

RESPONSE RATES AS AN EFFECTIVE TOOL IN MANAGING DATA QUALITY

Marilyn M. Seastrom

ABSTRACT

Survey response rates serve as one key measure of the quality of a data set. However, they are only useful to a statistical agency in the evaluation of ongoing data collections if they are based on a predefined set of formulas and definitions that are uniformly applied across all data collections.

In anticipation of a revision of the current NCES Statistical Standards several agency-wide audits of statistical practices were undertaken in the late 1990's. In particular, a compendium documenting major survey design parameters of NCES surveys was drafted. Related to this, NCES conducted a targeted audit of the consistency in response rate calculations across surveys.

Although NCES has had written statistical standards since 1988, the audit of the reported response rates from 50 survey components in 14 NCES surveys revealed considerable variability in procedures used to calculate response rates. During the course of the response rate audit, the Statistical Standards Program staff concluded that the organization of the 1992 Standards made it difficult to find all of the information associated with response rates in the standards. In fact there are references to response rate in a number of separate standards scattered throughout the 1992 Statistical Standards.

1. NCES RESPONSE RATE STANDARDS

The 1992 NCES Statistical Standards include five separate standards that address different aspects of response rates. The first of these standards on survey design states that "The survey must be designed to achieve the **target parameters**," which are:

- For universe surveys, response rates of at least 95 percent for unweighted data and 90 percent for weighted data;
- For cross-sectional surveys, response rates of at least 85 percent and stratum response rates of at least 80 percent;
- For standalone longitudinal surveys, response rates of at least 90 percent in the base year and for each follow-up, and stratum response rates of at least 85 percent;
- For longitudinal surveys drawn from a cross-sectional survey, response rates of at least 85 percent in the base year and 90 percent for each follow-up; and
- The response rate for each critical variable must be at least 90 percent.

The second standard, on achieving Acceptable survey response rates, stresses the need for high survey response rates to ensure that the survey data are representative of the population being studied. Included are guidelines for ways to encourage participation and guidelines for carrying out successful nonresponse follow-up.

The third standard focuses on imputations for item nonresponse. In particular, missing data should be imputed for all key items in a survey, imputation procedures should be internally consistent, imputed values on a data file should be clearly identified, and if published estimates are based on unimputed data, the proportion missing should be reported.

The fourth standard specifies computational formulas for response rates. If the data collection is a sample survey, the response rates should be calculated using data weighted with the sample base weights. In hierarchical designs, the base weight at specific levels of analysis will be the product of all applicable sample base weights, known as the overall response rate.

$$RO = \frac{\text{weighted \# of completed interviews}}{\text{weighted \# of sampled} - \text{weighted \# out-of-scope units}}$$

The *unit response rate* is the ratio of the weighted number of completed interviews to the weighted number of selected sample respondents minus the weighted number considered out-of-scope. Unable-to-contact sample units known to be in-scope are classified as noninterviews; and unable-to-contact sample units of unknown status can be allocated between the nonrespondent and out-of-scope categories.

The *item response rate* is the ratio of the weighted number of in-scope responses for an item to the weighted number of respondents for which the question was intended:

$$RI = \frac{\text{weighted \# of respondents with in-scope responses}}{\text{weighted \# of completed interviews for which question intended}}$$

The fifth response rate standard defines an acceptable level of total response. Here the *total nonresponse* is defined as a measure of both *unit* and *item nonresponse*:

$$TNR = (1 - (\text{overall unit response rate} * \text{item response rate}))$$

Total nonresponse should not exceed 30 percent for any analyzed variable. If the nonresponse falls above this level, or conversely the total response rate falls below 70 percent, the representativeness of the sample must be evaluated by a nonresponse bias study. If nonresponse for the entire sample is unacceptably high, but is acceptable for specific strata or for an analytic sub-universe, the analysis may proceed for the sample subsets with acceptable levels of nonresponse.

2. EVALUATION OF COMPLIANCE WITH NCES RESPONSE RATE STANDARDS

Given the inconsistencies that were discovered during the response rate audit, case flow charts were developed for recent Center data collections and a consistent computational formula was applied across all surveys, in order to evaluate performance against the standards.

In the audits, the initial sample size drawn for a survey was first identified, and the counts of cases were followed through each step of data preparation. Ineligible cases and nonresponding cases were recorded and accounted for in developing unweighted estimates of response rates. In most cases, response rates computed using the base weights were available as well.

Since experience suggests that response rates vary by survey type, the results are displayed separately for: universe data collections; cross-sectional sample surveys, disaggregated by first and second levels of hierarchical designs; random digit dialing surveys, disaggregated by the screener and the respondent (second) level; and longitudinal surveys; disaggregated by hierarchical levels and waves of data collection

2.1 Overall Response Rates

Response rates are generally high in NCES universe data collections, as evidenced by the fact that 92 percent of the universe collections studies had response rates of 90 percent or better (Table 1). Similarly, the response rates for cross-sectional surveys that go directly to a respondent (first level response rates) are respectably high, with 78 percent at 80 percent or better.

Response rates deteriorate a bit when hierarchical designs are employed. Here typically what happens is that the survey first goes to a school to get access to students or teachers, and the student or teacher response rate is compounded by the schools willingness to participate. In this case, the overall response rate for the teacher or student is the product of the respondent level response rate and the school response rate. Most of the surveys reported in the first level response rates have a second component or level as well. In these cases, the overall response rate was observed to slip below the 70 percent NCES reporting standard 28 percent of the time.

Random digit dial (RDD) surveys face the unique challenge of engaging a household member on the phone for a long enough period of time to complete a "screener" survey to determine if there are any eligible survey

participants in the household. NCES has an ongoing RDD household-based survey that was administered six times during the 1990's. RDD surveys tend to have lower response rates than universe or list driven data collections. To date, the screener response rates in the NCES RDD survey have stayed over the 70 percent threshold, with 40 percent between 80 and 89 percent.

These relatively low screener rates present a problem when it comes to computing overall response rates for the second level respondents in these RDD surveys. No matter how high the respondent level rates may be, they are depressed by screener response rates that fall in the low 70's or even the 80's. This is at least a partial explanation of why the overall response rates (second level) for these surveys are low. In fact, less than one-third of the overall response rates at the second level fall above 70 percent and they are in the 70 to 79 percent range. (Since each of the surveys included at least two survey components, there are more response rates at the second level than in the screener.)

As a result of these low overall response rates, the survey managers for NHES routinely conduct nonresponse bias analyses to ensure that the data are representative of the intended population. These analyses are conducted before any results based on these data are reported.

Table 1. Percentage distribution of weighted overall response rates in NCES data collections, by type of collection

	LT 60	60-64	65-69	70-79	80-89	90-100	Number
Universe data collections	0	0	0	0	8	92	12
Cross-sectional sample surveys							
1 st level	0	0	0	22	43	35	40
2 nd level	12	4	12	48	20	4	25
3 rd level	0	0	0	66	33	0	3
Random digit dialing surveys							
Screener	0	0	0	60	40	0	5
2 nd level	21	21	29	29	0	0	14
Longitudinal surveys							
Base year 1 st level	0	33	0	0	66	0	3
Base Year 2 nd level	50	0	25	25	0	0	4
1st wave	66	0	0	33	0	0	6
2 nd wave	66	0	0	17	17	0	6
3 rd wave	100	0	0	0	0	0	1

NCES also conducts a series of longitudinal surveys. Although some of these surveys have multiple levels within each follow-up or wave of data collection, this analysis focuses on the student response rates. The postsecondary education longitudinal studies follow a design that uses a cross-sectional survey as the base year survey, with longitudinal samples drawn from the base year. Both of the two studies considered here, had institutional participation rates in the base year in the 80 to 89 percent range. However, when the student response rates were combined with the institutional response rates, the overall response rates for base year students were lower.

The one elementary and secondary level longitudinal survey included in this analysis suffered from an unfortunately low initial school participation rate (64 percent). In an effort to overcome this start, the survey administrators conducted a follow-up survey of nonparticipating schools. This gave them the information needed to conduct a bias analysis based on the actual characteristics of the nonparticipating schools. Fortunately, this effort produced results that gave the researchers the confidence to proceed with the study (that is to say they found no serious bias in the participating sample).

Nevertheless, the overall response rates for respondents of longitudinal studies are low. One-third of the overall response rates in the first follow-up wave of data collection fall between 70 and 79 percent, and two thirds are

below 60 percent. In the second follow-up wave of data collection, one-third of the rates were over 70 percent and two-thirds fell below 70 percent.

2.2 Respondent Level Response Rates

In an effort to further understand sources of nonresponse, the same framework was used to analyze the individual unit response rates at stages of data collection. By definition, the distributions of the response rates for the universe data collections and the response rates for cross-sectional sample surveys that go directly to the respondent (first level) do not change.

There is, however, a clear change in the response rates of surveys with even one level of hierarchy. Instead of observing three-quarters of the overall response rates below 80 percent, only 16 percent of the individual unit response rates fall below 80 percent (table 2). Once the surveys get to the actual respondent, over 80 percent of the surveys have response rates at 80 percent or above.

Table 2. Percentage distribution of weighted unit response rates in NCES data collections

	LT 60	60-64	65-69	70-79	80-89	90-100	Number
Universe data collections	0	0	0	0	8	92	12
Cross-sectional sample surveys							
1 st level	0	0	0	22	43	35	40
2 nd level	0	0	4	12	48	36	25
3 rd level	0	0	0	0	66	33	3
Random digit dialing surveys							
Screeners	0	0	0	60	40	0	5
2 nd level	0	0	0	7	71	21	14
Longitudinal Surveys							
Base year: 1 st level	0	33	0	0	66	0	3
Base year: 2 nd level	0	0	0	25	25	50	4
1st wave	17	0	0	0	33	50	6
2 nd wave	17	0	0	17	17	50	6
3 rd wave	0	0	0	0	0	100	1

In the case of RDD surveys, the response rates at the individual level are all 70 percent or above, with 92 percent of them at 80 percent or above. In the longitudinal surveys, the individual respondent participation was quite high, with 75 percent of the unit response rates at or above 80 percent. The individual level response rates remain strong in subsequent waves of data collection, with 83 percent above the 70 percent level in both the first and second follow-ups.

For the most part these comparisons of the individual level response rates with the overall response rates computed for the same levels suggest that low response rates associated with the initial participation rates are significant contributors to low overall response rates. In the case of hierarchical cross-sectional surveys, the unit response rates for second level respondents are at least as high or higher than the unit response rates for the first level respondents. For random digit dial surveys, the response rates at the screener stage are clearly the primary source of low overall response rates, as evidenced by the fact that 92 percent of the second level unit response rates are at 80 percent or above. Similarly, in the case of at least one of the longitudinal studies, the low initial cooperation rate was the primary contributor to the low overall survey response rates. In this case, many of the low overall response rates at the second and third levels were accompanied by relatively high individual level unit response rates at each of those levels.

3. DEVELOPING DRAFT REVISED STANDARDS

In the summer of 2000, the Statistical Standards Program at NCES convened a Steering Committee to oversee and direct the standards revision process. The Steering Committee, endorsed by NCES Senior Management, is made up of the Chief Mathematical Statistician, four Senior Mathematical Statisticians from the Statistical Standards Program, and Senior Technical advisors from each of the four major data collection Divisions within NCES.

The Steering Committee met and agreed to keep the current format in which each standard starts with a purpose, includes definitions if needed, and may have multiple subcomponent standards. Many of the standards are supplemented with guidelines, the distinction being that standards are must do items and guidelines are more in the category of best practices that would enhance the quality of the product, but may be taken under advisement subject to time and budgetary constraints.

The Steering Committee established working groups for 15 topics, included among these was response rates. In total, 60 of the 105 member NCES staff agreed to participate in one or more of the standards revision working groups. A quality improvement staff member from Westat facilitated each working group. In addition, external experts were made available to each working group if needed in the process of their deliberations.

Working groups began meeting in early August 2000. Throughout the following months, each working group met regularly, with most meeting two to four times a month for one and one-half hour per meeting, until they felt they had a complete draft of the standard(s) assigned to their group. Several of the working groups that focused on processes completed their tasks quickly, within a month or so. While other working groups focused on statistical issues met throughout the fall, winter, and early spring to complete drafts of their standards. The Response Rate Working Group was one of the groups that met weekly throughout the entire period to develop a draft set of standards for response rates.

Response Rate Standards

Much of the work of the Response Rate Working Group involved reorganizing and better defining the materials contained in the 1992 standards. However, there were some new issues tackled as well. The Response Rate Working Group drafted four separate standards: Computation and Reporting of Response Rates, Survey Response Rate Parameters, Achieving Acceptable Response Rates, and Nonresponse Bias Analysis.

The first draft standard in this set defines unit, overall, item, and total nonresponse, the related response rate calculations, and provides examples of each type of rate. The response rate formulas follow the model established by the American Association of Public Opinion Research (AAPOR) in their monograph "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys." (AAPOR, 2000).

In contrast, the 1992 standards included definitions and incomplete formulas for unit, overall, and item response rates in the Standard for Computation of Response Rates; and the definition of total nonresponse appeared in a separate standard on analysis (the logic being that the specific analysis drives the total response rates)

The second draft standard in this set identifies target response rate parameters for survey design. While every survey may not meet these targets, they are the goals that must be used when a study is designed. This draft standard also indicates the achieved response rate levels at which nonresponse bias analyses are either recommended or required. Where a recommended analysis represents the working groups definition of best practices, and a required analysis is viewed as a minimum requirement.

In drafting this standard, the working group members recognized differences inherent in different data collection methodologies and specified different target parameters for universe surveys (95%), longitudinal sample surveys (90%), cross-sectional sample surveys from a frame (85%), and random-digit dial sample surveys (75% for screener and 95% for components). This working group also specified a target item response rate of 90% for variables that are used as reporting categories.

In considering the achieved response rates, nonresponse bias analyses are recommended for rates below 90% for universe surveys and 85% for the overall unit level response rates for all cross-sectional surveys and for item response rates for key analysis variables. This working group concurred with the 1992 standard calling for a mandatory nonresponse bias analysis if the total response rate drops below 70%. The difference is that the new

draft standards moves this requirement to the standard on other response rate parameters, rather than the standard on analysis.

The draft Standard for Achieving Acceptable Response rates includes much of the information from the 1992 standard. In several cases, guidelines were moved to standards because they describe activities that are required by law and thus are mandatory (e.g. referenced to the Privacy Act of 1974, the National Education Statistics Act of 1994, or the OMB paperwork reduction act). In other cases the language was clarified and new guidelines were added to be more complete and comprehensive.

The draft standard for Nonresponse Bias Analysis is entirely new. The 1992 standards simply called for such an analysis to be completed in cases in which the total response rate fell below 70%. This standard was not enforced uniformly until the late 1990s, and then it was only enforced if the overall unit response rate fell below 70%; in other words the item nonresponse component was not taken into account in decisions as to whether a nonresponse bias analysis was required.

Also, since the standards did not identify any procedures for conducting nonresponse bias analyses, there has been considerable variability in the level of effort, the amount of detail included, and the analytic approaches that were used. In the face of this ambiguity, the Nonresponse Working Group drafted a Nonresponse Bias Analysis standard. The purpose of this standard is “To identify the existence of potential bias due to unit and item nonresponse.” The potential magnitude of nonresponse bias is defined as the product of the response rate and the differences in values of a characteristic between respondents and nonrespondents.

The first subcomponent standard reiterates the requirement for an evaluation of the potential magnitude of nonresponse bias prior to releasing the data when the unit or total response rates fall below 70 percent. The second subcomponent standard specifies that the base weights must be used in assessing the magnitude of unit nonresponse bias for rates below 70 percent. This subcomponent includes five guidelines that describe analytic approaches that may be used with a discussion of the pros and cons of each alternative. The third subcomponent standard addresses the situation where a total response rate is low due primarily to low item response rates. The guidelines here describe analytic approaches more appropriate for assessing the potential bias associated with a low item response rate.

Another new component of the response rate standards that was not addressed in the 1992 standards, the treatment of substitute cases, was identified in the internal review of response rates and the subsequent meeting with external experts convened by NISS. Some data collection programs have used substitute cases based on matched pairs or nearest neighbors to boost the number of cases with the desired distribution of characteristics. The working group added this topic to the standard on the computation of response rates. Specifically the draft standard states:

When substitutions are used, reported response rates should be based only on the initial sample (i.e. without any substituted cases included). Substitutions occur when matched cases are used to replace nonrespondents or ineligible. A matched case occurs when a similar respondent is identified from the frame or an independent sample and added to the fielded sample.

To avoid substitutions, an independent random sample of the population or sampling strata can be released and used in its entirety. In this case reported response rates should be based on the original and the added sample cases.

4. REVIEW OF THE DRAFT STANDARDS

The standards review process for NCES includes both internal and external reviews (Figure 1). There are several steps involved in the internal review and comment process. Each draft standard was made available to NCES staff in an online folder on the Internet. The posting of each standard was announced at the start of a 30-day comment period. In addition to the online comments, a seminar was scheduled midway through the comment period for each draft standard to give the working group an opportunity to present their standard(s) and receive feedback from fellow staff members. This process began in early February and ended in early May.

Each working group was asked to reconvene following the end of the comment period for their standard(s) to consider the comments received and incorporate any necessary changes. These revised draft standards were then submitted to the Steering Committee for review (Figure 2). The Steering Committee currently reviewed each standard to ensure responsiveness to the peer comments, and performed overall consistency edits for style and formatting and identify any remaining gaps.

The Senior Management Team reviewed the draft standards, and necessary changes were incorporated before they were shared with external reviewers. Several rounds of external review will take place during the fall of 2001. First, the data collection and research firms that consult with NCES will be asked to have members of their statistical staffs attend briefing seminars on groups of the standards. The standards will be made available in advance of each seminar, and these experts will be invited to offer comments at the seminars and will be given an end date for the submission of additional comments.

As a next step, two standing NCES advisory committees will be asked to review and comment—the Data Analysis Committee (DAC) for the National Assessment of Educational Progress (NAEP) and the Advisory Committee for Educational Statistics (ACES). The members of each of these Advisory Committees will be given copies of the revised draft standards prior to their meetings with NCES staff. At each of the meetings, NCES staff will brief the reviewers and accept comments. These reviewers will also be given end dates for the submission of comments.

Finally, the National Institute for Statistical Sciences has been commissioned to convene an independent panel of external statistical experts to review and comment on the draft standards before NCES Senior Management and the Commissioner for Education Statistics formally adopt them.

5. SUMMARY

Clearly NCES is still in the midst of the standards revision process. We anticipate completing the project by the end of 2001. Hopefully this discussion has provided a view of the multistage process involved in our standards revision; as well as some insights into the complexities and issues covered by the response rate working group.

REFERENCES

The American Association for Public Opinion Research. (2000). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, Michigan: AAPOR.

United States Department of Education (1987). *Center for Education Statistics Standards & Policies*. Office of Educational Research and Improvement. Washington, DC.

United States Department of Education (1992). *NCES Statistical Standard*. (NCES 92-021r) (nces.ed.gov/statprog/1992stand.asp).