Proposed Standards and Guidelines for Statistical Surveys

Introduction

Section 1 Development of Concepts, Methods, and Design

1.1 Survey Planning

1.2 Survey Design

1.3 Survey Response Rates

1.4 Pretesting Survey Systems

Section 2 Collection of Data

2.1 Developing Sampling Frames

2.2 Required Notifications to Potential Survey Respondents

2.3 Data Collection Methodology

Section 3 Processing and Editing of Data

3.1 Data Editing

3.2 Nonresponse Analysis and Response Rate Calculation

3.3 Coding

3.4 Data Protection

3.5 Evaluation

Section 4 Production of Estimates and Projections

4.1 Developing Estimates and Projections

Section 5 Data Analysis

5.1 Analysis and Report Planning

5.2 Inference and Comparisons

Section 6 Review Procedures

6.1 Review of Information Products

Section 7 Data Dissemination

7.1 Releasing Information

7.2 Data Protection and Disclosure Avoidance for Dissemination

7.3 Survey Documentation

7.4 Documentation and Release of Public-Use Microdata

Glossary

Introduction

These standards apply to Federal statistical surveys whose purposes include the description, estimation, or analysis of the characteristics of groups, segments, activities, or geographic areas. The development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support those purposes are covered by these standards.

Government policy establishing standards for Federal statistical programs serves both the interests of the public and the needs of the government. Standards, and the accompanying guidelines presenting recommended best practices to fulfill the goals of the standards, document the professional basis upon which a Federal agency expects to be judged by its stakeholders and the level of quality and effort expected in all statistical activities. Standards and guidelines provide a means to ensure consistency among and within statistical activities conducted across the Federal Government. Finally, standards and guidelines ensure that users of Federal statistical

information products are provided with details on the methods and principles employed in the development, collection, processing, analysis, and dissemination of statistical information.

Some of these standards also apply to the compilation of statistics based on information collected from individuals or firms (such as tax returns or the financial and operating reports required by regulatory commissions), applications/registrations, or administrative records. These standards have only limited relevance for information collections carried out by agencies in administering specific programs (such as summaries for production status, or fiscal accounting reports from the States or grant-in-aid programs).

In 2002, the U.S. Office of Management and Budget (OMB), in response to Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554), issued government-wide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." Federal statistical agencies worked together to draft a common framework to use in developing their individual Information Quality Guidelines. That framework, published in the June 4, 2002 *Federal Register* Notice, "Federal Statistical Organizations' Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Disseminated Information" (67 FR 38467-38470), serves as the organizing framework for the standards and guidelines presented here.[1] The framework: for these standards and guidelines includes

[1] The *Federal Register* notice included eight areas where statistical organizations set standards for performance. The framework utilized here combines "Development of concepts and methods" with "Planning and design of surveys and other means of collecting data" into the single section on "Development of concepts, methods, and design." The standards for these activities were closely linked and attempting to separate them into two separate

- Development of concepts, methods, and design
- Collection of data
- Processing and editing of data
- Production of estimates and projections
- Data analysis
- Review procedures
- Data dissemination.

Within this framework, there are 20 standards designed for general application to Federal statistical survey activities. Each standard is accompanied by guidelines that present best practices useful in fulfilling the goals of the standard. These standards and guidelines for Federal statistical surveys support agencies in achieving the Information Quality Guidelines' requirements for ensuring and maximizing the utility, objectivity, and integrity of information disseminated.

It should be recognized that these standards focus on ensuring high quality statistical surveys that result in information products satisfying an agency's and OMB's Information Quality Guidelines. The standards and guidelines are meant to provide general guidance and are not intended to substitute for the extensive existing literature on statistical and survey theory, methods, and operations. When undertaking a survey, an agency should engage knowledgeable and experienced survey practitioners to effectively achieve the goals of the standards. Persons

---

sections would have resulted in some duplication of standards between sections. The only other change is the title of Section 7, which was shortened to "Data Dissemination" for convenience rather than "Dissemination of data by published reports, electronic files, and other media requested by users" as it originally appeared in the *Federal Register* notice.

involved should have knowledge and experience in survey sampling theory, survey methodology, field operations, data analysis and technological aspects of surveys.

Under the OMB Information Quality Guidelines, quality is an encompassing term comprising utility, objectivity, and integrity.

> **Utility** refers to the usefulness of the information to its intended users. The usefulness of information disseminated by Federal agencies should be considered from the perspective of subject matter specific users, researchers, policymakers, and the public. Utility is achieved by continual assessment of information needs, anticipating emerging requirements, and developing new products and services.

To ensure that information disseminated by Federal agencies meets the needs of the intended users, agencies rely upon internal reviews, analyses, and evaluations along with feedback from advisory committees, researchers, policymakers, and the public. In addition, all information products should be clearly and correctly presented in plain language. The target audience for each product should be clearly identified, and the product's contents should be readily accessible to that audience.

In all cases, the goal is to maximize the usefulness of information and minimize the costs to the government and the public. When disseminating its information products, Federal agencies should utilize a variety of efficient dissemination channels so that the public, researchers, and policymakers can locate information in an equitable and timely fashion.

The specific standards that contribute directly to the utility and the dissemination of information include those on Survey Planning (1.1), Survey Design (1.2), Pretesting Survey Systems (1.4), Review of Information Products (6.1), Releasing Information (7.1), Survey Documentation (7.3), and Documentation and Release of Public-Use Microdata (7.4).

> **Objectivity** refers to whether information is accurate, reliable, and unbiased, and is presented in an accurate, clear, and unbiased manner. It involves both the content of the information and the presentation of the information. This includes complete, accurate, and easily understood documentation of the source of the information, with a description of the sources of any errors that may affect the quality of the data, when appropriate. Objectivity is achieved by using reliable information sources and appropriate techniques to prepare information products.

Standards relating to the production of accurate, reliable, and unbiased information include the data collection standards—Survey Response Rates (1.3), Developing Sampling Frames (2.1), Required Notifications to Potential Survey Respondents (2.2), and Data Collection Methodology (2.3); the standards on processing and editing data—Data Editing (3.1), Nonresponse Analysis and Response Rate Calculation (3.2), Coding (3.3), and Evaluation (3.5); the standard on Developing Estimates and Projections (4.1); and standards on data analysis—Analysis and Report Planning (5.1) and Inference and Comparisons (5.2).

There is also a set of standards related to presenting results in an accurate, clear, and unbiased manner; included in this group are standards on Review of Information Products (6.1) and on reporting and documentation — Survey Documentation (7.3) and Documentation and Release of Public-Use Microdata (7.4).

> **Integrity** refers to the security or protection of information from unauthorized access or revision. Integrity ensures that the information is not compromised through corruption or falsification.

Federal agencies have a number of statutory and administrative provisions governing the protection of information. Examples that may affect all Federal agencies include the Privacy Act; the Computer Security Act of 1987; the Freedom of Information Act; OMB Circular Nos. A-123, A-127, and A-130; the Federal Policy for the Protection of Human Subjects; the Government Information Security Reform Act; and the E-Government Act of 2002. The standards on Required Notifications to Potential Survey Respondents (2.2), Data Protection (3.4) during processing and editing, and Data Protection and Disclosure Avoidance for Dissemination (7.2) directly address statistical issues concerning the integrity of data.

The application of standards to the wide range of Federal statistical activities and uses should not be a mechanical process. Rather, it requires judgment that balances such factors as the uses of the resulting information and the efficient allocation of resources. Some surveys are extremely large undertakings requiring millions of dollars, and the resulting general-purpose statistics have significant, far-reaching effects. (Examples of major Federal information programs, many based

on statistical surveys, are the Principal Economic Indicators.[2]) Other statistical activities may be more limited and focused on specific program areas (e.g., customer satisfaction surveys, program evaluations, and research).

For each statistical survey in existence when these standards are issued and for each new survey, the sponsoring agency should evaluate compliance with applicable standards. The agency should establish compliance goals for applicable standards if a survey is not in compliance. An agency should use major survey revisions or other significant survey events as opportunities to address areas in which a survey is not in compliance with applicable standards.

Federal agencies are encouraged to adhere to all standards for every statistical survey. However, standards cannot be applied uniformly or precisely in every situation. Consideration must be given to the importance of the uses of the information as well as the quality required to support those uses. The provision of standards and guidelines cannot substitute for agency judgment about the most appropriate expenditure of funds. If funding or other contingencies make it impossible for all standards to be met, agencies should evaluate the potential improvement in data quality that would result from adherence to the standards and discuss what options were considered and why the final design was selected.

In instances where the strict application of a standard is impractical or infeasible, the agency should consider alternative methods of achieving the standard's purpose. The agency should include in the standard documentation for the survey, or in an easily accessible public venue, the

[2] For the list of principal economic indicators and their release dates see http://www.whitehouse.gov/omb/inforeg/statpolicy.html#sr

reasons why the standard could not be met and what actions the agency has taken or will take to address any resulting issues.[3]

The following standards and guidelines are not designed to be completely exhaustive of all efforts that an agency may undertake to ensure the quality of its statistical information. Agencies are encouraged to develop additional, more detailed standards focused on their specific statistical activities.

The standards are presented in seven Sections. For each standard, a list of key words for relevant concepts that are defined in the accompanying glossary and guidelines that represent best practices that may be useful in fulfilling the goals of the standard are presented. Agencies conducting surveys should also consult draft guidance issued by OMB entitled *Questions and Answers When Designing Surveys for Information Collections.* That document was developed by OMB to assist agencies in preparing their Information Collection Requests for OMB review under the Paperwork Reduction Act (PRA). The PRA requires that all Federal agencies obtain approval from OMB prior to collecting information from the public.

SECTION 1 DEVELOPMENT OF CONCEPTS, METHODS, AND DESIGN

Section 1.1 Survey Planning

Standard 1.1: Agencies initiating a new survey or major revision of an existing survey must

[3] In cases where the agency determines that ongoing surveys are not in compliance with the standards, the documentation should be updated at the earliest possible time.

develop a written plan that sets forth a justification, including goals and objectives, potential users, and the decisions the survey may inform; key survey estimates; the precision required of the estimates (e.g., the size of differences that need to be detected); the tabulations and analytic results that will inform decisions and other uses; related and previous surveys; when and how frequently users need the data; and the level of detail needed in tabulations, confidential microdata, and public-use data files.

Key Terms: bridge study, confidentiality, consistent data series, crosswalk study, data series, effect size, individually identifiable data, key variables, measurement error, microdata, minimum substantively significant effect (MSSE), pretest, public-use data file, respondent burden, survey system

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 1.1.1: Surveys (and related activities such as focus groups, cognitive interviews, pilot studies, field tests, etc.) are collections of information subject to the requirements of the Paperwork Reduction Act of 1995 (P.L. 104-13, 44 U.S.C. 3501 et seq.) and OMB's implementing regulations (5 CFR Part 1320, Controlling Paperwork Burdens on the Public). An initial step in planning a new survey or a revision of an existing survey should be to contact the sponsoring agency's Chief Information Officer or other designated official to ensure the survey work is done in compliance with the law and regulations. OMB approval will be required before the agency may collect information from 10 or more persons in a 12-month period. A useful

reference document regarding the approval process is OMB's *Questions and Answers When Designing Surveys for Information Collections*.

Guideline 1.1.2: Planning is an important prerequisite when designing a new survey or survey system, or implementing a major revision of an ongoing survey. Key planning and project management activities include the following:

1. A justification for the survey, including the rationale for the survey, relationship to prior surveys, survey goals and objectives, hypotheses to be tested, and definitions of key variables. Consultations with potential users to identify their requirements and expectations are also important at this stage of the planning process.
2. A review of related studies, surveys, and reports of Federal and non-Federal sources to ensure that part or all of the data are not available from an existing source, or could not be more appropriately obtained by adding questions to existing Federal statistical surveys. The goal here is to spend Federal funds effectively and minimize respondent burden. If a new survey is needed, efforts to minimize the burden on individual respondents are important in the development of new items.
3. A review of the confidentiality and privacy provisions of the Privacy Act, OMB Guidance for Implementing the Confidential Information Protection and Statistical Efficiency Act of 2002 (forthcoming), and OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002, and all other relevant laws, regulations, and guidance when planning any surveys that will collect individually-identifiable data from any survey participant.
4. A review of all survey data items, the justification for each item, and how each item can best be measured (e.g., through questionnaires, tests). Agencies should assemble reasonable evidence

that these items are valid and can be measured both accurately and reliably, or develop a plan for testing these items to assess their reliability and validity.

5. A plan for pretesting the survey or survey system, if applicable (see Section 1.4).
6. A plan for quality assurance during each phase of the survey process to permit monitoring and assessing performance during implementation. The plan should include contingencies to modify the survey procedures if design parameters appear unlikely to meet expectations (for example, if low response rates are likely). The plan should also contain general specifications for an internal project management system that identifies critical activities and key milestones of the survey that will be monitored, and the time relationships among them.
7. A plan for evaluating survey procedures, results, and measurement error (see Section 3.5).
8. An analysis plan that identifies analysis issues, objectives, key variables, minimum substantively significant effect sizes, and proposed statistical tests.
9. An estimate of resources and target completion dates needed for the survey cycle.
10. A dissemination plan that identifies target audiences, proposed major information products, and the timing of their release.

Guideline 1.1.3: Consistent data collection procedures for ongoing data collections help maintain a consistent data series over time.

Guideline 1.1.4: Continuous improvement efforts sometimes result in a trade-off between the desire for consistency and a need to improve a data collection. If changes are needed in key variables or survey procedures for a data series, consider the justification or rationale for the changes in terms of their usefulness for policymakers, conducting analyses, and addressing

information needs. Develop adjustment methods, such as crosswalks and bridge studies that will be used to preserve trend analyses and inform users about the effects of changes.

Section 1.2 Survey Design

Standard 1.2: Agencies must develop a survey design, including defining the target population, designing the sampling plan, specifying the data collection instrument and methods, developing a realistic timetable and cost estimate, and selecting samples using generally accepted statistical methods (e.g., probabilistic methods that can provide estimates of sampling error). Any use of nonprobability sampling methods (e.g., cut-off or model-based samples) must be justified statistically and be able to measure estimation error. The size and design of the sample must reflect the level of detail needed in tabulations and other data products, and the precision required of key estimates. Documentation of each of these activities and resulting decisions must be maintained in the project files for use in documentation (see Standards 7.3 and 7.4).

Key Terms: bias, confidentiality, cut-off sample, domain, effective sample size, estimation error, frame, imputation, key variables, model-based sample, nonprobabilistic methods, nonsampling error, power, precision, probabilistic methods, probability of selection, response rate, sampling error, sampling unit, strata, target population, total mean square error, variance

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 1.2.1: The survey design should include the proposed target population, response rate goals, frequency and timing of collection, data collection methods, sample design, sample size, precision requirements, and, where applicable, an effective sample size determination based on power analyses for key variables.

Guideline 1.2.2: The sample design should ensure the sample will yield the data required to meet the objectives of the survey. The sample design should include the following: identification of the sampling frame and the adequacy of the frame; the sampling unit used (at each stage if a multistage design); sampling strata; power analyses to determine sample sizes and effective sample sizes for key variables by reporting domains (where appropriate); criteria for stratifying or clustering, sample size by stratum, and the known probabilities of selection; expected yield by stratum; estimated efficiency of sample design; response rate goals (see Standard 1.3); estimation and weighting plan; variance estimation techniques appropriate to the survey design; and expected precision of estimates for key variables.

Guideline 1.2.3: When a nonprobabilistic sampling method is employed, the survey design documentation should include a discussion of what options were considered and why the final design was selected, an estimate of the potential bias in the estimates, and the methodology to be used to measure estimation error. The documentation should detail the selection process and demonstrate that units not in the sample are impartially excluded on objective grounds.

Guideline 1.2.4: The survey instrument should include a pledge of confidentiality, if applicable, along with instructions required to complete the survey. A clear, logical, and easy-to-follow

flow of questions is a key element of a successful survey.

Guideline 1.2.5: The anticipated data collection plans should include frequency and timing of data collections; methods of collection for achieving acceptable response rates; training of enumerators and persons coding and editing the data; and cost estimates, including the costs of pretests, nonresponse follow-up, and evaluation studies.

Guideline 1.2.6: Whenever possible, total mean square error should be constructed in approximate terms, and accuracy should be evaluated by comparing with other information sources. If probability sampling is used, sampling error should be estimated. If nonprobability sampling is used, the estimation error should be computed.

Guideline 1.2.7: Potential nonsampling errors including measurement errors due to interviewers, respondents, instruments, and mode; nonresponse error; coverage error; and processing error should be carefully considered, and when possible, the effects estimated.

Section 1.3 Survey Response Rates

Standard 1.3: Agencies must design the survey to achieve the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data collection costs, to ensure that survey results are representative of the target population so that they can be used with confidence to inform decisions. Nonresponse bias analyses must be conducted when unit or item response rates suggest the potential for bias to occur.

Key Terms: cross-sectional, key variables, longitudinal, nonresponse bias, response rates, stage of data collection, substitution, target population, universe

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 1.3.1: Calculate sample survey unit response rates without substitutions.

Guideline 1.3.2: Design data collections that will be used for sample frames for other surveys (e.g., the Decennial Census, and the Common Core of Data collection by the National Center for Education Statistics) to meet a target unit response rate of at least 95 percent, or provide a justification for a lower anticipated rate.

Guideline 1.3.3: Prior to data collection, identify expected unit response rates at each stage of data collection, based on content, use, mode, and type of survey.

Guideline 1.3.4: Plan for a nonresponse bias analysis if the expected unit response rate is below 80 percent.

Guideline 1.3.5: Plan for a nonresponse bias analysis if the expected item response rate is below 70 percent for any items used in a report.

Section 1.4 Pretesting Survey Systems

Standard 1.4: Agencies must pretest all components of a survey to ensure that measurement error is controlled and that the components function as intended when implemented in the full-scale survey.

Key Terms: cognitive interview, edit, estimation, field test, focus group, frame, imputation, instrument, pilot test, pretest, response analysis survey, response rates, stage of data collection, survey system

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 1.4.1: Use cognitive testing and usability testing to test new components of a survey system prior to a field test of the survey system (for example, focus groups and cognitive interviews). Results from these tests should be incorporated into the final design.

Guideline 1.4.2: Use field tests prior to implementation of the full-scale survey when some or all components of a survey system cannot be successfully demonstrated through previous work. The design of a field test should reflect realistic conditions, including those likely to pose difficulties for the survey. Elements to be tested include, for example, frame development, sample selection, questionnaire design, data collection, item feasibility, electronic data collection capabilities, edit specifications, data processing, estimation, file creation, and tabulations. A

complete test of all components (sometimes referred to as a dress rehearsal) may be desirable for highly influential surveys.

SECTION 2 COLLECTION OF DATA

Section 2.1 Developing Sampling Frames

Standard 2.1: Agencies must ensure that the frames for the planned sample survey or census are appropriate for the study design and are evaluated against the target population for quality.

Key Terms: bias, coverage, estimation, frame, frame populations, target populations

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 2.1.1: Describe target populations and associated survey or sampling frames. Include the following items in this description:

1. The manner in which the frame was constructed;
2. Any exclusions that have been applied to target and frame populations;
3. Coverage issues such as alternative frames that were considered, coverage rates (an estimation of the missing units on the frame), multiple coverage rates if some addresses target multiple populations (such as schools and children or households and individuals),

what was done to improve the coverage of the frame, and how data quality and item nonresponse on the frame may have affected the coverage of the frame;

4. Any estimation techniques used to improve the coverage of estimates such as post-stratification procedures; and
5. Other limitations of the frame including the timeliness of the frame.

Guideline 2.1.2: Conduct periodic evaluations of coverage rates and coverage of the target population in survey frames that are used for recurring surveys, for example, at least every 5 years.

Guideline 2.1.3: Coverage rates in excess of 95 percent overall and for each major stratum are desirable. If coverage rates fall below 85 percent, conduct an evaluation of the potential bias.

Guideline 2.1.4: Consider using frame enhancements, such as frame supplementation or dual-frame estimation, to increase coverage.

For more information on developing survey frames, see *FCSM Statistical Policy Working Paper 17, Survey Coverage*.

Section 2.2 Required Notifications to Potential Survey Respondents

Standard 2.2: Agencies must ensure that each collection of information instrument clearly states the reasons the information is planned to be collected; the way such information is planned

to be used to further the proper performance of the functions of the agency; whether responses to the collection of information are voluntary or mandatory (citing authority); the nature and extent of confidentiality to be provided, if any, citing authority; an estimate of the average respondent burden together with a request that the public direct to the agency any comments concerning the accuracy of this burden estimate and any suggestion for reducing this burden; and a statement that an agency may not conduct and a person is not required to respond to an information collection request unless it displays a currently valid OMB control number.

Key Terms: confidentiality, mandatory, respondent burden, voluntary

---

The following guideline represents best practices that may be useful in fulfilling the goals of the standard:

Guideline 2.2.1: Depending on respondent burden as well as the scope and nature of questions to be asked, agencies should provide adequate informational materials to respondents. These may include a pre-notification letter, brochure, set of questions and answers, or an 800 number to call that does the following:

1. Informs potential respondents that they have been selected to participate in a survey;
2. Informs potential respondents about the name and nature of the survey; and
3. Provides any additional information to potential respondents that the agency is required to supply (e.g., see 5 CFR 1320.8(b)(3)).

Section 2.3 Data Collection Methodology

Standard 2.3: Agencies must design and administer their data collection instruments and methods in a manner that achieves the best balance between maximizing data quality and controlling measurement error while minimizing respondent burden and cost.

Key Terms: imputation, item nonresponse, nonresponse bias, required response item, respondent burden, response analysis survey, response rates, target population, validation studies

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 2.3.1: Encourage respondents to participate to maximize response rates and improve data quality. This can be accomplished by training interviewers and other staff who may have contact with respondents in techniques for obtaining respondent cooperation and building rapport with respondents. Techniques for building rapport include respect for respondents' rights, follow-up skills, knowledge of the goals and objectives of the data collection, and knowledge of the uses of the data.

Guideline 2.3.2: Design the data collection instrument in a manner that minimizes respondent burden, while maximizing data quality. The following strategies may be used to achieve these goals:

1. Questions are clearly written and skip patterns easily followed;
2. The questionnaire is of reasonable length;

3. The questionnaire is pretested to identify problems with interpretability and ease in navigation; and
4. Methods to reduce item nonresponse are adopted.

Guideline 2.3.3: Since high response rates improve data quality through ensuring that results are representative of the target population, the following data collection strategies can also be used to achieve high response rates:

1. Ensure that the data collection period is of adequate and reasonable length;
2. Send materials describing the data collection to respondents in advance, when possible;
3. Plan an adequate number of contact attempts; and
4. When appropriate, consider use of respondent incentives.

Guideline 2.3.4: The way a data collection is designed and administered also contributes to data quality. The following issues are important to consider:

1. Given the characteristics of the target population, the objectives of the data collection, the resources available, and time constraints, determine the appropriateness of the method of data collection (e.g., mail, telephone, Internet);
2. Collect data at the most appropriate time of year, where relevant;
3. Establish the data collection protocol to be followed by the field staff;
4. Provide training for field staff on new protocols, with refresher training on a routine, recurring cycle;
5. Establish best practice mechanisms to minimize interviewer falsification, such as protocols for monitoring interviewers and reinterviewing respondents;

6. Conduct response analysis surveys or other validation studies for new data collection efforts that have not been validated;
7. Establish protocols that minimize measurement error, such as conducting response analysis surveys to ensure records exist for data elements requested for business surveys, establishing recall periods that are reasonable for demographic surveys, and developing computer systems to ensure Internet data collections function properly; and
8. Quantify nonsampling errors to the extent possible.

Guideline 2.3.5: Develop protocols to monitor data collection activities, with strategies to correct identified problems. The following issues are important to consider:

1. Implement quality and performance measurement and process control systems to monitor data collection activities and integrate them into the data collection process. These processes, systems, and tools will provide timely measurement and reporting of all critical components of the data collection process, on the dimensions of progress, response, quality, and cost, and enable managers to identify and resolve problems and ensure that data collection is completed successfully. Additionally, these measurements will provide survey designers and data users with indicators of survey performance and resultant data quality.
2. Use internal reporting systems that provide timely reporting of response rates and the reasons for nonresponse throughout the data collection. These systems should be flexible enough to identify important subgroups with low response rates for more intensive follow-ups.
3. If response rates are low and it is impossible to conduct more extensive procedures for the full sample, select a random subsample of nonrespondents for the more intensive data collection method. This subsample permits a description of nonrespondents' characteristics,

provides data needed for nonresponse bias analysis, and allows for possible weight adjustments or for imputation of missing characteristics.

4. Determine a set of required response items to obtain when a respondent is unwilling to fully cooperate. These items may then be targeted in follow-up to meet the minimum standard for unit response. These items may also be used in a nonresponse bias analysis that compares characteristics of respondents and nonrespondents using the sample data for those items. These required response items may also be used for item nonresponse imputation systems.

SECTION 3 PROCESSING AND EDITING OF DATA

Section 3.1 Data Editing

Standard 3.1: Agencies must edit data appropriately, based on available information, to mitigate or correct detectable errors.

Key Terms: editing

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 3.1.1: Check and edit data to mitigate errors. Data editing is an iterative and interactive process that includes procedures for detecting and correcting errors in the data. Editing uses available information and some assumptions to derive substitute values for

inconsistent values in a data file. When electronic data collection methods are used, data are usually edited both during and after data collection. As appropriate, check data for the following and edit if errors are detected:

1. Responses that fall outside a prespecified range (e.g., based on expert judgment or previous responses) or, for categorical responses, are not equal to specified categories;
2. Consistency, such as the sum of categories matches the reported total, or responses to different questions are logical;
3. Contradictory responses and incorrect flow through prescribed skip patterns;
4. Missing data that can be directly filled from other portions of the same record (including the sample frame);
5. The omission and duplication of records; and
6. Inconsistency between estimates and outside sources.

Guideline 3.1.2: Possible actions for failed edits include the following:

1. Automated correction within specified criteria;
2. Data verified by respondent, and edit overridden;
3. Corrected data provided by respondent;
4. Corrected data available from other sources;
5. If unable to contact respondent, and after review by survey staff, an imputed value may be substituted for a failed edit; and
6. Data edit failure overridden after review by survey staff.

Guideline 3.1.3: Code the data set to indicate any actions taken during editing, and/or retain the unedited data along with the edited data.

For more information on data editing, see *FCSM Statistical Policy Working Paper 18, Data Editing in Federal Statistical Agencies*, and *FCSM Statistical Policy Working Paper 25, Data Editing Workshop and Exposition*.

Section 3.2 Nonresponse Analysis and Response Rate Calculation

Standard 3.2: Agencies must appropriately measure, adjust for, report, and analyze unit and item nonresponse to assess their effects on data quality and to inform users. Response rates must be computed using standard formulas to measure the proportion of the eligible sample that is represented by the responding units in each study, as an indicator of potential nonresponse bias.

Key Terms: bias, cross-wave imputation, cross-sectional, eligible sample unit, frame, imputation, item nonresponse, key variables, longitudinal, longitudinal analysis, missing at random, missing completely at random, multivariate analysis, multivariate modeling, nonresponse bias, overall unit nonresponse, probability of selection, response rates, stages of data collection, unit nonresponse, wave, weights

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 3.2.1: Calculate all response rates, unweighted and weighted, based either on the probability of selection or, in the case of establishment surveys, on the proportion the establishment represents of the total industry.

Guideline 3.2.2: Calculate unit response rates (RRU) as the ratio of the number of completed interviews (I) to the number of in-scope sample cases (AAPOR, 2004). There are a number of different categories of cases that comprise the total number of in-scope cases:

I = number of completed interviews;

R = number of refused interview cases;

O = number of eligible sample units not responding for reasons other than refusal;

NC = number of noncontacted sample units known to be eligible;

U = number of sample units of unknown eligibility, with no interview; and

e = estimated proportion of sample units of unknown eligibility that are eligible.

The unit response rate represents a composite of these components:

$$RRU = \frac{I}{I + R + O + NC + e(U)}$$

Guideline 3.2.3: Calculate the overall unit response rates for cross-sectional analysis (RRO^C) as the product of two or more unit-level response rates when a survey has multiple stages:

$$RRO^C = \prod_{i=1}^{K} RRU_i$$

(It is the product, not the sum)

where K = the number of stages and C denotes cross-sectional.

When sample is drawn with probability proportionate to size (PPS), then the interpretation of RRO^C can be improved by weighting the $k_1 \ldots k_{K-1}$ stages by their size. This is especially helpful if nonresponse is related to the size of the sample units.

Guideline 3.2.4: Calculate longitudinal response rates for each wave. Use special procedures for longitudinal surveys where previous nonrespondents are eligible for inclusion in subsequent waves. The overall unit response rate used in longitudinal analysis (RRO^L) reflects the proportion of all eligible respondents in the sample who participated in all waves in the analysis, and includes the response rates from all stages of data collection used in the analysis:

$$RRO^L = \prod_{k=1}^{K} \frac{I_k^L}{I_k^1 + R_k^1 + O_k^1 + NC_k^1 + e_k(U_k^1)}$$

where:

K = the last stage of data collection used in the analysis;

I^L = the number of responding cases common to all waves in the analysis

R^1_k = Refusals at wave 1 at stage k

so that $I^1_k + R^1_k + O^1_k + NC^1_k + e_k(U^1_k)$ is the entire sample entered at wave 1

Guideline 3.2.5: Calculate item response rates (RRI) as the ratio of the number of respondents for whom an in-scope response was obtained (I^x for item x) to the number of respondents who were asked to answer that item. The number asked to answer an item is the number of unit-level respondents (I) minus the number of respondents with a valid skip for item x (V^x). When an abbreviated questionnaire is used to convert refusals, the eliminated questions are treated as item nonresponse:

$$RRI^{x} = \frac{I^{x}}{I - V^{x}}$$

Guideline 3.2.6: Calculate the total item response rates (RRT^{x}) for specific items as the product of the overall unit response rate (RRO) and the item response rate for item x (RRI^{x}):

$$RRT^{x} = RRO * RRI^{x}$$

Guideline 3.2.7: When calculating a response rate with supplemented samples, base the reported response rates on the original and the added sample cases. However, when calculating response rates where the sample was supplemented using matched pairs selected during the initial sample selection, calculate unit response rates without the substituted cases included (i.e., only the original cases are used).

Guideline 3.2.8: Given a survey with an overall unit response rate of less than 80 percent, conduct an analysis of nonresponse using unit response rates as defined above, with an assessment of whether the data are missing completely at random. For a multistage (or wave) survey, focus the analysis on each stage, with particular attention to the "problem" stages. Make comparisons between respondents and nonrespondents across subgroups using available sample frame variables. In the analysis of unit nonresponse, consider a multivariate modeling of response using respondent and nonrespondent frame variables to determine if nonresponse bias exists. Comparison of the respondents to known characteristics of the population from an external source can provide an indication of possible bias, especially if the characteristics in question are related to the survey's key variables.

Guideline 3.2.9: If the item response rate is less than 70 percent, conduct an item nonresponse analysis to determine if the data are missing at random at the item level for at least the items in question, in a manner similar to that discussed in Guideline 3.2.8.

Guideline 3.2.10: In those cases where the analysis indicates that the data are not missing at random, the amount of potential bias should inform the decision to publish individual items.

Guideline 3.2.11: For data collections involving sampling, adjust weights for unit nonresponse, unless unit imputation is warranted. The unit nonresponse adjustment should be internally consistent, based on theoretical and empirical considerations, appropriate for the analysis, and make use of the most relevant data available.

Guideline 3.2.12: Base decisions regarding whether or not to adjust or impute data for item nonresponse on how the data will be used, the assessment of nonresponse bias that is likely to be encountered in the review of collections, prior experience with this collection, and the nonresponse analysis discussed in this Section. When used, imputation and adjustment procedures should be internally consistent, based on theoretical and empirical considerations, appropriate for the analysis, and make use of the most relevant data available. If multivariate analysis is anticipated, care should be taken to use imputations that minimize the attenuation of underlying relationships.

Guideline 3.2.13: In the case of imputing longitudinal data sets, use cross-wave imputations or cross-sectional imputations.

Guideline 3.2.14: Clearly identify all imputed values on a data file (e.g., code them).

Section 3.3 Coding

Standard 3.3: Agencies must add codes to collected data to identify aspects of data quality from the collection (e.g., missing data) in order to allow users to appropriately analyze the data. Codes added to convert information collected as text into a form that permits immediate analysis must use standardized codes, when available, to enhance comparability.

Key Terms: coding, quality assurance process

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 3.3.1: Insert codes into the data set that clearly identify missing data and cases where an entry is not expected (e.g., skipped over by skip pattern). Do not use blanks and zeros as codes to identify missing data, as they tend to be confused with actual data.

Guideline 3.3.2: When converting text data to codes to facilitate easier analysis, use standardized codes, if they exist. Use the Federal coding standards listed below, if applicable. Provide cross-referencing tables to the Federal standard codes for any legacy coding that does not meet the Federal standards. Develop other types of codes using existing Federal agency

practice or standard codes from industry or international organizations, when they exist. Current Federal standard codes include the following:

1. FIPS Codes. The National Institute of Standards and Technology maintains Federal Information Processing Standards (FIPS) required for use in Federal information processing in accordance with OMB Circular No. A-130. The following FIPS should be used for coding (see www.itl.nist.gov/fipspubs/index.htm for the most recent versions of these standards):

   5-2 Codes for the Identification of the States, the District of Columbia and the Outlying Areas of the United States, and Associated Areas

   6-4 Counties and Equivalent Entities of the United States, Its Possessions, and Associated Areas

   9-1 Congressional Districts of the United States

   10-4 Countries, Dependencies, Areas of Special Sovereignty and Their Principal Administrative Divisions

2. NAICS Codes. The North American Industry Classification System (NAICS) should be used to classify establishments. NAICS was developed jointly by the United States, Canada, and Mexico to provide new comparability in statistics about business activity across North America. NAICS coding has replaced the U.S. Standard Industrial Classification (SIC) system (for more information, see www.census.gov/epcd/www/naics.html).

3. SOC Codes. The Standard Occupational Classification (SOC) system should be used to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data (for more information, see www.bls.gov/soc).

4. Race and Ethnicity. Classification of race and ethnicity, as well as methods of collection, should comply with OMB's Standards for Maintaining, Collecting, and Presenting Federal

Data on Race and Ethnicity (for more information, see www.whitehouse.gov/omb/inforeg/statpolicy.html).

5. Statistical Areas. The Standards for Defining Metropolitan and Micropolitan Statistical Areas should be used for collecting, tabulating, and publishing Federal statistics for geographic areas (for more information, see www.whitehouse.gov/omb/inforeg/statpolicy.html).

Guideline 3.3.3: When setting up a manual coding process to convert text to codes, create a quality assurance process that verifies at least a sample of the coding to determine if a specific level of coding accuracy is being maintained.

Section 3.4 Data Protection

Standard 3.4: Agencies must implement safeguards throughout the production process to ensure that survey data are handled to avoid disclosure.

Key Terms: confidential, individually-identifiable

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 3.4.1: For surveys that include confidential data, establish procedures and mechanisms to ensure the information's protection during the production, use, storage,

transmittal, and disposition of the survey data in any format (e.g., completed survey forms, electronic files, and printouts).

Guideline 3.4.2: Ensure that

1. Individually-identifiable survey data are protected;
2. Data systems and electronic products are protected from unwarranted intervention; and
3. Data files, network segments, servers, and desktop PCs are electronically secure from malicious software and intrusion using best available information resource security practices, which are periodically monitored and updated.

Guideline 3.4.3: Ensure controlled access to data sets so that only specific, named individuals working on a particular data set can either read or write, or both read and write, that data set. Data set access rights are to be periodically reviewed by the project manager responsible for that data set in order to guard against unauthorized release or alteration.

For more information on data protection, see *FCSM Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology*, and forthcoming OMB guidance on implementation of the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).

Section 3.5 Evaluation

Standard 3.5: Agencies must evaluate the quality of the data and make the evaluation public (through technical notes and documentation included in reports of results or through a separate report) to allow users to interpret results of analyses, and to help designers of recurring surveys focus improvement efforts.

Key Terms: coverage error, instrument, item nonresponse, measurement error, nonresponse error, nonsampling error, sampling error, weights

---

The following guideline represents best practices that may be useful in fulfilling the goals of the standard:

Guideline 3.5.1: Include an evaluation component in the survey plan that evaluates survey procedures, results, and measurement error (see Section 1.1). Review past surveys similar to the one being planned to determine likely sources of error, appropriate evaluation methods, and problems that are likely to be encountered. Address the following areas:

1. Potential sources of error, including
    - Coverage error (including frame errors);
    - Nonresponse error; and
    - Measurement error, including sources from the instrument, interviewers, and collection process;
2. Data processing error (e.g., keying, coding, editing, and imputation error);
3. How sampling and nonsampling error will be measured, including variance estimation and studies to isolate error components;

4. How total mean square error will be assessed;
5. Methods used to reduce nonsampling error in the collected data;
6. Methods used to mitigate nonsampling error after collection;
7. Post-collection analyses of the quality of final estimates; the data and estimates derived from the data should be compared to other independent collections of similar data, if available; and
8. Make evaluation studies public to inform data users.

For more information on evaluations, see *FCSM Statistical Policy Working Paper 15, Measurement of Quality in Establishment Surveys,* and *FCSM Statistical Policy Working Paper 31, Measuring and Reporting Sources of Error in Surveys*.

SECTION 4 PRODUCTION OF ESTIMATES AND PROJECTIONS

Section 4.1 Developing Estimates and Projections

Standard 4.1: Agencies must use accepted theory and methods when deriving direct survey-based estimates, as well as model-based estimates and projections that use survey data. Error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections. Agencies must plan and implement evaluations to assess the quality of the estimates and projections.

Key Terms: design effect, direct survey-based estimates, estimation, model, model-based estimate, model validation, population, post-stratification, projection, raking, ratio estimation, sensitivity analysis, strata, variance, weights

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 4.1.1: Develop direct survey estimates according to the following practices:

1. When data from a sample survey are used to calculate population estimates, employ weights appropriate for the sample design. However, an agency may employ an alternative method (e.g., ratio estimators) to calculate population estimates if the agency has evaluated the alternative method and determined that it leads to acceptable results.
2. Use auxiliary data to improve precision and/or reduce the error associated with direct survey estimates.
3. Calculate variance estimates by a method appropriate to a survey's sample design taking into account probabilities of selection, stratification, clustering, and the effects of nonresponse, post-stratification, and raking. The estimates must reflect any design effect resulting from a complex design.

Guideline 4.1.2: Develop model-based estimates according to accepted theory and practices (e.g., assumptions, mathematical specifications).

Guideline 4.1.3: Develop projections in accordance with accepted theory and practices (e.g., assumptions, mathematical specifications).

Guideline 4.1.4: Any model used for developing estimates or projections should be subjected to the following:

1. Sensitivity analysis to determine if changes in key model inputs cause key model outputs to respond in a sensible fashion;
2. Model validation to analyze a model's performance by comparing the results to available independent information sources; and
3. Demonstration of reproducibility to show that, given the same inputs, the model produces similar results.

Guideline 4.1.5: Prior to producing estimates, establish criteria for determining when the error (both sampling and nonsampling) associated with a direct survey estimate, model-based estimate, or projection is too large to publicly release the estimate/projection.

Guideline 4.1.6: Document methods and models used to generate estimates and projections to help ensure objectivity, utility, transparency, and reproducibility of the estimates and projections. (For details on documentation, see Section 7.3). Also, archive data and models so the estimates/projections can be reproduced.

For more information on developing model-based estimates, see *FCSM Statistical Policy Working Paper 21, Indirect Estimators in Federal Programs.*

SECTION 5 DATA ANALYSIS

Section 5.1 Analysis and Report Planning

Standard 5.1: Agencies must develop a plan for the analysis of survey data prior to the start of a specific analysis to ensure that statistical tests are used appropriately and that adequate resources are available to complete the analysis.

Key Terms: key variables, response rates

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 5.1.1: Include the following in the analysis plan:

1. An introduction that describes the purpose, the research question, relevant literature, data sources (including a brief description of the survey data and any limitations of the data), key variables to be used in the analysis, type of analysis, and significance level to be used;

2. Table and figure shells that support the analysis; and

3. A framework for technical notes including, as appropriate, the history of the survey program, data collection methods and procedures, sample design, response rates and the treatment of missing data, weighting methods, computation of standard errors, instructions for constructed

variables, limitations of the data, and sources of error in the data.

Guideline 5.1.2: Include standard elements of project management in the plan, including target completion dates and the resources needed to complete each activity.

Section 5.2 Inference and Comparisons

Standard 5.2: Agencies must base statements of comparisons and other statistical conclusions derived from survey data on acceptable statistical practice.

Key Terms: Bonferroni adjustment, covariance, estimates, hypothesis test, multiple comparisons, *p* value, standard error, statistical significance, Type I error

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 5.2.1: Specify the criterion for judging statistical significance for tests of hypotheses (Type I error) before conducting the testing.

Guideline 5.2.2: Before including statements in information products that two characteristics being estimated differ in the actual population, make comparison tests between the two estimates, if either is constructed from a sample. Use methods for comparisons appropriate for the nature of the estimates. In most cases, this requires estimates of the standard error of the estimates and, if the estimates are not independent, an estimate of the covariance between the

two estimates.

Guideline 5.2.3: When doing multiple comparisons with the same data between subgroups, include a note with the test results indicating whether or not the significance criterion (Type I error) was adjusted and, if adjusted, by what method (e.g., Bonferroni, modified Bonferroni, Tukey).

Guideline 5.2.4: When performing comparison tests, test and report only the differences that are substantively meaningful (i.e., don't necessarily run a comparison between every pair of estimates; run only those that are meaningful within the context of the data, and report only differences that are large enough to be substantively meaningful, even if other differences are also statistically significant).

Guideline 5.2.5: Given a comparison that does not have a statistically significant difference, conclude that the data do not support a statement that they are different. If the estimates have apparent differences, but have large standard errors making the difference statistically insignificant, note this in the text or as a note with tables or graphs.

Guideline 5.2.6: Support statements about monotonic trends (strictly increasing or decreasing) in time series using appropriate tests. If extensive seasonality, irregularities, known special causes, or variation in trends are present in the data, take those into account in the trend analysis.

Guideline 5.2.7: If part of an historical series is revised, data for both the old and the new series

should be published for a suitable overlap period for the use of analysts.

SECTION 6 REVIEW PROCEDURES

Section 6.1 Review of Information Products

Standard 6.1: Agencies are responsible for the quality of information that they disseminate and must institute appropriate content/subject matter, statistical, and methodological review procedures to comply with OMB and agency Information Quality Guidelines.

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 6.1.1: All information products should undergo a content/subject-matter review. Those conducting the review should have appropriate expertise in the subject matter, operation, or statistical program discussed in the document. Among the areas that reviewers should consider are the following:

1. Subject-matter literature is referenced in the document if appropriate;
2. Information is factually correct; and
3. Information is presented clearly and logically, conclusions follow from analysis, and no anomalous findings are ignored.

Guideline 6.1.2: All information products should undergo a statistical and methodological review. Those conducting the review should have appropriate expertise in the methodology described in the document. Among the tasks that reviewers should consider are the following:

1. Review assumptions and limitations for accuracy and appropriateness;
2. Ensure that appropriate statistical methods are used and reported;
3. Review calculations and formulas for accuracy and statistical soundness;
4. Review data and presentations of data (e.g., tables) to ensure disclosure risk avoidance, as necessary;
5. Review contents, conclusions, and technical (statistical and operational areas) recommendations to ensure that they are supported by the methodology used; and
6. Ensure that data sources and technical documentation, including data limitations, are included or referenced.

Guideline 6.1.3: Information products disseminated electronically should be reviewed for Section 508 compliance for accessibility by persons with disabilities. Ensure that any product that is disseminated via special software is tested for accessibility and interpretability prior to dissemination.

SECTION 7 DATA DISSEMINATION

Section 7.1 Releasing Information

Standard 7.1: Agencies must release information intended for the general public according to a dissemination plan that provides for equivalent, timely access to all users and provides information to the public about any planned or unanticipated data revisions.

Key Terms: estimate, forecast, key variables, model, nonsampling error, variance

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 7.1.1: Dissemination procedures for major information products include the following:

1. Develop schedule and mode for the release of information products;
2. Inform targeted audiences; and
3. Ensure equivalent, timely access to all users.

Guideline 7.1.2: Protect information against any unauthorized prerelease, and release information only according to established release procedures.

Guideline 7.1.3: If revisions to estimates are planned, establish a schedule for anticipated revisions, make it available to users, and identify initial releases as preliminary.

Guideline 7.1.4: Establish a policy for handling unscheduled corrections due to previously unrecognized errors. The policy may include threshold criteria (e.g., the correction will change a

national level total value by more than one percent or a regional value by more than five percent) identifying conditions under which data will be corrected and disseminated.

Guideline 7.1.5: When information products are disseminated, provide users access to the following information:

1. Definitions of key variables;
2. Source information, such as a survey form number and description of methodology used to produce the information or links to the methodology;
3. Quality-related documentation such as conceptual limitations and nonsampling error;
4. Variance estimation documentation;
5. Time period covered by the information and units of measure;
6. Data taken from alternative sources;
7. Point of contact to whom further questions can be directed;
8. Software or links to software needed to read/access the information and installation/operating instructions, if applicable; and
9. Date the product was last updated.

Guideline 7.1.6: Information products derived using models should adhere to the following:

1. Forecasts and derived estimates should be clearly identified; and
2. Descriptions of forecasting models or derivation procedures should be accessible from the product along with any available evaluation of its accuracy.

Guideline 7.1.7: Establish criteria for instances when information will not be publicly disseminated (e.g., underlying data are of insufficient quality).

For more information on electronic dissemination of statistical data, see *FCSM Statistical Policy Working Paper 24, Electronic Dissemination of Statistical Data.*

Section 7.2 Data Protection and Disclosure Avoidance for Dissemination

Standard 7.2: Agencies must release all information products in accordance with the survey pledge to the respondents and all applicable Federal legislation.

Key Terms: confidentiality, data protection, disclosure

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 7.2.1: For survey information collected under a pledge of confidentiality, employ sufficient procedures and mechanisms to protect any individually-identifiable data.

Guideline 7.2.2: Do not publicly reveal parameters associated with disclosure limitation rules.

For more information, see *FCSM Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology,* and forthcoming OMB guidance on the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).

Section 7.3 Survey Documentation

Standard 7.3: Agencies must produce survey documentation that includes those materials necessary to understand how to properly analyze data from each survey, as well as the information necessary to replicate and evaluate each survey's results (See also Standard 1.2). Survey documentation must be readily accessible to users, unless it is necessary to restrict access to protect confidentiality.

Key Terms: coverage, editing, imputation, instrument, nonsampling error, response rates, sampling error, sampling unit, strata, variance

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 7.3.1: Survey system documentation includes all information necessary to properly analyze the data. Along with the final data set, documentation, at a minimum, includes the following:

1. Description of variables used to uniquely identify records in the data file;

2. Description of the sample design, including strata and sampling unit identifiers to be used for analysis; and
3. Final instrument(s) or a facsimile thereof for surveys conducted through a computer-assisted telephone interview (CATI) or computer-assisted personal interview (CAPI) or Web instrument that includes the following:
    - All items in the instrument (e.g., questions, check items, and help screens);
    - Items extracted from other data files to prefill the instrument (e.g., dependent data from a prior round of interviewing); and
    - Items that are input to the post data collection processing steps (e.g., output of an automated instrument);
4. Definitions of all variables, including all modifications;
5. Data file layout;
6. Descriptions of constructed variables on the data file that are computed from responses to other variables on the file;
7. Unweighted frequency counts;
8. Description of sample weights, including adjustments for nonresponse and benchmarking and how to apply them;
9. Description of how to calculate variance estimates appropriate for the survey design;
10. Description of all editing and imputation methods applied to the data and how to remove imputed values from the data;
11. Descriptions of known data anomalies and corrective actions;
12. Description of the magnitude of sampling error associated with the survey, and how it was calculated;

13. Description of the sources of nonsampling error associated with the survey (e.g., coverage, measurement);
14. Unit response rates (weighted and unweighted);
15. Overall response rates (weighted and unweighted); and
16. Item response rates for variables with rates below 70 percent.

Guideline 7.3.2: To ensure that a survey can be replicated and evaluated, the internal archived portion of the survey system documentation, at a minimum, includes the following:

1. Survey planning and design decisions, including OMB Information Collection Request package;
2. Field test design and results;
3. Selected sample;
4. Sampling frame;
5. Justifications for the items on the survey instrument, including why the final items were selected;
6. All instructions to respondents and/or interviewers either about how to properly respond to a survey item or how to properly present a survey item;
7. Description of the data collection methodology;
8. Sampling plan and justifications, including any deviations from plan;
9. Data processing plan specifications and justifications;
10. Final weighting plan specifications, including calculations for how the final weights were derived, and justifications;
11. Final imputation plan specifications and justifications;

12. Data editing plan specifications and justifications;

13. Evaluation reports;

14. Descriptions of models used for indirect estimates and projections;

15. Analysis plans;

16. Time schedule for revised data; and

17. Documentation made publicly available in conjunction with the release of data.

Guideline 7.3.3: Where appropriate, develop and implement methods for bounding or estimating the nonsampling error from each source identified in the evaluation plan.

Guideline 7.3.4: For recurring surveys, produce a periodic evaluation report, such as a methodology report, that itemizes all sources of identified error. Where possible, provide estimates or bounds on the magnitudes of these errors; discuss the total error model for the survey; and assess the survey in terms of this model.

Guideline 7.3.5: All survey documentation should be retained by an agency according to its records disposition and archival policy.

For more information on measuring and reporting sources of errors in surveys, see *FCSM Statistical Policy Working Paper 31, Measuring and Reporting Sources of Error in Surveys.*

Section 7.4 Documentation and Release of Public-Use Microdata

Standard 7.4: Agencies that release microdata to the public must include documentation clearly describing how the information is constructed and provide the metadata necessary for users to access and manipulate the data (See also Standard 1.2). Public-use microdata documentation and metadata must be readily accessible to users.

Key Terms: microdata, public-use microdata, record layout, stage of the data collection

---

The following guidelines represent best practices that may be useful in fulfilling the goals of the standard:

Guideline 7.4.1: Provide complete documentation for all data files. See Section 7.3 for additional information on file documentation.

Guideline 7.4.2: Provide a file description and record layout for each file. All variables must be clearly identified and described.

Guideline 7.4.3: All microdata products and documentation should be accessible by users with generally available software.

Guideline 7.4.4: All imputed values should be clearly identified on the data file.

Guideline 7.4.5: Public-use microdata should be released as soon as practicable to ensure timely availability for data users.

Guideline 7.4.6: All microdata products and documentation should be retained by an agency according to its records disposition and archival policy.

GLOSSARY

-B-

Bias is the deviation of the average survey value from the true population value. Bias refers to systematic errors that affect any sample taken under a specific design with the same constant error.

Bonferroni adjustment is a procedure for guarding against an increase in the probability of a Type I error when performing multiple significance tests. To maintain the probability of a Type I error at some selected value alpha, each of the m tests to be performed is judged against a significance level, alpha/m.

A **bridge study** continues an existing methodology concurrent with a new methodology for the purpose of defining the relationship between the new and old estimates.

-C-

Coding involves converting information into numbers or other symbols that can be more easily counted and tabulated.

Cognitive interviews are use to develop questionnaires. In a cognitive interview, respondents are required to report aloud everything they are thinking as they attempt to answer a survey question.

Confidentiality involves the protection of individually identifiable data from unauthorized disclosures.

A **consistent data series** maintains comparability over time by keeping an item fixed, or by incorporating appropriate adjustment methods in the event an item is changed.

Covariance is a characteristic that indicates the strength of relationship between two variables. It is the expected value of the product of the deviations of two random variables, x and y from their respective means.

Coverage refers to the extent to which all elements on a frame list are members of the population, and to which every element in a population appears on the frame list once and only once.

Coverage error refers to the discrepancy between statistics calculated on the frame population and the same statistics calculated on the target population. Undercoverage errors occur when target population units are missed during frame construction, and overcoverage errors occur when units are duplicated or enumerated in error.

A **crosswalk study** delineates how categories from one classification system are related to categories in a second classification system.

A **cross-sectional** sample survey is based on a representative sample of respondents drawn from a population at one point in time.

Cross-sectional imputations are based on data from a single time period.

Cross-wave imputations are imputations based on data from multiple time periods. For example, a **cross-sectional imputation** for a time 2 salary could simply be a donor's time 2 salary. Alternatively, a cross-wave imputation could be the change in a donor's salary from time 1 to time 2 multiplied by the time 1 nonrespondent's salary.

A **cut-off sample** is a nonprobability sample that consists of the units in the population that have the largest values of a key variable (frequently the variable of interest from a previous time period). For example, a 90% cut-off sample consists of the largest units accounting for at least 90% of the population total of the key variable. Sample selection

is usually done by sorting the population in decreasing order by size, and including units in the sample until the percent coverage exceeds the established cut-off.

-D-

Data protection involves techniques that are used to insure that confidential individually identifiable data are not disclosed.

Data series are repeated collections of sequential cross-sectional or longitudinal data characteristics of the target population over time.

The **design effect (DEFF)** is the ratio of the true variance of a statistic (taking the complex sample design into account) to the variance of the statistic for a simple random sample with the same number of cases. Design effects differ for different subgroups and different statistics; no single design effect is universally applicable to any given survey or analysis.

Direct survey-based estimates are intended to achieve efficient and robust estimates of the true values of the target populations, based on the sample design and resulting survey data.

Disclosure means the public release of individually identifiable data that were obtained under a pledge of confidentiality.

Domain refers to a defined universe of knowledge, skills, abilities, attitudes, interests, or other human characteristics.

-E-

Editing is a procedure that uses available information and some assumptions to derive substitute values for inconsistent values in a data file.

Effect size refers to the standardized magnitude of the effect or the departure from the null hypothesis. For example, the effect size may be the amount of change over time, or the difference between two population means, divided by the appropriate population standard deviation. Multiple measures of effect size can be used (e.g., standardized differences between means, correlations, and proportions).

The **effective sample size**, as used in the design phase, is the sample size under a simple random sample design that is equivalent to the actual sample under the complex sample design. In the case of complex sample designs, the actual sample size is determined by multiplying the effective sample size by the anticipated design effect.

An **eligible sample unit** is a unit selected for a sample, which is confirmed to be a member of the target population.

Estimates result from the process of providing a numerical value for a population parameter on the basis of information collected from a survey and/or other sources.

Estimation is the process of using data from a survey and/or other sources to provide a value for an unknown population parameter (such as a mean, proportion, correlation, or effect size), or to provide a range of values in the form of a confidence interval.

Estimation error is the difference between a survey estimate and the true value of the target population.

-F-

In a **field test,** all or some of the survey procedures are tested on a small scale that mirrors the planned full-scale implementation.

A **focus group** involves a semi structured group discussion of a topic.

Forecasts involve the specific projection that an investigator believes is most likely to provide an accurate prediction of a future value of some process.

A **frame** is a mapping of the universe elements (i.e., sampling units) onto a finite list (e.g., the population of schools on the day of the survey).

The **frame population** is the set of elements that can be enumerated prior to the selection of a survey sample.

-H-

Hypothesis testing draws a conclusion about the tenability of a stated value for a parameter. For example, sample data may be used to test whether an estimated value of a parameter (such as the difference between two population means) is sufficiently different from zero that the null hypothesis, designated H_0 (no difference in the population means), can be rejected in favor of the alternative hypothesis, H_1 (a difference between the two population means).

-I-

Imputation is a procedure that uses available information and some assumptions to derive substitute values for missing values in a data file.

Individually identifiable data refers specifically to data from any list, record, response form, completed survey, or aggregation about an individual or individuals from which information about particular individuals or their schools/education institutions may be revealed by either direct or indirect means.

Instrument refers to an evaluative device that includes tests, scales, and inventories to measure a domain using standardized procedures. It is commonly used in surveys to refer to the device used to collect data, such as a questionnaire or data entry software.

Item nonresponse occurs when a respondent fails to respond to one or more relevant item(s) on a survey.

-K-

Key variables include survey-specific items for which aggregate estimates are commonly published from a study. They include, but are not restricted to, variables most commonly used in table row stubs. Key variables also include important analytic composites and other policy-relevant variables that are essential elements of the data collection. They are first defined in the initial planning stage of a survey, but may be added to as the survey and resulting analyses develop. For example, a study of student achievement might use gender, race-ethnicity, urbanicity, region, and school type (public/private) as key reporting variables.

-L-

A **longitudinal** sample survey follows the experiences and outcomes over time of a representative sample of respondents (i.e., a cohort) who are defined based on a shared experience (e.g., shared birth year or grade in school).

Longitudinal analysis involves the analysis of data from a study in which subjects are measured repeatedly over time.

-M-

Response to a **mandatory survey** is required by law.

Measurement error is the difference between observed values of a variable recorded under similar conditions and some fixed true value (e.g., errors in reporting, reading, calculating, or recording a numerical value). Response bias is the deviation of the survey estimate from the true population value that is due to measurement error from the data collection. Potential sources of response bias include the respondent, the instrument, and the interviewer.

A **microdata** file includes the detailed responses for individual respondents.

The **minimum substantively significant effect (MSSE)** is the smallest effect, that is, the smallest departure from the null hypothesis, considered to be important for the analysis of key variables. The minimum substantively significant effect is determined during the design phase. For example, the planning document should provide the minimum change in key variables or perhaps, the minimum correlation, r, between two variables that the survey should be able to detect for a specified population domain or subdomain of analytic interest. The MSSE should be based on a broad knowledge of the field, related theories, and supporting literature.

Missing at random, for a given survey variable, refers to a situation in which the probability that a unit is missing that variable is independent of its value, but may not be independent of another variable being measured.

Missing completely at random occurs when values are missing because individuals drop out of a study in a process that is independent of both the observed measurements and those that would have been available had they not been missing.

A **model** is a formalized set of mathematical expressions quantifying the process assumed to have generated a set of observations.

A **model-based estimate** is produced by a model.

Model-based samples are selected to achieve efficient and robust estimates of the true values of the target populations under a chosen working.

Model validation involves testing a model's predictive capabilities by comparing the model results to "known" sources of empirical data.

Multiple comparisons involve a detailed examination of the differences between a set of means.

Multivariate analysis is a generic term for many methods of analysis that are used to investigate **multivariate data**.

Multivariate data include data for which each observation consists of values for more than one random variable.

Multivariate modeling provides a formalized mathematical expression of the process assumed to have generated the observed multivariate data.

-N-

Nonprobabilistic methods—see ""probabilistic methods."

Nonresponse bias occurs when the observed value deviates from the population parameter due to differences between respondents and nonrespondents. Nonresponse bias is likely to occur as a result of not obtaining 100 percent response from the selected cases.

Nonresponse error is the overall error observed in estimates caused by differences between respondents and nonrespondents. It consists of a variance component and nonresponse bias.

Nonsampling error includes measurement errors due to interviewers, respondents, instruments, and mode; nonresponse error; coverage error; and processing error.

-O-

Overall unit nonresponse reflects a combination of unit nonresponse across two or more levels of data collection, where participation at the second stage of data collection is conditional upon participation in the first stage of data collection.

-P-

The ***p* value** is the probability of the observed data, or data showing a more extreme departure from the null hypothesis, occurring when the null hypothesis is true.

In a **pilot test,** a laboratory or a very small-scale test of a questionnaire or procedure is conducted.

Population—see "target population."

Post-stratification is applied to survey data, in which sample units are stratified after data collection using information collected in the survey and auxiliary information to adjust weights to population control totals.

The **power** $(1 - b)$ of a test is defined as the probability of rejecting the null hypothesis when a specific alternative hypothesis is assumed. For example, with $b = 0.20$ for a particular alternative hypothesis, the power is 0.80, which means that 80 percent of the time the test statistic will fall in the rejection region if the parameter has the value specified by the alternative hypothesis.

Precision of survey results refers to how closely the results from a sample can reproduce the results that would be obtained from a complete count (i.e., census) conducted using the same techniques. The difference between a sample result and the result from a complete census taken under the same conditions is known as the precision of the sample result.

A survey **pretest** involves experimenting with different components of the questionnaire or survey design or operationalization prior to full-scale implementation. This may involve **pilot testing**, that is a laboratory or a very small-scale test of a questionnaire or procedure, or a **field test** in which all or some of the survey procedures are tested on a small scale that mirrors the planned full-scale implementation.

Probabilistic methods for survey sampling are any of a variety of methods for sampling that give a known, non-zero, probability of selection to each member of the target population. The advantage of probabilistic sampling methods is that sampling error can be calculated. Such methods include: random sampling, systematic sampling, and

stratified sampling. They do not include: convenience sampling, judgment sampling, quota sampling, and snowball sampling.

Probability of selection in a survey is the probability that a given sampling unit will be selected, based on the probabilistic methods used in sampling.

A **projection** is an estimate of a future value of a characteristic based on current trends.

A **public-use data file or public-use microdata** includes a subset of data that have been coded, aggregated, or otherwise altered to mask individually identifiable information, and thus is available to all external users. Unique identifiers, geographic detail, and other variables that cannot be suitably altered are not included in public-use data files.

-Q-

Quality assurance processing includes any procedure or method that is aimed at maintaining or improving the reliability or validity of the data.

-R-

Raking is a multiplicative weighting technique that uses iterative proportional fitting. That is, weights are obtained as the product of a number of factors contributed by auxiliary variables.

In **ratio estimation**, an auxiliary variate x_i, correlated with y_i, is obtained for each unit in the sample. The population total X of the x_i must be known. In practice, x_i is often the value of y_i at some previous time when a complete census was taken. The goal is to obtain increased precision by taking advantage of the correlation between y_i and x_i. The ratio estimate of Y the population total of y_i is YR = (y/x), where y and x are the sample totals of y_i and x_i, respectively.

A **record layout** is a description of the data elements on the file (variable names, data types, and length of space on the file) and their physical locations.

Required response items include the minimum set of items required for a case to be considered a respondent.

Respondent burden is the estimated total time and financial resources expended by the survey respondent to generate, maintain, retain, and provide survey information.

A **Response analysis survey** is a study of the capability of respondents to accurately provide the data requested for a survey.

Response rates calculated using base weights measure the proportion of the sample frame that is represented by the responding units in each study.

-S-

Sampling error is the error associated with nonobservation, that is, the error that occurs because all members of the frame population are not measured. It is the error associated with the variation in samples drawn from the same frame population. The variance equals the square of the sampling error.

Sampling units are the basic components of a sample frame. Everything covered by a sample frame must belong to one definite sampling unit, or have a measurable probability of belonging to a specific unit. The sampling unit may contain, for example, houses, people, or businesses.

Sensitivity analysis is designed to determine how the variation in the output of a model (numerical or otherwise) can be apportioned, qualitatively or quantitatively, to changes in input parameter values and assumptions. This type of analysis is useful in ascertaining the capability of a given model, as well its robustness and reliability.
Stage of data collection includes any stage or step in the sample identification and data collection process in which data are collected from the identified sample unit. This includes information obtained that is required to proceed to the next stage of sample selection or data collection (e.g., school district permission for schools to participate or schools providing lists of teachers for sample selection of teachers).
Standard error is the standard deviation of the sampling distribution of a statistic. Although the standard error is used to estimate sampling error, it includes some nonsampling error.
Strata are created by partitioning the frame and are generally defined to include relatively homogeneous units within strata.
Statistical significance is attained when a statistical procedure applied to a set of observations yields a *p* value that exceeds the level of probability at which it is agreed that the null hypothesis will be rejected.
A **statistical survey** is a data collection whose purposes include the description, estimation, or analysis of the characteristics of groups, organizations, segments, activities, or geographic areas. A statistical survey may be a census or may collect information from a sample of the **target population**.
Substitutions are done using matched pairs, in which the alternate member of the pair does not have an independent probability of selection.

-T-
The **target population** is any group of potential sample units or persons, businesses, or other entities of interest.
The **total mean square error** is a measure of the combined overall effect of sampling and nonsampling error on the estimate.
Type I error is made when the tested hypothesis, H0, is falsely rejected when in fact it is assumed true. The probability of making a Type I error is denoted by alpha (α). For example, with an alpha level of 0.05, the analyst will conclude that a difference is present in 5 percent of tests where the null hypothesis is true.

-U-
Unit nonresponse occurs when a respondent fails to respond to all required response items (i.e., fails to fill out or return a data collection instrument).
A **universe** survey involves the collection of data covering all known units in a population (i.e., a census).

-V-
Validation studies are conducted to independently verify that the data collection methodology employed will obtain accurate data for the concept studied.
Variance estimates—see "sampling error."

Response to a **voluntary** survey is not required by law.

-W-

A **wave** is a round of data collection in a longitudinal survey (e.g., the base year and each successive followup are each waves of data collection).

Weights are relative values associated with each sample unit that are intended to correct for unequal probabilities of selection for each unit due to sample design. Weights most frequently represent the relative portion of the population that the unit represents. Weights may be adjusted for nonresponse.