dominance of a small number of establishments often leads to extraordinary efforts to collect information about the largest entities. Considerable amounts of data about firms and organizations are usually available in the public domain. Thus, imputation through various means is a frequent partner to the standard weight adjustments.

Some research analysts ignore missing data and only consider the reported data. However, this practice can introduce major errors in estimates for establishments.

■ Data Collection Method

The Subcommittee developed a questionnaire to obtain unweighted and weighted response rates for each year from 1981 to 1991, information on nonresponse adjustments, and data on selected design features believed to have an effect on nonresponse in establishment surveys. These features include:

- ❑ mode and frequency of data collection,
- ❑ method of follow-up,
- ❑ length of survey (measured as the amount of time needed by the survey respondent to complete the survey), and

- reporting requirement (mandatory or voluntary).

Information helpful for interpreting response trends was also collected, such as definitions of response rates and major changes to the survey. Additional data on survey operational costs, survey readability, survey layout, and other factors possibly related to nonresponse were not solicited, to avoid overburdening the respondents.

The selection of the sample of establishment surveys was complicated by the lack of a computerized database of surveys. This forced the Subcommittee to purposively select a sample of surveys to represent different topics, populations, modes of data collection, and survey lengths. Only surveys that were either conducted or contracted out by the Federal Government were considered for the study.

Questionnaires were mailed to the sponsors of the surveys. To avoid revealing problems with a specific data collection, the sponsors were promised that the identification of their surveys would be confidential. Subcommittee members served as "shepherds" over the questionnaires, collecting and verifying the reported information. Generally, several players are involved in the statistical design of a survey and in the processing of the survey data. Therefore, the sponsors were asked to consult with individuals responsible for the various aspects of the survey, as necessary.

Twenty-one establishment surveys were selected for the Subcommittee's study. The following section presents the results from an analysis of the response rate data for these surveys.

■ Trends in Response for Establishments

Completed questionnaires were returned for all of the 21 selected surveys. For about one-half of the polled surveys, response information was reported for every data collection period from 1981 through 1991. For the remaining surveys, complete response data were not available for various reasons. In some cases the historical records were not maintained or the information was not readily available. For three surveys, the absence of viable data was because a full canvass was not conducted for certain years: survey statistics were generated based on a small sample using a benchmark estimator. In addition, one of the surveys was newly instituted and one was discontinued during the period covered by this study.

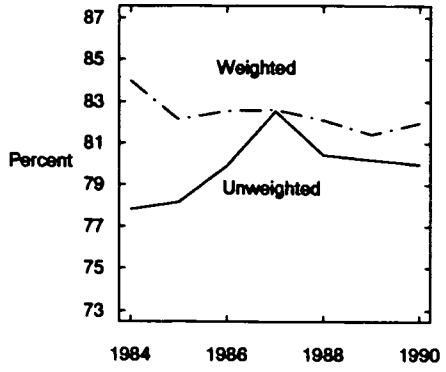
For most surveys, both weighted and unweighted rates were provided. In those cases where only unweighted rates were reported, the survey sponsors had determined that the weighted rates were not sufficiently meaningful to justify the additional expense of calculating and documenting the weighted rates. In the exceptional case of two surveys, weighted rates were provided, but unweighted rates were not.

To analyze the response rates of the surveys over the period of interest, it is more meaningful to limit the analysis to those surveys which reported response rates for several collection periods. Including surveys having only a few collection periods in the analysis might obscure any significant trend or, alternatively, might indicate a trend which does not truly exist. For this reason, the following analysis of time trends in the response rates of establishment surveys will only cover the nine surveys for which both weighted and unweighted rates were reported for six or more collection periods.

Only five of the nine surveys reported rates for the period from 1981 through 1983; only four were able to report rates for 1991. For these reasons, response rates were interpolated and estimated for a small number of selected observations. Rates for monthly and quarterly surveys were averaged to obtain an annual rate.

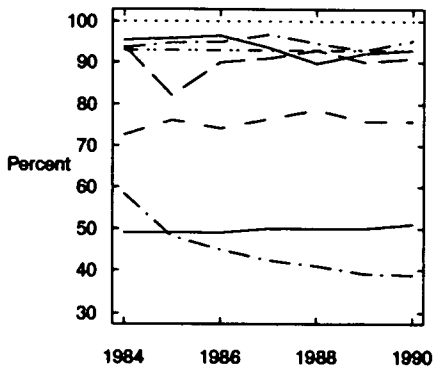
Figure 1 shows the average weighted and unweighted annual response rates for the selected nine surveys from 1984-1990. As may be readily seen, the weighted rate was slightly decreasing, but fairly stable over the period covered by the data,

Figure 1. Average Response Rate
Nine Selected Surveys



while the unweighted rate was slightly increasing. Response rates for the surveys (Figure 2) were also examined separately. Except for one survey, there appeared to be no substantive change over the period studied.

Figure 2. Weighted Response Rate
Nine Selected Surveys



■ Characteristics and Response Rates

Prior to actual data collection and analysis, a number of survey characteristics were thought to be related to survey response rates. Using the collected data, correlation coefficients were calculated to quantify the relative strengths of such relationships. Both Pearson (product moment) and Spearman (rank) correlations were calculated, and contingency tables were tabulated to evaluate the relationships.

The characteristics which were found to be most highly correlated with weighted response rate were: mandatory/voluntary reporting status; the average time required to complete the survey questionnaire; the use of alternates for nonrespondents; and the survey sample size. Of these four characteristics, the correlations for mandatory/voluntary reporting status and average time required to complete the survey questionnaire are noteworthy. The correlations for the other two characteristics are not meaningful. It was not possible to assess the effect of the use of alternate units for nonrespondents, since only two of the polled Federal data collections reported their use. The sample size of one of the reporting surveys dwarfed the sample size of the remaining surveys, thus virtually guaranteeing the appearance of a relationship between survey sample size and response rate.

Sixteen of the surveys reported weighted response rates. As expected, the nine surveys having mandatory reporting experienced an average weighted rate of 92.3%, which was much higher than the 70.5% average rate for the seven voluntary surveys. Rates for the mandatory surveys ranged from 85% to 95%; voluntary rates ranged from 58% to 80%.

The average time for respondents to complete the questionnaire ranged from six minutes to nearly 20 hours. Since this average time is a good measure of the burden which is being placed on the survey respondent, it is somewhat surprising that response rates were higher for the surveys requiring more time to complete. For surveys requiring one hour or less, the average rate was 74.7%, while those requiring more than one hour had a rate of 93.2%. The difference in the rates, however, may be attributed to the fact that the majority of the surveys requiring more than one hour to complete were mandatory, while nearly all of the shorter surveys were voluntary.

The type of data collection unit (establishment, company, etc.) which was the target of the survey was also believed to be related to response rate. The Subcommittee's questionnaire permitted the survey sponsors to respond by specifying more than one type

of unit. Several of the survey sponsors indicated that, in fact, more than one type of data collection unit was utilized. As a result of the diversity of responses, it is not possible to provide definitive statements with regard to the relationship between collection unit and response rates.

Intuitively, the methods of initial data collection and nonresponse follow-up collection should be related to the response rate: personal interview surveys would be expected to have a high response rate. Of the 16 surveys providing weighted rates, 9 reported the use of a single method for the initial data collection: 7 mail, 1 administrative records, and 1 personal interview. Thus, the distribution of the sample does not permit meaningful analysis.

The more frequent (monthly and quarterly) surveys appear to have lower response rates than the less frequent surveys (annual, periodic, and other). There were, however, only three monthly and two quarterly surveys in the sample. In addition, four of the five more frequent surveys were voluntary collections, which required an average of less than one hour to respond.

In conclusion, the great majority of the Federally sponsored data collections which were surveyed fell into one of two categories. The first category consisted of voluntary surveys having a smaller respondent burden. The second category consisted of mandatory surveys with a higher respondent burden. The average response rate for the first category was considerably less than the rate for the second category. The distribution of surveys for each category is shown in Figure 3. "Short" surveys are those having an estimated respondent burden of one hour or less; "long" surveys took more than one hour. Figure 4 is a box and whisker plot showing the distribution of unweighted response rates by reporting status. (Only 19 surveys are represented in Figures 3 and 4, since two of the 21 did not provide an estimated respondent burden.)

Figure 3. Sample Distribution by Status and Burden

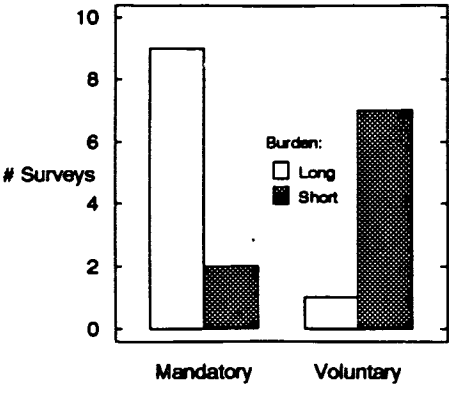
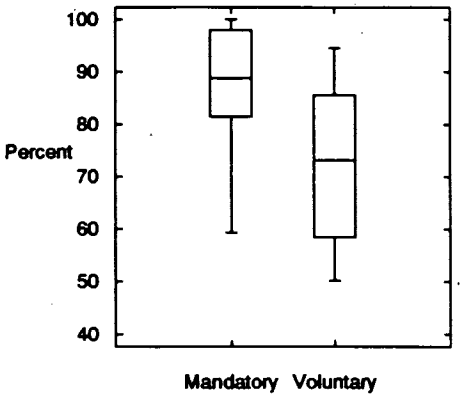


Figure 4. Unweighted Response Rate Distribution by Status



■ Conclusions and Recommendations

The data collected by the Subcommittee does not support a hypothesized decrease in response rates for establishment surveys during the period studied; neither did it suggest improvement. The inability to detect a trend may be due to the small size of the sample, the fact that only 11 years' data were collected, the non-random method of sample selection, or to other factors, such as changes in survey priorities and budgets. Assessing the true impact on response rate of a change in a survey's budget and methods is a very difficult task.

Empirical evidence indicates that the largest discernable factor affecting response rate is the reporting status -- i.e., mandatory versus voluntary reporting (Tulp, D. R., Jr., et al., 1991; Tulp, 1992). None of the contacted agencies reported a change in survey reporting status during the period. Although there might be a slight downward trend in response rates for the decade, any such trend could not be substantial.

Any future study of the response experience of establishment surveys should be more thorough: it should review a longer time frame, include considerably more surveys in the sample, and employ a stratified sample selected from a frame which lists all Federally sponsored surveys.

The first step in managing response policy is to establish definitions of response and nonresponse components appropriate for each survey. The second is to maintain historical response data and document related changes in survey design and processes. Next is implementing system-wide reviews of the recorded response data. The last step is periodically publishing survey response rates.

The data collected during this study indicate that all of the agencies contacted recognize the importance of high response rates and are attempting to "Collect the survey data as fully and accurately as possible, using callbacks and follow-ups as needed to do so." (Madow, et al., 1983). The responses of the agencies, however, do not indicate a strong commitment to the necessity for documenting response components and their rates. In many cases, survey-specific definitions of response and response components do not appear to be readily available. In addition, few of the polled agencies are able to prepare response rates for important domains without some difficulty. We thus arrive at our principal recommendations:

Recommendation 1.-- Each agency should prepare and maintain survey-specific definitions of response and response components for every establishment survey sponsored by the agency. The definitions should be, to the greatest extent possible, consistent with definitions used for other Federally sponsored surveys. These definitions should be periodically disseminated to all survey personnel. Each agency should also develop and document rules for determining when item nonresponse becomes unit nonresponse.

Recommendation 2.-- Each agency should prepare and maintain records of weighted and unweighted response rates for every establishment survey sponsored by the agency. The records should include rates for those components which the agency considers important and should indicate the date on which the rate calculations were performed. The records (response documentation) should include a date and description of each significant change in survey methodology or sample design. The response records of related surveys should be maintained as coherent files, so that response data from the related surveys can be readily linked.

Recommendation 3.--Each agency should formally institute periodic reviews of the response data for all continuing surveys. The frequency of the review will depend on the frequency of the survey being reviewed. Surveys conducted more than once a year should be reviewed at least once a year. The reviews should be directed at: detecting changes in response patterns, auditing and assessing survey frame quality, and examining survey practices that affect response rates. All agency personnel involved, in any capacity, with a survey under review should be included in the review process. All involved personnel should be encouraged to make recommendations for improving response rates.

Recommendation 4.--Each agency should develop and implement a policy of periodically publishing survey response rates. Where possible, the rates and the survey statistics should be published together.

We recognize that, in practice, measuring the impact of survey nonresponse on all estimates is not feasible. However, it is imperative that those us-

ing the estimates are provided with an assessment of the overall survey response rate, the rate for key statistics and for important domains, and a description of the methods used to adjust the estimates for nonresponse.

Recommendation 5.--A computerized database of all Federally sponsored surveys should be developed. The database would include all relevant information such as: survey name, sponsoring agency, OMB Clearance Number, survey and reference periods, mandatory vs. voluntary reporting status, a brief description of major data elements collected, population and sample sizes, a description of the sample selection and rotation plan, and the source of the frame. Data used to implement the database should be annually collected from agencies using a standardized questionnaire. Surveys not requiring an OMB clearance and surveys of administrative records should also be included. The database would be used to facilitate future Committee efforts.

Recommendation 6 (Research).--Based on the Subcommittee's collected data and prior research efforts, we are recommending the three response rate projects described below for future research.

- ***Boundary for Item / Unit Nonresponse*** -- At what point should a respondent's refusal and/or seemingly inaccurate response to individual questions be judged to be unit nonresponse? The answer to this question depends on the subject matter. In the past, research has concentrated on decision criteria based on the logistics of the data editing, validation, etc. processes. This project would develop statistical guidance on such decisions based on an evaluation of the consequences for survey estimates.

- ***Familiarity Survey*** -- Numerous studies on nonresponse have preceded this effort. In addition, OMB's Office of Information and Regulatory Affairs has published the Statistical Policy Working Paper series on quality in Federal data (González, 1994). How familiar are these studies and their recommendations to those designing and conducting surveys? This study would assess survey managers' knowledge and inform them on available literature.

- ***Implementation of Prior Recommendations*** -- Have the recommendations of the Panel on Incomplete Data, the CASRO report, and other similar efforts been implemented? We propose an investigation to measure and evaluate the degree to which the needed changes have been made.

■ Acknowledgment

This paper is based on the work of the Subcommittee on Nonresponse, Office of Management and Budget's Federal Committee on Statistical Methodology, whose members included: Robert M. Groves, Joint Program on Survey Methodology (Chair); Susan Ahmed and Daniel Kasprzyk, National Center for Education Statistics; J. Donald Allen and Peter Basiotis, Department of Agriculture; David Belli, Bureau of Economic Analysis; Steve Botman, National Center for Health Statistics; Eileen Collins and Carolyn F. Shettle, National Science Foundation; Mick Couper, Pamela Powell-Hill, and Maria Reed, Bureau of the Census; Maria Elena Gonzalez, Office of Management and Budget; Patricia M. Guenther, Human Nutrition Service; Paul Hsen, Bureau of Labor Statistics; Ayah Johnson, Henry M. Jackson Foundation; Arthur Kennickell, Federal Reserve Board; Antoinette Ware Martin, Energy Information Administration; Paul B. McMahon and Fritz Scheuren, Internal Revenue Service; and Jeffrey Osmint, Bureau of Mines.

■ References

Bailar, B.A. and Lanphier, C.M. (1978), *Development of Survey Methods to Assess Survey Practices*, Washington, DC: American Statistical Association.

Council of American Survey Research Organizations (1982), *On the Definitions of Response Rates*, Port Jefferson, NY.

de Heer, W.F. and Israels, A.Z. (1992), "Response Trends in Europe," *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

General Accounting Office (1993), *Paperwork Reduction: Reported Burden Hour Increases Reflect New Estimates, Not Actual Change*, GAO/PEMD-94-3.

González, M. E. (1994), "Improving Data Quality Awareness in the United States Statistical Agencies," *The American Statistician*, February.

Madow, W. G.; Nisselson, H.; and Olkin, I. (eds.) (1983), *Incomplete Data in Sample Surveys.* Washington, DC: Panel on Incomplete Data, Committee on National Statistics, National Academy of Sciences.

Sarndal, C.E.; Swensson, B.; and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer Verlag.

Steeh, C.G. (1981), "Trends in Nonresponse Rates, 1959-1979." *Public Opinion Quarterly*, 45; 40-57.

Tulp, D. R. Jr. (1992), "Final Summary of Results for the 1990 Survey of Pollution Abatement Costs and Expenditures (PACE) Mandatory/Voluntary Study," Internal working paper, Industry Division, U. S. Bureau of the Census.

Tulp, D. R., Jr.; Hoy, C. E.; Kusch, G. L.; and Cole, S. J. (1991), "Nonresponse Under Mandatory vs. Voluntary Reporting in the 1989 Survey of Pollution Abatement Costs and Expenditures (PACE)," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 272-276.

U. S. Department of Commerce, Office of Federal Statistical Policy and Standards (1977), *Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics*, Statistical Policy Working Paper 4, Washington, DC. ■