

Reporting on Questionnaire Surveys in GAO Reports, OSM Sections, and E-Supplements

Note: *This guidance is designed to ensure that GAO policies on evidence and generally accepted government auditing standards are met. The guidance conforms to the generally accepted principles and practices of the appropriate disciplines. Statements that particular actions “should” be taken are practices that are expected to be followed, unless there are good reasons for not doing so. Before deviating from a practice expressed as a “should” statement, staff members must consult with an appropriate staff member in Applied Research and Methods (ARM) or a team specialist and must document the consultation.*

Abstract: GAO reports should provide adequate, clear information about their surveys. This information is usually concentrated in OS&M (objectives, scope, and methodology) sections of GAO reports and in e-supplement introductions. This guidance elaborates on principles found under the heading “Reporting on Survey Methods in GAO Products” from the ARM Web site document “Conducting Questionnaire Surveys.”

To conform to this guidance, almost all GAO reports should provide information on the following ten components of their survey research portion of their engagement:

1. objectives addressed by the survey,
2. population the survey is designed to represent,
3. sample frame and the sample design,
4. questionnaire development and testing and procedures,
5. questionnaire content and wording (exact question wordings or reproduction of instrument)
6. key data collection steps (e.g., modes, fieldwork dates, advance contacts, nonresponse followup, editing),
7. disposition of sample (e.g., number of respondents, ineligible)
8. response rate(s) and definitions of rate(s) calculated ,
9. sampling errors (if a probability sample), and
10. extent and direction (if known) of nonsampling errors (nonresponse bias, coverage, measurement, and processing error)

Each of the following sections include examples of appropriate reports statements. Most come from the OS&M sections of GAO reports. Some examples were modified from their report text to better demonstrate a principle. An appendix at the end of this guidance provides an example of an abbreviated survey description that might be appropriate for some testimonies or reports that make minimal use of a survey.

1. Objectives Addressed by the Survey

An OS&M section should clearly link relevant objectives of the GAO report to the general content of the questionnaire. At least one of the objectives should link to the survey. The report objectives need not be fully restated.

- Example 1.1: Brief Purpose Statement with Content Description** To learn about the states’ use of the food stamp options available under the Farm Bill, we conducted a Web-based survey of food stamp administrators in the 50 states and the District of Columbia.¹ For each of the eight Farm Bill options, we asked state officials to provide information on whether or not their state had chosen and implemented the option, reasons for choosing (or not choosing) the option, program challenges in implementing the option, changes because of the option, and potential improvements to the option.
- Example 1.2: Linking Study Objective to Sample Design Decision** To further identify barriers to college access, we sought to determine why the eligible applicants did not use the grant.² We chose to survey parents rather than the eligible applicants, because current contact information for the parents was readily available.
- Example 1.3: Statement within Full Introductory Paragraph** To determine the extent to which actions had been taken to ensure that fugitive felons do not receive Supplemental Security Income, food stamps, Temporary Assistance to Needy Families (TANF), or housing assistance benefits, we obtained information by using telephone interviews and e-mailed questionnaires from state officials who administered Food Stamp and TANF programs in each state and the District of Columbia.³ In our telephone survey, we collected data on the actions these programs had taken to implement the fugitive felon provisions in the Personal Responsibility and Work Opportunity Reconciliation Act of 1996.

¹*Food Stamp Program: Farm Bill Options Ease Administrative Burden, but Opportunities Exist to Streamline Participant Reporting Rules among Programs*, GAO-04-916 (Washington, D.C.: Sept. 16, 2004).

²*D.C. Tuition Assistance Grants: Program May Increase College Choices, but a Few Program Procedures May Hinder Grant Receipt for Some Residents*, GAO-02-265 (Washington, D.C.: Jan. 31, 2002).

³*Welfare Reform: Implementation of Fugitive Felon Provisions Should Be Strengthened*, GAO-02-716 (Washington, D.C.: Sept. 25, 2002).

Mini-OS&M Paragraph

The mini-OS&M paragraph needs to make note of the methodology used, but not to any elaborate degree. A final sentence after introducing the methodology, to the appendix OS&M is sufficient.

Using example 1.1a; For the first example listed above, the mini-OS&M might state: To answer the (x) research question we conducted a Web-based survey of food stamp administrators in the 50 states and the District of Columbia. For further information on our survey, see Appendix 1.

Using example 1.2a; for the mini-OS&M we might say: To answer the (x) research question we conducted a survey of parents of grant applicants. For further information on our survey, see Appendix 1.

Using example 1.3a; for the mini-OS&M we might say: To answer the (x) research question we used telephone interviews and e-mailed questionnaires from state officials who administered Food Stamp and TANF programs in each state and the District of Columbia. For further information on our survey and interviews, see Appendix 1.

As each of the examples illustrates, the point is to make a note of the methodology used, but then refer the reader to the appendix if they want more information. As the point of the appendix OS&M is to convey the appropriate amount of transparency, the point of the mini-OS&M is to convey the minimum amount of transparency that is appropriate for that circumstance.

2. Population the Survey is Designed to Represent

For the simplest study design, a report should identify the exact population that the study represents (the study population). This is often closely linked to the description of the sample frame and design - see wording examples in the next section.

In many GAO engagements, the population that is actually sampled from (the study population) differs from the population that we want to make estimates for (the target population), for reasons described below.

3. The Sample Frame and the Sample Design

The report should describe:

- the population list (the sample frame) from which the sample was drawn,
- the method used to draw the sample, and
- the size of the sample.

The basic structure of the sample design should be described. The report should state whether it was a probability or nonprobability design; if a probability design, the report should say whether it was a simple random sample or involved stratification, multiple stages, or variable

probabilities of selection. The reasoning behind the sample design should be clear. The sample frame description should explicitly identify the title and source of any lists and, if relevant, should identify the date or version of any databases that were the source of the frame. Any data reliability assessments of the sampling frame should be described. More details about how the sample design should be documented can be found in the ARM guidance paper “Documenting Sample Design and Estimates.”

Some parts of the target population may be missing or deliberately excluded from the study population because of errors in the sample frame, lack of contact information, complexity of the study plans, or the costs of data collection. In addition, the sample frame may contain duplicates or otherwise ineligible entries. In complex situations like these, the differences between the target and study populations should be clearly described and any limitations in the coverage or quality of the sampling frame should be disclosed. More details about describing probability samples and reporting their results in the body of reports are provided in the ARM guidance paper “Reporting Results from Probability Samples.”

For studies that do not use probability selection techniques, it is especially important to fully disclose any and all reasons that led to selecting particular cases and to explain why it was reasonable and credible to choose particular types of cases for the study’s specific purposes.

Example 3.1: Simple Design with a Perfect Study Population List

We drew a simple random sample of 1,234 grantees from the Department of Transportation’s TRANSIT database that contained all 23,456 organizations that had been awarded a TransMorePeople grant in fiscal year 2006 (October 1, 2005, to September 30, 2006).

Example 3.2: Design with a Study Population Smaller than the Target Population

To collect information on the extent to which hospitals’ uncompensated care costs were related to treating undocumented aliens, we mailed a questionnaire to a probability sample of 503 hospitals in 10 states—Arizona, California, Florida, Georgia, Illinois, New Jersey, New Mexico, New York, North Carolina, and Texas.⁴ We selected the 4 Southwest states—Arizona, California, New Mexico, and Texas—because uncompensated care costs for treating undocumented aliens have been a long-standing issue for hospitals in communities near the U.S.–Mexico border. We selected the 6 other states because large estimated numbers of undocumented aliens resided there in 2000, according to the Immigration and Naturalization Service (INS). In all, an estimated 78 percent of the population of undocumented aliens resided in these 10 states in 2000.

We sent our questionnaire to a randomly selected stratified sample of 503 of the 1,637 short-term, nonfederal, general medical and surgical care hospitals that had an emergency department—according to either the American Hospital Association’s annual survey database, fiscal

⁴Adapted from *Undocumented Aliens: Questions Persist about Their Impact on Hospitals’ Uncompensated Care Costs*, GAO-04-472 (Washington, D.C.: May 21, 2004).

year 2000, or the Centers for Medicare and Medicaid Services Provider of Service File, as of the end of 2000. Table 3 shows the characteristics of the study population from which the hospitals were sampled. We did not assess the reliability of these databases

From this population of 1,637 hospitals, we sampled 100 percent of the 53 hospitals in Arizona and 31 hospitals in New Mexico. The remaining 419 sampled hospitals were drawn within strata defined by the state, hospital-ownership, and county estimates of the presence of undocumented aliens. The hospital-ownership strata were divided between private and public hospitals to ensure that we had enough private hospitals to examine experiences in this type of facility.

Relatively large numbers of hospitals were drawn from counties with large proportions of aliens to obtain hospitals that were more likely to treat many undocumented aliens. For sampling purposes, we developed estimates of undocumented aliens as a percentage of the population by county by (1) dividing INS estimates of the number of undocumented aliens in each state by Census Bureau estimates of the number of foreign-born noncitizens in the state and (2) applying this ratio to Census Bureau estimates of the number of foreign-born noncitizens in each county.

Example 3.3: Stratified Probability Sample from a Sample Frame Larger Than the Population

We drew a stratified, national probability sample of school districts to study commercial activities in schools.⁵ The sample was based on the Department of Education’s comprehensive database of all U.S. school districts in the Common Core of Data Local Education Agency file for the 2000–2001 school year. We drew our initial sample of 271 from 14,553 local districts—that is, districts that were not administered by state or federal authorities.

In the course of our study, we learned that some of the 14,553 districts were special education and other units that did not have legal authority to establish formal policies. As a result, we estimate that our study population of local school districts with legal authority to regulate commercial activities consisted of 13,866 districts in the 50 states and the District of Columbia.

Our stratified probability sample of 271 districts was drawn from the three strata shown in table 3. The sample strata were designed to draw relatively large numbers of districts from states likely to include districts that had engaged in, or planned to engage in, one or more

⁵Adapted from *Commercial Activities in Schools: Use of Student Data Is Limited and Additional Dissemination of Guidance Could Help Districts Develop Policies*, GAO-04-810 (Washington, D.C.: Aug. 20, 2004).

specific activities involving the collection, disclosure, or use of student information for marketing, selling, or providing information to others for these purposes. Because we thought the activities of interest were low-incidence activities, we wanted to maximize our ability to examine situations involving the use of student data for commercial purposes. We defined the expected high-activity strata as states that we identified as having laws that permitted commercial activities when we performed our work in 2000. Each district was subsequently weighted in the analysis to correctly represent the total number of districts in the respective strata, including those that were not studied.

Example 3.4: Selection Based on Nonprobability Methods

To gain a range of views from states, we selected a nonprobability sample of 10 states, consisting of the 5 states that had had the most sites proposed to the National Priorities List (NPL) in the past 5 years (California, Florida, New Jersey, New York, and Texas) and the 5 that had had no sites proposed in the past 10 years (Arizona, Delaware, Nevada, North Dakota, and Wyoming).⁶ We could contact only a small number of states with the available resources. The 5 most active states were selected because they included 44 percent of the sites identified in the past 5 years. The 5 without sites were included to determine what issues, if any, states had with supporting the listing of sites on the NPL. Results from this nonprobability sample cannot be used to make inferences about all states, because the 40 states with smaller numbers of sites had no chance of being selected as part of the sample.

Example 3.5: Combination Selection for Reliability Assessment Subsample: Nonprobability for 18 Offices, Probability for 8 Decisions within the 18

We examined the *Federal Register* notices for all 101 decisions to obtain information we needed for our study.⁷ For the 40 decisions without this information, we asked the 18 field offices responsible for the decisions to provide the missing information. As a result, we obtained the required information for all 101 decisions. To assess the general accuracy of the information reported in the *Federal Register* notices, we selected a nonprobability sample of 8 of the 61 decisions for which complete information was available in the *Federal Register* notices.

⁶*Superfund Program: Current Status and Future Fiscal Challenges*, GAO-03-850 (Washington, D.C.: July 31, 2003).

⁷Adapted from *Endangered Species: Fish and Wildlife Service Uses Best Available Science to Make Listing Decisions, but Additional Guidance Needed for Critical Habitat Designations*, GAO-03-803 (Washington, D.C.: Aug. 29, 2003).

We selected these 8 decisions in the following way. To minimize the burden on the Service’s field staff, we limited our selections to decisions that were the responsibility of the 18 field offices that we had already intended to contact. We selected 3 decisions by randomly choosing one decision from each of the 3 field offices with the most decisions. We randomly selected the remaining 5 decisions from among the 23 decisions at the remaining 15 field offices. Since the 18 offices were not randomly selected, we could not use the results to make inferences about the population of all decisions—specifically, offices whose *Federal Register* notices were all complete had no chance of being selected as part of the sample.

4. Questionnaire Development and Testing Procedures

Reports should describe the questionnaire development and testing procedures. Reviews of a draft questionnaire by GAO, external subject matter experts, and independent GAO survey experts should be mentioned. Pretests should be described, including (1) whether the pretesting was done in person, over the telephone, or on the Web or by some combination of these methods; (2) the number of versions tested; (3) the total number of pretests; and (4) the characteristics of the pretest respondents and how they were selected. A description should also be provided if there was a large-scale pilot study of the entire data collection procedure with large numbers of respondents.

Special note should be made of any verification of data during pretesting—for example, the use of separate documentation to check respondents’ answers or of especially detailed or structured probes on some topics during pretests. Any relationship between the current and any previous questionnaires should be discussed, especially if the results from the current survey are to be compared with the results of the earlier versions.

Example 4.1: Multiple Reviews, Mixed Mode Pretesting with Detailed Selection Criteria, No Pilot or Related Study

After we drafted the questionnaire, we asked for comments from knowledgeable officials of the Association of Local Air Pollution Control Officials and from independent GAO survey professionals.⁸ We conducted pretests to check that (1) the questions were clear and unambiguous, (2) terminology was used correctly, (3) the questionnaire did not place an undue burden on agency officials, (4) the information could feasibly be obtained, and (5) the survey was comprehensive and unbiased.

We chose the four pretest sites to include major subgroups: states with “delegated” authority, states with and without local permitting authorities, and locations across a wide geographic area. We conducted two pretests in person and two over the telephone. We made changes to the content

⁸Adapted from *Clean Air Act: Key Stakeholders’ Views on Revisions to the New Source Review Program*, GAO-04-274 (Washington, D.C.: Feb. 2, 2004).

and format of the questionnaire after both reviews and after each of the first three pretests, based on the feedback we received.

Example 4.2: Internal Review, Pilot Study, Pretest on Population with Same Mode as Full Survey

For the pilot study, we selected a stratified random sample of 100 employers.⁹ We selected the 100 employers from four groups: employers who had 2 to 49 employees, 50 to 499 employees, 500 or more employees, and an unknown number of employees. We expected to use these stratification categories in the full survey.

Before we began telephoning for the pilot study, we revised the questionnaire to reflect comments from an independent reviewer within GAO. We revised the questionnaire again after the pilot study was complete, and we revised it again after we further pretested the questionnaire over the telephone with two additional businesses from the study population.

Example 4.3: External Review, Multimode Pretests with Verification

The practical difficulties of developing and administering a structured interview guide may introduce errors—from how a particular question is interpreted, for example, or from differences in the sources of information available to respondents when answering a question.¹⁰ Therefore, we included steps in developing and administering the structured interview guide to minimize such errors. We pretested the guide in person at one location and conducted a second pretest by telephone. We also obtained comments on a draft of the guide from federal officials knowledgeable about woody biomass.

We modified the structured interview guide after considering the questions and comments we received as a result of these steps. To verify that the information we needed to gather could be reliably obtained through structured interviews, we compared the pretest interviewees' responses with other information, such as contracts, third-party evaluations of users' activities, and financial analyses. Because the documentary evidence we reviewed agreed with the information from the structured interviews, we believe the data are sufficiently reliable for describing the factors that facilitate users' use of woody biomass in our review.

⁹Adapted from *Workforce Investment Act: Employers Are Aware of, Using, and Satisfied with One-Stop Services, but More Data Could Help Labor Better Address Employers' Needs*, GAO-05-259 (Washington, D.C.: Feb. 18, 2005).

¹⁰Adapted from *Natural Resources: Federal Agencies Are Engaged in Various Efforts to Promote the Utilization of Woody Biomass, but Significant Obstacles to Its Use Remain*, GAO-05-373 (Washington, D.C.: May 13, 2005).

5. Questionnaire Content and Wording

The report should make available a copy of the full questionnaire or the exact wording of at least the questions that are used in the report. Weighted frequencies for the questionnaire can also be presented, if appropriate. If the questionnaire is in the appendix, it should be mentioned in the OS&M and any other appropriate place in the text. If there is an e-supplement, a web link to this supplement should be provided. GAO guidance for preparing e-supplements is in “Guidance on E-Supplements” - accessed in the EAGLE, Section 3.1.1 “Determining Methods” under the Policy and Operational Guidance tab.

Example 5.1: Questionnaire Presented in Appendix The questionnaire used for this study is in appendix II.

Example 5.2: E-Supplement Link Detailed survey results are available at www.gao.gov/special.pubs/gao-04-337sp.¹¹

Example 5.3: Question Presentation Preserving Formatted Appearance To gauge the state’s ability to meet future requirements, we asked the following question:
32. How capable is your state’s financial assurance fund of meeting future demands upon it? [Check one].
 Able to meet all
 Able to meet most.....
 Able to meet some
 Not able to meet any.....

Example 5.4: Question Wording Presented in a Sentence in the Text Respondents were asked, “To what extent do you believe that your service consistently assigns people with the skills and experience to be effective program managers? (*Check only one answer.*) - - Very great extent, Great extent, Moderate extent, Little extent, No extent, No basis to judge.”

¹¹Survey of State and Local Air Quality Officials Opinions on the Impacts of the Environmental Protection Agency’s Revisions to the Clean Air Act’s New Source Review Program, GAO-04-337SP (Washington, D.C.: Feb. 2, 2004). See GAO, Month in Review Feb. 2004, www.gao.gov.

6. Key Data Collection Steps

The report should identify and discuss the data collection and follow-up methods and procedures. The data collection modes used for the survey should be identified, whether Web-based, postal mail, fax, e-mail, or personal or telephone interview. If several modes of data collection were used, some discussion of how questionnaires differed across the modes should be included. The collection procedure should describe the intended respondent—for example, the type of agency official contacted to provide information about a program.

Each follow-up and reminder procedure should also be described, indicating its timing, the mode (e-mail, letter, re-sent questionnaire, telephone call), the number of attempts, and the number of respondents contacted. Any other data collection procedures, such as advance contacts, should be discussed in this section as well. Finally, the time period in which respondents completed the questionnaires should be specified.

Example 6.1: Web-based Mode with Follow-Up E-Mails and Telephone Calls

We developed and administered a Web-based questionnaire accessible through a secure server.¹² When we completed the final survey questions and format, we sent an e-mail announcement of the survey to 282 Federal Communications Commission advisory committee members, including the committee chairmen, on March 17, 2004. They were notified that the questionnaire was available online and were given unique passwords and usernames on March 24, 2004. We sent follow-up e-mail messages on April 15, 2004, to those who had not yet responded. Then we contacted all remaining nonrespondents by telephone, starting April 22, 2004. The questionnaire was available online until June 11, 2004.

¹²Adapted from *Federal Communications Commission: Federal Advisory Committees Follow Requirements, but FCC Should Improve Its Process for Appointing Committee Members*, GAO-05-36 (Washington, D.C.: Dec. 10, 2004).

Example 6.2: Electronic MSWord Questionnaire Mode with Follow-Up E-Mails and Telephone Calls

We surveyed all National Parks Service units where air tour operators had applied for operating authority, including existing and new entrant operators.¹³ We sent the questionnaire by e-mail in an attached Microsoft Word form that respondents could return electronically after marking checkboxes or entering responses into open answer boxes. Alternatively, respondents could return it by mail after printing the form and completing it by hand. In an e-mail in advance of the questionnaire, we asked the relevant official at each park if he or she were the correct respondent and, if not, we asked for a referral to the official who was. Six officials indicated that someone else within the park would be a more appropriate respondent, and we addressed further correspondence to those individuals.

We sent the questionnaire with a cover letter on July 19, 2005. Two weeks later, we sent a reminder letter, attaching an additional copy of the questionnaire, to everyone who had not responded. We telephoned all respondents who had not returned the questionnaire after 4 weeks and asked them to participate. All questionnaires were returned by August 24, 2005.

Example 6.3: Mail Questionnaire with Follow-Up Letters and Telephone Calls

The Social Security Administration provided us with home addresses for all employees of Region X as of August 27, 2002.¹⁴ On September 18, 2002, we mailed 1,801 questionnaires to those addresses. We mailed a reminder postcard to all 1,801 addresses 1 week later, and 2 weeks later we made a follow-up mailing of the questionnaire to those who had not yet responded.

One person returned the questionnaire with an indication of being no longer employed at Region X. For four questionnaires that were returned as undeliverable, we could identify no correct address. Completed questionnaires were accepted until December 20, 2002.

In addition to follow-up to prompt nonrespondents to answer the survey, GAO may also follow-up to verify or supplement respondents' data. If GAO used a structured instrument for the primary data collection but the follow-up used less structured interviews, then the extent to which the follow-up interviews were structured should be indicated. Each follow-up program should be described, along with the number and mode of the follow up. The potential bias from calling only some respondents should be discussed.

¹³*National Parks Air Tour Management Act: More Flexibility and Better Enforcement Needed*, GAO-06-263 (Washington, D.C.: Jan. 27, 2006).

¹⁴Adapted from *Equal Employment Opportunity: SSA Region X's Changes to Its EEO Process Illustrate Need for Agencywide Procedures*, GAO-03-604 (Washington, D.C.: July 16, 2003).

Example 6.4: Re-Contact to Verify Responses

We conducted 17 follow-up discussions with Office of Small and Disadvantaged Business Utilization (OSDBU) directors—15 interviews and 2 e-mail discussions.¹⁵ The purpose of follow-up was to confirm the answers of respondents who said they did not view one or more functions listed in section 15(k) as a role of their office. On the basis of these discussions, we changed 23 of the original answers from “director not viewing a 15(k) function as an OSDBU role” to “director viewing the 15(k) function as an OSDBU role.”

We used two criteria to make these changes. The directors (1) explicitly stated that they wanted to change their answers and gave a reason for the change or (2) misunderstood the question. Thus, in our findings, we report these adjustments as a “yes” response.

Example 6.5: Re-Contact to Obtain Additional Information

To supplement the survey and to elaborate on survey responses, we selected 10 states in which to conduct follow-up telephone calls; we based our selection of these 10 on their answers to the questionnaire’s open-ended questions.¹⁶ The calls helped us obtain more specific examples of states’ experiences in preparing for the Child and Family Services Review; developing, funding, and implementing a program improvement plan; and working with the Administration for Children and Families to improve their child welfare systems.

7. Disposition of Sample

The outcomes from the sample should be described with a count of the number of respondents and a tabulation of the originally drawn sample by type of outcome. In the simplest case, this requirement can be met by reporting the number of usable returns, the number of sample elements drawn, and the ratio of the two (the response rate).

Many studies require more complex descriptions that consider issues raised by ineligible sample elements, sample elements of undetermined eligibility, elements that were never contacted, elements that refused, response dispositions that differ by sample strata, and returned questionnaires with only partially completed data.

Response rates are typically reported in conjunction with sample dispositions.

¹⁵*Small and Disadvantaged Businesses: Most Agency Advocates View Their Roles Similarly*, GAO-04-451 (Washington, D.C.: Mar. 22, 2004).

¹⁶GAO-04-333.

8. Response Rates

The type(s) of response rates that are appropriate to report may differ from sample to sample. In the simplest case, mentioned in the previous section, the response rate is the ratio of the number of usable responses over the number of sample elements drawn. A survey that estimates a population total of some quantity that varies widely across sampled units may call for weighted response rate. In such a case, a weighted response rate may more accurately reflect the level of participation - large units which contribute relatively more to the estimate of a total would have a larger “weight” on the response rate. In addition, there are other specialty rates that measure the level of contact versus the level of cooperation upon contact.

The report should explain the calculation of the response rate used.

Most of the issues in reporting on more complex samples are discussed in “Calculating and Reporting Response Rates” ARM guidance that contains examples of sample disposition tables for different types of surveys, as well as procedures for correctly calculating response rates at GAO. The nonsampling error section of this same guidance contains related discussions of caveats that should accompany low response rates and item response rates.

Example 8.1: Simple, Population Survey (50 States)

Questionnaires were completed by state water officials in 47 states, for a response rate of 94 percent.¹⁷ We telephoned the officials in the 3 remaining states (California, Michigan, and New Mexico), but they declined to participate within the study’s time period.

Example 8.2: Stratified Sample with Some Ineligible Elements

We had selected the names of 231 broker-dealers from our original sample frame of 3,781.¹⁸ We took steps to contact the 231 but, as table 6 shows, we found that 43 were not eligible because they were no longer independent firms. We received 164 usable broker-dealer responses from the remaining 188 eligible broker-dealers, for an unweighted response rate of 87 percent.

¹⁷Adapted from *Freshwater Supply: States’ Views of How Federal Agencies Could Help Them Meet the Challenges of Expected Shortages*, GAO-03-514 (Washington, D.C.: July 9, 2003).

¹⁸Adapted from *Anti-Money Laundering: Efforts in the Securities Industry*, GAO-02-111 (Washington, D.C.: Oct. 10, 2001).

Table 8: Disposition of Broker-Dealer Sample

Sample stratum	Population: independent broker-dealers		Initially drawn sample	Not eligible: not independent broker-dealers	Sample disposition			Response rate
	No. in sample frame	Revised population estimate			Eligible: independent			
					Nonresponse		Usable response	
Refusal	All other							
Large: \$230 million+	166	127	81	19	4	2	56	90%
Medium: \$1 million to \$230 million	1,472	1,256	75	11	4	4	56	88
Small: up to \$1 million	2,143	1,772	75	13	7	3	52	84
Total	3,781	3,155	231	43	15	9	164	87%

The 24 nonrespondents were divided between 15 refusals and 9 who had said they planned to respond but who had not responded by the time the study period ended. Because the response rate of the more numerous, smaller broker-dealer strata was lower, the overall weighted response rate was 86 percent. This total weighted response rate is the ratio of the weighted number of usable responses over the revised estimate of the population size.

9. Sampling Errors

Sampling errors, or other measures of sampling precision such as confidence intervals, should be provided for probability samples. The OS&M section or associated appendixes may provide sampling errors for each result individually. Commonly, however, we make a blanket statement giving either the largest sampling error for any statistic in the report or a range that encompasses almost all statistics together, with a caveat that we have noted in the report any statistics with larger sampling errors.

Reports differ in the extent to which they define or interpret sampling errors. In testimony and reports for which sampling errors have minor importance, we may report only their size, without more detailed interpretation. Where sampling errors have great importance, they should be explained carefully. Examples below give alternative sampling error statements.

A statistician should review all sampling error statements to ensure that technical terms—sampling error, confidence interval, margin of error, and the like—are used appropriately. ARM’s “Glossary of Statistical Terms” defines some common statistical terms. More details about reporting sampling

errors in reports, as well as in OS&M sections, are provided in ARM’s guidance paper “Reporting Results from Probability Samples.”

Example 9.1: Minimum Acceptable Statement with Only Percentages

The percentages in this report are subject to sampling errors of as much as plus or minus 10 percentage points at the 95 percent confidence level.

Example 9.2: Alternative Statement for Testimony Citing a Single Result

Our sample survey finding that 23 percent were affected is surrounded by a 95 percent confidence interval that extends from 18 to 28 percent.

Example 9.3: Short Blanket Statement for Both Percentages and Means

All percentage estimates from our survey are surrounded by 95 percent confidence intervals of no more than plus or minus 9 percentage points, unless otherwise noted. All other numerical results have margins of error of no more than plus or minus 6 percent of the value of those numerical estimates, unless otherwise noted.

Example 9.4: Extended Statement Explaining and Disclosing Sampling Errors

Because we followed a probability procedure based on random selections, our sample is only one of a large number of samples that we might have drawn. Since each sample could have provided different estimates, we express our confidence in the precision of our particular sample’s results as a 95 percent confidence interval.

For example, our estimate that 48 percent of the agency officials are in compliance is surrounded by a 95 percent confidence interval of plus or minus 7 percentage points that extends from 41 percent to 55 percent. This is the interval that would contain the actual population value for 95 percent of the samples we could have drawn. As a result, we are 95 percent confident that each confidence interval in this report includes the true values in the study population.

10. Nonsampling Errors

The OS&M section should contain a statement about all sources of survey error, including nonsampling error. The discussion of nonsampling error should cover nonresponse, coverage, measurement, data processing, and any known or strongly suspected errors in the survey. There should be some discussion of steps we took to reduce nonsampling error, such as pretesting, questionnaire editing, verifying the keypunching, and checking internal consistency. If nonresponse bias analyses were conducted, they should be described. More details about reporting nonsampling errors are in the ARM guidance paper

“Evaluating and Reporting Nonsampling Errors in Surveys.” More details about reporting on nonresponse bias are in “**Addressing Nonresponse and Nonresponse Bias Issues in Surveys.**”

Example 10.1: General Statement

Because this was not a sample survey, it has no sampling errors.¹⁹ However, the practical difficulties of conducting any survey may introduce errors, commonly referred to as nonsampling errors. For example, difficulties in interpreting a particular question, sources of information available to respondents, or entering data into a database or analyzing them can introduce unwanted variability into the survey results. We took steps in developing the questionnaire, collecting the data, and analyzing them to minimize such nonsampling error.

For example, social science survey specialists designed the questionnaire in collaboration with GAO staff who had subject matter expertise. Then, we pretested the draft questionnaire with a number of state officials to ensure that the questions were relevant, clearly stated, and easy to understand. When we analyzed the data, an independent analyst checked all computer programs. Since this was a Web-based survey, respondents entered their answers directly into the electronic questionnaire, eliminating the need to key data into a database, minimizing error.

Example 10.2: Reducing Nonsampling Error in Data Entry, Processing, and Analysis

Finally, surveys may be subject to error in entering, processing, and analyzing data.²⁰ We verified the accuracy of a small sample of keypunched records by comparing them with their corresponding questionnaires, and we corrected the errors we found. Less than 0.5 percent of the data items we checked had random keypunch errors that would not have been corrected during data processing. Analysis programs were also independently verified.

Example 10.3: Nonresponse Analysis

We compared key characteristics of respondents and nonrespondents.²¹ We performed an analysis to determine whether respondents and nonrespondents differed significantly on several key characteristics. Separately for respondents and nonrespondents, we estimated the percentage of schools that participated in both the Direct Loan Program and the Federal Family Education Loan Program and the proportion of schools that participated in the Direct Loan Program for 6, 7, or 8 years. We performed this analysis for all Direct Loan schools and separately for each school type. For most of the comparisons, these characteristics did not differ significantly between respondents and nonrespondents.

¹⁹GAO-04-333.

²⁰GAO-02-111.

²¹*Direct Student Loan Program: Management Actions Could Enhance Customer Service*, GAO-04-107 (Washington, D.C.: Nov. 20, 2003).

Illustration of an Abbreviated Survey Description

Brief publications, such as typical testimonies or reports that make minimal use of a survey could meet survey reporting needs with a statement such as this:

“On the basis of 1,203 responses to our mail survey of a stratified probability sample, we estimate that approximately xx percent of the ##### nurses at Veterans Administration hospitals in the contiguous United States feel they are ‘satisfied’ (‘moderately satisfied’ or ‘very satisfied’ on a five-point scale) when asked ‘How satisfied are you with the care that the local VA hospital provides to disabled outpatients . . . ?’ Our survey achieved an 87 percent response rate and provides results that are accurate to within about plus or minus 4 percentage points (95 percent level of confidence). Our previous report (GAO-00-000) describes the survey in more detail.”