

QUESTIONS AND ANSWERS
WHEN DESIGNING SURVEYS
FOR INFORMATION COLLECTIONS

Office of Information and Regulatory Affairs
Office of Management and Budget

Last Modified: October 2016
Originally Published: January 2006

Table of Contents

PURPOSE OF THIS GUIDANCE.....	1
1. WHAT IS THE PURPOSE OF THIS GUIDANCE?	1
2. DOES THIS GUIDANCE APPLY TO ALL ICRs SUBMITTED TO OMB?.....	1
SUBMISSION OF ICRs TO OMB.....	3
3. WHEN SHOULD AN AGENCY BEGIN THE PRA PROCESS?	3
4. WHEN SHOULD AGENCIES TALK TO OMB ABOUT PLANS FOR A STUDY?	3
5. WHAT DOES IT MEAN FOR AN AGENCY TO CONDUCT OR SPONSOR AN INFORMATION COLLECTION?.....	4
6. WHEN ARE STUDIES INVOLVING THIRD PARTY OR INVESTIGATOR-INITIATED GRANTS SUBJECT TO PRA REVIEW?	4
7. ARE FOCUS GROUPS SUBJECT TO PRA REVIEW?	5
8. WHAT ARE GENERIC CLEARANCES AND WHEN ARE THESE USEFUL FOR AGENCIES?	6
9. WHAT NEEDS TO BE DONE FOR AN EMERGENCY CLEARANCE?	6
10. WHEN DO AGENCIES NEED TO COMPLETE PART B OF THE ICR SUPPORTING STATEMENT?.....	7
11. WHY DO AGENCIES NEED TO COMPLETE SOME OF PART B IF THEY ARE CONDUCTING QUALITATIVE RESEARCH STUDIES OR PROGRAM EVALUATIONS?	7
SCOPE OF THE INFORMATION COLLECTION	9
12. WHY IS THIS DATA COLLECTION NECESSARY AND HOW WILL THE INFORMATION BE USED?	9
13. HOW OFTEN SHOULD DATA BE COLLECTED?	10
14. WHAT IS INCLUDED IN THE CALCULATION OF BURDEN HOURS?	11
15. FOR ESTABLISHMENT SURVEYS OR PANEL SURVEYS, SHOULD BURDEN HOURS INCLUDE THE ORIGINAL COLLECTION OF ADMINISTRATIVE RECORDS THAT MAY HAVE TAKEN PLACE MONTHS OR YEARS BEFORE?	11
16. WHY ARE AGENCIES REQUIRED TO ESTIMATE THE BURDEN IN TERMS OF BOTH TIME AND COSTS?	12
CHOICE OF METHODS.....	13
17. HOW DOES THE CHOICE OF METHODS FOR THE STUDY RELATE TO THE RESEARCH QUESTIONS OR PURPOSE OF THE COLLECTION?	13
18. HOW DO THE USE OF THE INFORMATION AND CHOICE OF METHODS FOR THE STUDY RELATE TO THE AGENCY'S INFORMATION QUALITY GUIDELINES?	14
19. WHEN SHOULD AGENCIES CONSIDER CONDUCTING A SURVEY?	14
20. WHAT SHOULD AGENCIES CONSIDER WHEN DESIGNING AND CONDUCTING A SURVEY?	15
21. WHEN SHOULD AGENCIES CONSIDER CONDUCTING A QUALITATIVE STUDY?	16
22. WHEN SHOULD AGENCIES CONDUCT A PILOT STUDY, PRETEST, OR FIELD TEST?	17
23. WHEN SHOULD AGENCIES CONSIDER CONDUCTING FOCUS GROUPS OR COGNITIVE INTERVIEWS?	18
24. WHEN SHOULD AGENCIES CONSIDER USING CASE STUDY METHODS?	18
25. WHAT SHOULD AGENCIES CONSIDER WHEN DESIGNING AND CONDUCTING A CASE STUDY?	19
26. WHEN SHOULD AGENCIES CONSIDER USING EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS?.....	20
SAMPLING.....	23
27. WHAT IS THE DIFFERENCE BETWEEN A CENSUS AND A SAMPLE SURVEY AND WHEN IS EACH APPROPRIATE?	23
28. WHAT IS A SAMPLING FRAME AND WHAT IS THE COVERAGE OF THE SAMPLING FRAME?	24
29. IS A LIST OF INTERNET SUBSCRIBERS AVAILABLE AND ACCEPTABLE FOR USE AS A SAMPLING FRAME?	24
30. WHAT IS AN APPROPRIATE SAMPLE DESIGN TO ENSURE THE SAMPLE DRAWN REPRESENTS THE POPULATION OF INTEREST?	25
31. ARE PROBABILITY SAMPLES ALWAYS THE BEST FOR SURVEYS OF ESTABLISHMENTS?	26
32. WHAT INFORMATION SHOULD AGENCIES PROVIDE ABOUT THEIR COMPLEX SAMPLE DESIGNS?	27
33. HOW LARGE SHOULD A SAMPLE BE FOR A STATISTICAL SURVEY?	27

SAMPLING (Continued)	23
34. CAN PRE-EXISTING SURVEY PANELS, SUCH AS INTERNET PANELS, BE USED TO OBTAIN REPRESENTATIVE SAMPLES?	29
35. WHAT ARE SOME COMMON NONPROBABILITY SAMPLES, AND WHY ARE THEY USED?	29
MODES OF DATA COLLECTION	32
36. WHAT ARE THE DIFFERENT MODES OF SURVEY DATA COLLECTION?	32
37. WHAT MODE OF DATA COLLECTION IS APPROPRIATE FOR A GIVEN SURVEY?	33
38. WHEN SHOULD AGENCIES CONSIDER A MIXED-MODE APPROACH?	34
39. HOW DOES GPEA AFFECT CHOICE OF MODES FOR SURVEY DATA COLLECTION?	34
40. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF MAIL SURVEYS?	35
41. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF TELEPHONE INTERVIEWING?	35
42. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF IN-PERSON INTERVIEWING?	37
43. WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF USING INTERNET SURVEYS?	37
44. HOW DOES THE DATA COLLECTION MODE AFFECT QUESTIONNAIRE DESIGN?	38
QUESTIONNAIRE DESIGN AND DEVELOPMENT	40
45. WHAT SHOULD AGENCIES DO WHEN DEVELOPING NEW QUESTIONNAIRES OR QUESTIONNAIRE ITEMS?	40
46. WHY SHOULD AGENCIES CONSIDER USING QUESTIONS PREVIOUSLY USED BY OTHER AGENCIES OR RESEARCHERS?	40
47. WHEN IS IT ACCEPTABLE TO DUPLICATE QUESTIONS USED ON OTHER SURVEYS?	41
48. WHAT TECHNIQUES CAN BE USED TO DEVELOP NEW QUESTIONS?	41
49. WHAT ROLE DOES PRETESTING PLAY IN QUESTIONNAIRE DEVELOPMENT?	43
50. WHAT DO AGENCIES NEED TO DO TO OBTAIN CLEARANCE FOR PRETESTING ACTIVITIES?	44
51. WHAT IS A GENERIC CLEARANCE FOR PRETESTING ACTIVITIES?	45
STATISTICAL STANDARDS	47
52. WHAT ARE OMB STATISTICAL CLASSIFICATIONS, DEFINITIONS, AND DATA SOURCES?	47
53. WHAT STATISTICAL CLASSIFICATIONS HAVE BEEN ADOPTED BY OMB?	47
54. WHAT STANDARD DEFINITIONS AND DATA SOURCES HAVE BEEN ADOPTED BY OMB?	49
55. WHAT ARE THE REQUIREMENTS FOR COLLECTING INDIVIDUAL DATA ON RACE AND ETHNICITY?	49
INFORMING RESPONDENTS ABOUT THEIR PARTICIPATION AND THE CONFIDENTIALITY OF THEIR DATA	51
56. WHAT SHOULD RESPONDENTS BE TOLD ABOUT THEIR PARTICIPATION IN AN INFORMATION COLLECTION?	51
57. WHAT IS A PLEDGE OF CONFIDENTIALITY AND HOW SHOULD A PLEDGE OF CONFIDENTIALITY BE MADE TO RESPONDENTS?	52
58. WHAT LEGAL AUTHORITY DOES AN AGENCY HAVE TO PROTECT THE CONFIDENTIALITY OF INFORMATION IT IS COLLECTING?	53
59. WHAT IS THE CONFIDENTIAL INFORMATION PROTECTION AND STATISTICAL EFFICIENCY ACT OF 2002 (CIPSEA)?	53
60. IF AN AGENCY DOES NOT COLLECT DATA UNDER CIPSEA, HOW CAN IT PROTECT THE CONFIDENTIALITY OF THE DATA?	54
61. WHAT MUST BE DONE TO PROTECT DATA THAT ARE GATHERED UNDER A PLEDGE OF CONFIDENTIALITY?	55
RESPONSE RATES AND INCENTIVES	56
62. WHY ARE RESPONSE RATES IMPORTANT?	56
63. HOW SHOULD RESPONSE RATES BE CALCULATED?	57
64. WHEN SHOULD WEIGHTED RESPONSE RATES BE REPORTED?	58
65. WHAT ARE TYPICAL RESPONSE RATES FOR FEDERAL GOVERNMENT STATISTICAL SURVEYS?	59
66. WHAT ARE ACCEPTABLE RESPONSE RATES FOR DIFFERENT KINDS OF SURVEY COLLECTIONS?	60
67. DO LONGITUDINAL AND MULTI-STAGE SURVEYS NEED TO ACHIEVE THE SAME LEVELS OF RESPONSE RATES AS OTHER SURVEYS?	61

RESPONSE RATES AND INCENTIVES (Continued)	56
68. ARE DIFFERENT RESPONSE RATES ACCEPTABLE FOR DIFFERENT MODES OF DATA COLLECTION?	62
69. HOW CAN RESPONSE RATES BE IMPROVED?	62
70. GIVEN THAT RANDOM DIGIT DIALING (RDD) TELEPHONE SURVEY RESPONSE RATES HAVE BEEN DECLINING, WILL OMB APPROVE ICRs WITH THIS METHODOLOGY?	64
71. HOW CAN AGENCIES EXAMINE POTENTIAL NONRESPONSE BIAS?	64
72. WHAT RESPONSE RATE ISSUES ARE INVOLVED WITH USING SAMPLES DERIVED FROM PRE-EXISTING MULTIPURPOSE PANELS, SUCH AS INTERNET OR CONSUMER PANELS?	66
73. WHAT SHOULD AGENCIES DO TO ASSESS AND DEAL WITH NONRESPONSE BIAS DUE TO ITEM NONRESPONSE?	67
74. WHAT ARE INCENTIVES?	68
75. WHY MUST AGENCIES PROVIDE A JUSTIFICATION TO GIVE INCENTIVES TO RESPONDENTS?	69
76. WHAT FACTORS SHOULD AGENCIES ADDRESS IN THEIR JUSTIFICATION TO GIVE INCENTIVES TO RESPONDENTS?	69
ANALYSIS AND REPORTING	72
77. WHAT INFORMATION SHOULD AGENCIES INCLUDE IN THEIR ANALYSIS PLANS?	72
78. WHAT PREDISSEMINATION REVIEW DO AGENCIES NEED TO DO FOR REPORTS BASED ON SURVEYS OR STATISTICAL COLLECTIONS?	73
STUDIES USING STATED PREFERENCE METHODS	74
79. WHAT ARE STATED PREFERENCE METHODS?	74
80. WHAT SHOULD AGENCIES CONSIDER WHEN DESIGNING QUESTIONS FOR STATED PREFERENCE STUDIES?	74
81. WHAT FACTORS SHOULD BE CONSIDERED WHEN DESIGNING OR EVALUATING STUDIES USING STATED PREFERENCE METHODS?	75
GLOSSARY OF ABBREVIATIONS	77
GLOSSARY OF TERMS	78
APPENDICES	87

PURPOSE OF THIS GUIDANCE

Federal agencies conduct or sponsor a wide variety of information collections to gather data from businesses, individuals, schools, hospitals, and State, local, and tribal governments. Information collections employing surveys are frequently used for general purpose statistics, as well as for program evaluations or research studies that answer more specific research questions. Data collected by Federal agencies are widely used to make informed decisions and to provide necessary information for policy makers and planners. The collection of this information can take many forms and is accomplished in a variety of ways.

The Paperwork Reduction Act of 1995 (PRA) requires agencies to submit requests to collect information from the public to the Office of Management and Budget (OMB) for approval. This guidance is designed to assist agencies and their contractors in preparing Information Collection Requests (ICRs), which may be commonly known as PRA submissions or “OMB clearance packages,” for surveys used for general purpose statistics or as part of program evaluations or research studies.

1. What is the purpose of this guidance?

OMB is often asked about the ICR review process and what its expectations are, especially for collections involving surveys. These Q&As are designed to answer many of the frequently asked questions to help agencies better understand OMB’s expectations for survey information collection requests. This improved understanding should assist agencies in identifying and documenting information for inclusion in their ICRs, and should facilitate the review process.

This guidance seeks to highlight a wide range of issues that agencies need to consider when designing their surveys. Different sections of this guidance provide a very brief overview of the literature on statistical sampling and different survey methodology topics; each section provides some useful references for more information on these issues. The goal of this guidance is to help agencies to better plan and document their information collections that use surveys.

Conducting a high quality survey is a complex undertaking, and this guidance cannot (and is not intended to) take the place of professional survey methodologists and statisticians that agencies will need to consult in designing, executing, and documenting their surveys. For agencies that do not have these professionals on staff or involved in a particular collection, this guidance points out some key areas where professional consultation will be needed.

2. Does this guidance apply to all ICRs submitted to OMB?

The next two sections of this guidance (on submission of ICRs to OMB and scope of the information collection) cover some general requirements under the PRA that can generally be applied to any information collection request an agency makes. However, the focus of this guidance is on conducting surveys for general purpose statistics or as part of program evaluations or research studies.

Purpose

Surveys represent only a small percentage of all ICRs that OMB reviews. Most ICRs submitted to OMB are mandatory recordkeeping requirements, applications, or audits that are not used for statistical purposes. Because surveys require that careful attention be paid to a variety of methodological and statistical issues, agencies are required to complete Part B of the ICR supporting statement to more fully document how the survey will be conducted and analyzed (see question #10). The focus of this guidance is to assist agencies in planning surveys and documenting their proposed surveys in their ICRs.

SUBMISSION OF ICRs TO OMB

This section covers some basic questions related to the Paperwork Reduction Act (PRA) submissions that agencies prepare and submit to OMB including process issues, what is and is not covered by the PRA, and when agencies need to complete Part B of the Information Collection Request (ICR) supporting statement. Agencies should consult the OMB regulations implementing the PRA (5 C.F.R. § 1320) for more detailed and complete information.

3. When should an agency begin the PRA process?

The PRA requires that the agency publish a 60-day notice in the *Federal Register* to obtain public comment on the proposed collection, prior to submitting the information collection to OMB.¹ At the time this notice is published, agencies must have at least a draft survey instrument available for the public to review. Agencies should state in their ICRs whether any comments were received from the public, and the comments should be addressed in the ICR that is submitted to OMB.

When submitting the ICR to OMB, agencies are required to place a second notice in the *Federal Register*, allowing a 30-day public comment period and notifying the public that OMB approval is being sought and that comments may be submitted to OMB. This notice runs concurrent with the first 30 days of OMB review, and OMB has a total of 60 days after receipt of the ICR to make its decision.² Thus, agencies need to allow at least 120 days for consideration of initial public comments, the second public comment period and OMB review, plus additional time for preparation of the ICR, as well as time lags for publication of *Federal Register* notices.

Agencies may also have requirements for internal review or higher level reviews (e.g., departmental) that need to be factored into the schedule for planning a survey. A six month period, from the time the agency completes the ICR to OMB approval, is fairly common for planning purposes but varies considerably across agencies depending on internal review procedures. Thus, starting the process early can be very important to ensure timely data collection. Survey managers should consult with their agency paperwork clearance officers to ascertain what they need to do and the time required to meet agency and OMB requirements. In rare instances, the PRA does provide for expedited processing if an agency can justify an Emergency Collection (see question #9).

4. When should agencies talk to OMB about plans for a study?

The PRA and its implementing regulations provide a formal basis for OMB review of agency information collection requests. However, they do not preclude informal consultation with OMB desk officers prior to the submission of an ICR. Consultation with OMB prior to submission of an ICR is not required as part of the PRA and typically does not occur. However, if an agency is proposing a significant new collection about which it expects OMB may have questions or concerns, the agency is encouraged to consult with its OMB desk officer about the particular

¹ 5 C.F.R. § 1320.8(d)(1)

² 5 C.F.R. § 1320.10(a)

collection in advance of submitting the ICR to OMB. When an agency is planning a new, large survey data collection, a major revision to an ongoing survey, or large-scale experiments or tests, agencies and OMB frequently find it helpful for the agency to brief OMB on the nature of the planned collection and the proposed methodology. In this less formal context, OMB and agency staff can discuss potential areas of concern, including the need for further detail and justification. This kind of early consultation can considerably reduce the likelihood that major unexpected concerns about survey methodology or statistical sample design will arise during OMB review, and it allows more time for the agency to consider alternatives if necessary. Agencies can then address any issues identified by OMB in their ICRs. While this informal consultation does not affect the timing of the formal OMB review process under the PRA, it can be of benefit in identifying some issues much earlier and may avoid delays that could otherwise occur.

5. What does it mean for an agency to conduct or sponsor an information collection?

An agency conducts or sponsors an information collection if the agency collects the information using its own staff and resources, or causes another agency or entity to collect the information, or enters into a contract or cooperative agreement with another person or contractor to obtain the information.³ If the agency requests the collection directly or indirectly through another entity or contractor or exercises control over those collecting the information, the agency is conducting or sponsoring the collection (see also question #6).

6. When are studies involving third party or investigator-initiated grants subject to PRA review?

Collections of information conducted through investigator-initiated grants (e.g., in response to a Request for Applications (RFA)) are generally not subject to OMB review under the PRA. However, information collections by a Federal grant recipient are subject to PRA review if (1) the grant recipient is conducting the collection at the specific request of the agency, or (2) the terms and conditions of the grant require specific approval by the agency for the collection or collection procedures.⁴ If either of these conditions is met, the sponsoring agency needs to seek and obtain OMB approval, and the grantee needs to display the OMB control number on the collection instrument.

For example, the National Science Foundation has many program areas that support basic research on a wide variety of topics. Proposals are reviewed by scientific panels and funding may be provided to a university researcher to study some topic, which may include a survey. Although the National Science Foundation funded the research, it did not specifically request the survey, nor does the agency approve the collection or the collection procedures. However, if another agency gives the same researcher a grant to design and conduct a survey that the agency reviews and approves, then this collection would be covered by the PRA. Agencies are encouraged to discuss specific cases with their OMB desk officers prior to collecting the information to determine whether the collection is subject to OMB review under the PRA.

³ 5 C.F.R. § 1320.3(d)

⁴ 5 C.F.R. § 1320.3(d)

7. Are focus groups subject to PRA review?

There is no exemption for focus groups in the PRA. Agencies conducting focus groups must comply with the requirements detailed in 5 C.F.R. § 1320.3(c): “Collection of information means...the obtaining...of information by or for an agency by means of identical questions posed to, or identical reporting, record-keeping, or disclosure requirements imposed on, ten or more persons....” It then goes on to clarify “ten or more persons refers to the persons to whom a collection of information is addressed by the agency within any 12 month period.” Thus, focus groups are covered unless the total number of persons participating within a 12-month period is fewer than ten. For example, an agency conducting three focus groups of nine persons would be subject to the PRA because the total number of participants is greater than 10.

Although each focus group may not be asked the exact same questions in the same order, focus groups should be treated as information collections under the PRA if the same information is being sought from the groups. For example, an agency that is developing questions for a survey may convene a few focus groups in different areas of the country (or composed of people with different characteristics) and may have fairly wide ranging discussions on the topic of the survey in order to hear how the participants think about that topic and the vocabulary they use. Because the flow of discussion in the different groups may lead to different areas in more depth or at different points in the discussion, some parts of the protocol may not have been necessarily followed verbatim or may have occurred at a different point in one focus group than another. However, the same information was still being sought by the agency and the collection is subject to the PRA, regardless of whether the exact questions or probes were used or used in the exact same order with each group.

When agencies submit their ICRs for focus groups to OMB, they should include the protocols or scripts for the discussion. Agencies that routinely conduct focus groups as part of their development of questionnaires (e.g., pretesting) may find it useful to obtain a generic clearance for focus groups (see questions #8, #50, #51).

In addition to using focus groups for pretesting, an agency may conduct focus groups as part of its collection of other information and in conjunction with other methods of data collection as part of an overall research study. For example, some program participants may participate in a focus group as part of a program evaluation that also includes other collections, such as surveys of program administrators and staff. In these cases, it is important that the focus groups are included and described in the ICR in the context of the collection the agency is conducting so that OMB can appropriately evaluate the entire scope of the study and the practical utility of the information the agency will obtain. Thus, agencies should include the respondent burden associated with the focus groups in the ICR along with the protocols or script for the focus groups.

8. What are generic clearances and when are these useful for agencies?

A generic clearance is a plan for conducting more than one collection of information using very similar methods. The review of this plan occurs in two stages: (1) a full PRA review of the generic clearance ICR, which includes the general approach and methodology, at least once every three years, and (2) an expedited review of the individual collections that fall within the scope of the generic clearance. A generic clearance is considered only when the agency is able to demonstrate that there is a need for multiple, similar collections, but that the specifics of each collection cannot be determined until shortly before the data are to be collected.

Collections that are appropriate for consideration as generic include methodological tests, focus groups, or other pretesting activities (see question #51), as well as many customer satisfaction surveys. For example, an agency may want to use a “core” satisfaction survey with its many customer groups, but may want to customize the questionnaire for different groups by including some specific questions related to a particular service or publication they use.

Each collection under the generic clearance must be well defined in the overarching ICR approved by OMB in terms of its sample or respondent pool and research methodology, and each individual collection should clearly fit within the overall plan. Individual collections should not raise any substantive or policy issues or go beyond the methods specified in the generic ICR. Any individual collection that would require policy or methodological review is inappropriate for expedited review under the generic clearance and must go through the full PRA process. For example, a generic clearance is not appropriate for the collection of influential information (see question #18) and is probably not appropriate for large collections involving many respondents and high respondent burden. Agencies are encouraged to consult with their OMB desk officers before developing a generic clearance to determine whether their plans are appropriate for this type of clearance.

9. What needs to be done for an emergency clearance?

Agencies may submit an emergency ICR if the collection is both needed sooner than would be possible using normal procedures and is essential for the agency’s mission. In addition, the agency must demonstrate that the time to comply with the public comment provisions of the PRA would do any of the following: (1) result in public harm; (2) prevent the agency from responding to an unanticipated event; (3) prevent or disrupt the collection; or (4) cause the agency to miss a statutory or court-ordered deadline. This type of clearance should only be sought if the agency could not have reasonably foreseen the circumstances requiring collection; it is not a substitute for inadequate planning.

Agencies submitting an emergency ICR must publish a *Federal Register* notice stating the collection is being reviewed under emergency processing procedures unless OMB waives this publication requirement. The emergency ICR must contain all of the information that would be submitted with a normal ICR. Agencies must also specify the date by which they would like OMB to act on the ICR. Approval for an emergency collection is valid for a maximum of six months. If longer approval is needed, the agency must also initiate the normal PRA approval

process to take effect when the emergency clearance expires. Agencies are strongly encouraged to consult with their OMB desk officers prior to submitting a request for emergency clearance.

10. When do agencies need to complete Part B of the ICR Supporting Statement?

Agencies are instructed to complete Part B if they are using statistical methods, such as sampling, imputation, or other statistical estimation techniques; most research collections or program evaluations should also complete Part B.⁵ If an agency is planning to conduct a sample survey as part of its information collection, Part B of the ICR supporting statement must be completed, and an agency should also complete relevant portions of Part B when conducting a census survey (collections that are sent to the entire universe or population under study). For example, an agency doing a census of a small, well-defined population may not need to describe sampling procedures requested in Part B, but it should address what pretesting has taken place, what its data collection procedures are, how it will maximize response rates, and how it will deal with missing unit and item data.

Agencies conducting qualitative research studies or program evaluations, including case studies or focus groups, should also complete the relevant sections of Part B to provide a more complete description of the use of the information and the methods for collecting the information (see question #11).

11. Why do agencies need to complete some of Part B if they are conducting qualitative research studies or program evaluations?

Agencies need to specify how they plan to use the information they are collecting and how they will collect the information in order for OMB to properly evaluate an ICR that uses qualitative methods. There are elements of Part B that are not covered elsewhere in the justification that agencies should answer to appropriately describe the information collection. For example, an agency conducting case studies should specify in Part B:

- how the different sites and/or respondents will be selected,
- whether the agency intends to generalize beyond the specific sites and/or respondents selected,
- what pretesting has been done, and
- what different methods will be used to collect the information, e.g., in-person interviews, focus groups, observations, etc. and the protocols that will be followed to ensure high quality data are obtained.

In addition, as noted in questions #21 and #24, agencies will need to justify why they are not using statistical methods if their research questions are most appropriately addressed by a survey or other quantitative study.

⁵ See the instructions for supporting statements in Appendix A.

Useful Resources

Office of Management and Budget (August 1995). 5 C.F.R. § 1320 Controlling Paperwork Burdens on the Public; Regulatory Changes Reflecting Recodification of the Paperwork Reduction Act. *Federal Register*, 60, No. 167, 44978-44996.

SCOPE OF THE INFORMATION COLLECTION

This section addresses questions related to the content of the Information Collection Requests (ICRs) submitted to OMB. The Paperwork Reduction Act (PRA) requires agencies to demonstrate the practical utility of the information that they propose to collect and to balance this against the burden imposed on the public. Thus, it is critical that agencies describe the need for the information and how it will be used. Without a clear justification, OMB cannot approve the collection. The burden on the public must also be completely accounted for and minimized to the extent practicable while still yielding useful information. Again, agencies should consult the OMB regulations implementing the PRA (5 C.F.R. § 1320) for more detailed and complete information.

12. Why is this data collection necessary and how will the information be used?

The PRA requires that agencies address how the information the agency is proposing to collect is necessary for the performance of the functions of the agency. First, agencies should identify legal or administrative requirements that authorize the collection and should include copies of the authorizing statute and regulations in their ICRs. Second, agencies must also justify why the information is needed and how it furthers the agency's goals.

When appropriate, agencies should also highlight the knowledge gaps that the information collection is designed to address, including a brief review of existing information and the relevant scientific literature. If an agency proposes a research study or program evaluation that is designed to address specific research questions, it must demonstrate a direct connection between the information needs and the specific research questions. Thus, agencies should provide sufficient background information to support the need for the research questions (including a brief review of the relevant scientific literature) and how the study will meet that need. Agencies must also ensure that the collection does not duplicate other information accessible to the agency (but see question #47). If the information is a continuation of a prior collection, agencies should document how the information has been used and the continuing need for the collection.

The PRA also requires that the agency demonstrate the practical utility of the collection and the use the agency will make of the information. The supporting statement should always include a careful discussion of what the agency hopes to achieve by collecting the information and the quality of information that will be obtained employing the proposed design.

Agencies must also evaluate their intended use of information from the proposed collection in light of the OMB's information quality guidelines for utility, integrity, and objectivity⁶ as well as the agency's information quality guidelines. Based on that evaluation, agencies should be able to state in their ICRs that the proposed collection of information will result in information that will be collected, maintained, and used in a way consistent with the OMB and agency information quality guidelines (also see question #18).

⁶ Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by Federal agencies, 67 **FR** 8452-8460

13. How often should data be collected?

When submitting an information collection request (ICR) to OMB, agencies are required to describe the consequences of collecting data less frequently than proposed. While less frequent data collection reduces burden on the public, more frequent data collection can provide more current and useful information. For example, in longitudinal and panel surveys, more frequent collections allow for shorter reference periods between reports, which may reduce bias. The goal is to strike a balance between the need for current information and the need to reduce public reporting burden.

Most Federal data collections fall into one of two categories: continuing or one-time. Continuing data collections have an established frequency of collection (monthly, quarterly, annually, biannually, etc.). A one-time collection is conducted without the intention of collecting the same information again, or without an established collection pattern. The frequency of data collection is an issue when establishing a new continuing data collection, when renewing a continuing data collection, or when repeating a prior one-time survey.

When determining the frequency of data collection, the agency should consider the following:

- *The timeliness of estimates requires high frequency collections.* For example, the monthly unemployment rate is a key economic indicator, and the data must be collected monthly. Some collections are required by law to be collected at specific frequencies, e.g., the Decennial Census occurs every 10 years, and the Economic Censuses are conducted every 5 years (for years ending in 2 and 7).
- *There is a reasonable expectation of significant change in key statistics between collections.* For example, an agency may wish to conduct a customer satisfaction survey every year; however, if the agency has not made any changes in its programs, there would be no expectation for change. If the agency has started a new customer relations program, then a repeat of the customer satisfaction survey could be used to measure the effectiveness of that change. Another consideration in evaluating the frequency of a collection is the potential for seasonal variation. The need to capture cyclical patterns might justify either monthly or quarterly collection.
- *The frequency of collection has an effect on data quality.* For example, the Survey of Income and Program Participation (SIPP) is a longitudinal survey that captures a month-by-month accounting of income and governmental transfers such as Social Security, welfare, food stamps, etc. Pretesting of two different collection periods showed significant differences in the data quality between the three-month and six-month time frames, requiring the use of a shorter period. SIPP adopted a four-month time frame for data collection and reference period, which provided nearly the same quality of data as the three-month time frame with a 25 percent reduction in respondent burden.
- *Reduced frequency would have an adverse impact on agency programs.* If an agency program requires data with a specified frequency, the agency needs to detail how the data will be used and how the agency would be hindered by less frequent information.

14. What is included in the calculation of burden hours?

Burden hours are a measure of the time it takes respondents to review instructions, search data sources, complete and review their responses, and transmit or disclose information. Estimating burden for household surveys is typically done by timing the completion of interviews done in previous administrations of the survey or in pretests (using 9 or fewer persons) and developing an average time.

Estimating burden for establishment surveys is more complicated because respondents often have to search for information before answering the survey questions. Agencies must first identify all the steps a respondent takes in order to comply with the survey request, and then estimate the time for each step to arrive at a total burden per respondent. The aggregate burden of an ICR is the average burden per respondent multiplied by the number of expected respondents and should be reported in section A.12 of the ICR.

15. For establishment surveys or panel surveys, should burden hours include the original collection of administrative records that may have taken place months or years before?

Generally, surveys of business establishments ask a respondent to aggregate and report data that the establishment already has somewhere in files or databases. Burden hours for these surveys should include only the time it takes to locate the source data and aggregate them. The estimate should not include the time originally taken to collect information in administrative records that were compiled by the establishment for its own purposes, such as accounting records. For example, there are a variety of reporting and recordkeeping requirements in the equal employment opportunity arena. These reports usually ask for summary demographic and job data on employees, and respondents often obtain the data needed from existing personnel files, records, or databases. Agencies **SHOULD NOT** count the time involved in the original collection of the demographic data from the employees but **SHOULD** count the time it takes to access the personnel files, aggregate the requested data, and report the data on the agency form.

For panel or longitudinal surveys, agencies **SHOULD** count the time it takes respondents to begin their participation in a panel in the initial ICR for the recruitment and baseline collection. However, this time **SHOULD NOT** be counted in subsequent ICRs that concern later collections. Agencies **SHOULD** count only the hours associated with the collection of information described in the current ICR. For example, the Survey of Income and Program Participation selects respondents to participate in interviews every four months (called waves) for the duration of a panel—usually 3 or 4 years. Each wave has a set of core questions used in all waves and a topical module that differs from one wave to the next. In essence, each wave is treated as a unique survey and the burden associated with answering all the questions in a wave is reported. In this case, the agency **SHOULD** count the burden of recruitment and the initial collection in the ICR for wave 1; however, the agency **SHOULD NOT** count the original recruitment of the individual into the survey panel in the ICRs for later waves.

16. Why are agencies required to estimate the burden in terms of both time and costs?

The term "burden" means the "time, effort, or financial resources" the public expends to provide information to or for a Federal agency, or otherwise fulfill statutory or regulatory requirements.⁷ Currently, agencies separately estimate the "hour burden" and "cost burden" of each particular information collection in their supporting statements in A.12 and A.13, respectively. This ensures that both types of burden are taken into account.

Thus, for establishment surveys, in addition to the hour burden for reviewing instructions, searching data sources, completing and reviewing responses, and transmitting or disclosing information, there may also be capital, operation, and maintenance costs associated with generating and maintaining the information. Agencies should include costs that respondents incur for developing, acquiring, installing, and utilizing technology and systems for the purposes of collecting, validating, verifying, processing, maintaining, disclosing, and providing information, as well as costs incurred by respondents adjusting to changes from previous instructions, and training personnel to be able to respond to a collection. These costs may be borne directly by the respondent or indirectly by their subordinates, agents, or contractors.

The PRA requires that the agency demonstrate the practical utility of the collection and demonstrate that the burden of the collection both in terms of hours and other costs is justified given the agency's need for the information and the use the agency will make of the information.

Useful Resources

Office of Management and Budget (August 1995). 5 C.F.R. § 1320 Controlling Paperwork Burdens on the Public; Regulatory Changes Reflecting Recodification of the Paperwork Reduction Act. *Federal Register*, 60, No. 167, 44978-44996.

⁷ 44 U.S.C. § 3502(2); 5 C.F.R. 1320.3(b).

CHOICE OF METHODS

This section is intended as a broad overview of many specialized methodologies. One can often find entire textbooks devoted to one or more methods; thus, the purpose of this section is simply to call attention to some basic considerations agencies should explain and justify in their ICRs when proposing to conduct studies that use these methods. The method selected must also be appropriate for the intended use of the information. Agencies should consult with experts in the particular methods to design and implement their studies.

17. How does the choice of methods for the study relate to the research questions or purpose of the collection?

The methodology for the study should be driven by the kinds of questions the agency needs to answer or the general purpose of the collection. Sometimes agencies collect information for general statistical purposes that may be used by a wide variety of different parties to address many different questions. In this case, the design of the survey or study should reflect these multiple uses and be clear about its strengths and limitations for different purposes, and agencies should consult with appropriate stakeholders and experts when designing their studies to ensure the relevant questions are addressed. In other cases, agencies need to answer very specific questions, and the design needs to be appropriately focused to answer those questions well.

Agencies should carefully consider the kinds of questions the information collection needs to answer and the strengths and limitations of different methods to answer those questions. For example, if an agency wishes to know whether a program caused some change to occur in those served by the program, appropriate methods, such as an experimental design, will need to be employed. In this case, agencies will need to do considerable advance planning to randomly assign participants to experimental or control conditions to evaluate the program. If an experimental design is not possible or practical, then a quasi-experimental design or other design may be used by the agency. Agencies need to justify how their choice of methodology will be able to provide the information needed to address the research question and discuss the limitations as well as the strengths of the methodology for the particular purpose (see question #26).

In order to address complex and multi-faceted research questions, an agency may need to plan a program of research and use multi-method approaches to obtain all the information needed. When more than a single study is planned to address the research questions, the agency should include a brief description of the complete research program (including studies not yet approved) in its ICRs and refer to previously approved collections (by their OMB number) to explain how the study fits into the larger program and support how the complete program will provide the agency with the information it needs. If the agency is supplementing its survey collection with administrative or other available data, this should also be described. In addition, the general timeframe for the other components should be included.

18. How do the use of the information and choice of methods for the study relate to the agency's information quality guidelines?

A primary consideration for an agency in designing its information collections should be how the agency intends to use the information it is gathering. Agencies use information collected through surveys for a variety of purposes. Some information is intended to be “influential information.” As defined in OMB and agency Information Quality Guidelines, “influential” means that “an agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions.” The Information Quality Guidelines require that agencies hold the information they designate as “influential” to a higher standard of reproducibility and transparency than information that is not defined as influential. For example, some survey results directly or indirectly feed into Principal Federal Economic Indicators that are widely watched and have broad impact on government, business, and individual decisions. In other situations, one agency may use the information collected by another agency to support health and safety assessments that in turn affect both public and private sector decisions.

As part of their ICRs, agencies report how they intend to use the information they are proposing to gather. Agencies should explain how the methods they have chosen to employ will yield information of sufficient quality for its intended purpose. For example, if an agency wishes to generalize the results of a survey beyond the particular cases sampled, it must utilize appropriate statistical sampling methods (see question #30) to yield information that has sufficient precision and accuracy (see question #33). Because more rigorous methods often entail higher cost, agencies need to carefully consider the resources that will be required to obtain information of sufficient quality for the intended uses. Agencies should be able to certify explicitly in their ICRs that the proposed collection of information will result in information that will be collected, maintained, and used in a way consistent with the OMB and agency information quality guidelines, or they should not propose to collect the information.

19. When should agencies consider conducting a survey?

When the research question or purpose of the study is to produce descriptive information about a population, agencies should consider conducting a survey. Surveys may be conducted to provide general purpose statistics on the national (or some target) population, or they may be used as part of a research study, experiment, or program evaluation. For example, an evaluation of a federally funded school program may be done by conducting surveys of school principals, teachers, and district administrators to obtain information from each about the implementation or results of the program. However, surveys are often only one source of information that an agency may need, especially when conducting program evaluations. Agencies should also examine how they can obtain other appropriate outcome measures, including the use of administrative records.

When properly done with an appropriate sample design, a survey can provide broad descriptive information about a population and subgroups, as well as information about relationships among variables or constructs that are being measured. Generally, the results from surveys are only

descriptive or correlational. When surveys are used in the context of an experimental design, quasi-experimental design, or longitudinal study, stronger causal inferences may be warranted; however, agencies will need to carefully consider the limitations of the study and other potential explanations when drawing causal conclusions.

Because they are designed to gather standardized information from an often relatively large number of persons or entities, surveys may not be able to provide the degree of detail that can be obtained through qualitative or case study methods. Furthermore, the standardization of questions requires that the concepts that are being measured be well known and understood, and shown to be reliable and valid. Thus, it may be premature to conduct a survey when an agency is in a more exploratory mode, trying to develop research questions, or understand the characteristics that need to be measured. It is not appropriate for agencies to conduct developmental activities to define a concept and then attempt to use those same findings to test hypotheses (see question #21). A separate survey is needed to test the hypothesis.

20. What should agencies consider when designing and conducting a survey?

The quality of a survey design can be judged by the strategies that are taken to prevent, adjust for, and measure potential problems and sources of error in surveys. How well a survey is designed and conducted can lead to either more or less variance (or noise) or bias (or systematic errors) in results. Well-designed and conducted surveys anticipate potential problems and try to prevent or minimize the impact of different sources of error as much as possible. Additionally, good surveys make efforts to measure and adjust for errors that are not controlled. The best surveys are those that check and verify each step of the research process. Common sources of error in surveys include sampling (due to measuring only a subset of the population), coverage (due to mismatches between the population and the lists used to draw the sample), nonresponse (due to failure to measure some sampled units), measurement (due to mismatches between data sought and data provided), and processing (due to editing or imputation). These topics are dealt with in greater detail in the following sections of this guidance.

For example, measurement errors can be reduced through careful questionnaire design and pretesting (see Questionnaire Design and Development). A field test comparing alternative versions (or revised versions) of key questions may provide insights into sensitivity of answers to alternative wording (see questions #22, #23, and #49). Agencies can also reinterview a subsample of respondents to measure instability in responses. Sometimes, survey results can also be checked against administrative records; however, there may be differences in definition and coverage between the information available from records and the survey that need to be carefully considered when assessing the results of the comparison. Similarly, potential nonresponse bias can be reduced by following a variety of strategies to maximize response rates or repair imbalances in the respondent pool (see questions #69 and #70). Bias can be measured in special nonresponse bias studies (see question #71) and adjustments can be made to weights to attempt to reduce bias.

Agencies designing and conducting surveys need to consider all of the potential sources of errors and plan to adequately prevent, measure, and adjust for them. Conducting a high quality survey

requires careful planning and sufficient resources to yield quality data that have practical utility for the agency. Agencies should carefully document and justify the adequacy of their survey methods in their ICRs. Specifically, agencies should provide information about the target population, the sampling frame used and its coverage of the target population, the design of the sample (including any stratification or clustering), the size of the sample and the precision needed for key estimates, the expected response rate (see question #63), the expected item non-response rate for critical questions, the exact wording and sequence of questions and other information provided to respondents, data collection methods and procedures, and the training of interviewers (if applicable). In addition, agencies need to take into account what is known about the different sources of error in their analysis and interpretation of the results from the survey. Experts in survey methodology within and outside the agencies can be helpful to inform this process. Agencies should be transparent and report in their ICRs the methods they plan to use, what is known about the different sources of error, and the impact of the errors on the analytic results.⁸

21. When should agencies consider conducting a qualitative study?

An agency may want to consider a qualitative study under a variety of circumstances. In contrast to gathering numerical information or data that can be quantified, a qualitative study uses unstructured interviews, notes, or observations that are typically difficult to quantify. Qualitative studies can be useful for exploratory investigations such as when very little is known about a problem or the implementation of a program. A qualitative study in this case may be a good first step to understanding the scope of a problem or identifying the key issues for more systematic study. A variety of methods may be used in a qualitative study, including focus groups, unstructured interviews, or semi-structured interviews with “experts,” stakeholders, or other participants. Case studies may also be conducted (see question #24). Typically, these methods attempt to obtain insights through the intensive study of a relatively small number of people, institutions, or establishments. Respondents are usually purposively chosen because of their knowledge, experience, or status.

In a qualitative study, typically, different persons or entities may be chosen because they “represent” a particular kind of person or entity, but the sample is usually not representative—in a statistical sense—of any larger population. However, the obtained information may be very useful in generating hypotheses that can be tested more systematically with other methods such as quantitative surveys. Sometimes qualitative studies are done in conjunction with or as a component of a larger quantitative study to obtain further insights or context for the results; however, these qualitative interpretations can be prone to misinterpretation and over-generalization. Although qualitative studies can also be done using statistical sampling (see question #30) and rigorous designs to generalize results, this is rarely done.

Agencies should demonstrate how a qualitative study will meet their information needs. Agencies need to acknowledge the limitations of data gathered using these methods and not generalize the data beyond those persons or entities that were interviewed. These studies should

⁸ For further information see *Statistical Policy Working Paper #31, Measuring and Reporting Sources of Error in Surveys* available at www.fcsn.gov/reports/.

usually be considered preliminary, and will often need to be followed with a larger-scale, representative study.

22. When should agencies conduct a pilot study, pretest, or field test?

Agencies should always consider conducting pretests (small trials of the measurement process) or pilot studies (larger trials yielding statistical information) when planning for a new information collection or changing methods and procedures for an ongoing survey. These kinds of tests may provide critical information necessary to ensure the quality of the data and smoothness of operations needed in the full-scale information collection. They can provide essential information to the agency and result in higher data quality than would have been achieved without them and may be the only vehicle for measuring the effects of different changes an agency is considering implementing. Thus, agencies will need to weigh the importance and use of pretests against the time and resources needed to conduct them.

Pilot studies can be useful when there are a number of issues the agency needs more information about before a full-scale study can be reasonably implemented. A pilot study may help an agency narrow down the research questions or provide rough estimates (and variances) that can be used to guide sample size determinations. An agency may also use a pilot study to examine potential methodological issues and decide upon a strategy for the main study. A pilot test may also be conducted before a large-scale study in order to test and refine the implementation procedures for the full-scale study.

Agencies may want to conduct pretests when developing new questionnaires to see how respondents actually answer questions and identify potential data quality problems, such as high item nonresponse rates. Agencies may also conduct pretests to gather data to refine questionnaire items and scales and assess reliability or validity. Sometimes agencies may also use a field test or experiment (a study to compare the effects of two or more procedures or questionnaires) when planning a change in methodology or questions in an ongoing survey. This enables comparisons and often provides quantifiable data to decide among the different methods or questions to use. An agency may further want to consider conducting a field test experiment on a representative sample to measure the effect of the change in methods or questions on resulting estimates.

Agencies can request clearance for pretests, pilot studies, or field tests separately or as part of their ICR for the full-scale collection (also see questions #50 and #51). However, in many cases it makes more sense for these to be separate requests, especially when conducting pilot studies for new collections. Agencies are encouraged to discuss whether it is appropriate to submit these studies separately or in combination with the full-scale study with their OMB desk officers prior to submitting the ICR to OMB.

23. When should agencies consider conducting focus groups or cognitive interviews?

Agencies should consider using focus groups or cognitive interviews when planning for a new information collection or when altering questions on an ongoing survey. Developing effective new questions or revising existing questions can be more difficult than most people anticipate, and questions need to be constructed so that respondents can answer them and provide useful data for the agency.

Focus groups (groups of 8-12 persons engaged in a semi-structured conversation led by a moderator) can be a useful first step in questionnaire development that can allow an agency to better understand what respondents think about a topic and what terms they use. Agencies can learn the language that respondents use when discussing the topic and as a result integrate more common terms and phrases into the design of survey questions.

Focus groups are also often used as part of a case study or in conjunction with a sample survey or program evaluation to gain insights and perspectives on the operation of a program or to provide more detailed information to help illustrate the results from the survey (see question #24).

In a cognitive interview, respondents are asked to think aloud as they answer questions and to identify anything that confuses them. Cognitive interviews can be a valuable tool when an agency has developed proposed questions and needs to understand better how respondents interpret them. Respondents are often asked to paraphrase a question so that researchers learn whether a respondent understands the question and interprets it as intended. Good questionnaire development is aided by survey methodologists who are trained in these methods. Further information on methods for developing questionnaires is provided in the section on Questionnaire Design and Development (questions #45 to #51).

24. When should agencies consider using case study methods?

A case study is a research methodology that is widely used in a variety of contexts. One good definition of a case study is:

a method for learning about a complex instance, based on a comprehensive understanding of that instance obtained by extensive description and analysis of that instance taken as a whole and in its context. (GAO 1990, p. 14)

Case studies can be useful when one is dealing with a complex program about which relatively little is known or understood. A case study can thus serve a useful purpose as a preliminary study for the agency to learn some of the characteristics of how the program is implemented or operating or what its possible effects might be. This can be a useful *first step* in evaluating a program because it leads to the generation of hypotheses about the program and its implementation, as well as a preliminary assessment of how more systematic research can be designed and implemented to evaluate the program.

Case studies can also provide important insights when used in conjunction with other research methods such as sample surveys. For example, an agency may conduct a large representative sample survey of program sites to gain knowledge about their characteristics. A number of sites may also be selected for case studies to help provide additional understanding about the way the program functions or is implemented, and thus illuminate the quantitative results from the survey. These case studies may, for example, include direct observational components that are not feasible in a large scale national study.

More specifically, case studies can provide vital insights about how programs are implemented in different local areas. For programs that deliver their services through State and local agencies, the Federal Government often sets general standards regarding administration, evaluation, and funding. Developing a comprehensive picture of how a federally-regulated program is administered, for example, may require site-specific observation and investigation. Data from specific sites can serve several purposes depending on the study design including:

- developing explanatory hypotheses on program characteristics and outcomes, which can be tested in future statistical studies;
- preparing guidance for field offices on how services may be delivered more effectively;
- providing qualitative explanatory information on the range of program characteristics and outcomes, which complement quantitative results obtained through a statistically valid, generalizable study; and
- illustrating findings of the main study through real-world examples.

25. What should agencies consider when designing and conducting a case study?

There are a number of limitations of the case study method that agencies should consider. In some situations, these limitations can make it difficult to conduct the research. In others, they can make it difficult to generalize the results. Limitations include:

- the case study sites are typically not selected in a manner that allows one to generalize to the population under study;
- too few sites are typically visited to get a comprehensive or generalizable picture;
- results observed at a site may be due to other factors besides the program being studied, and there is often no control group or randomized assignment to the program;
- site visits are expensive; they require significant travel and preparation costs; and
- data from site visits are often qualitative and anecdotal in nature.

When designing or evaluating a case study, the following questions should be considered:

- *Who is conducting the case study?* The role of the investigator is very prominent in case study methods and the training, experience, and thoroughness of the investigators visiting a site can have a large impact on the quality of the data that are gathered.
- *How are the sites selected?* How sites are chosen will have direct implications for the kinds of conclusions that can be drawn from the research. Although probability methods are essential for generalizable survey samples (see question #30), the small samples that are typically used in case studies cannot usually be meaningfully generalized to any population. However, the results from case studies are typically not intended to describe

the population in the same way as those from sample surveys. It may be useful to select cases to represent the diversity and important variations of programs that exist.

- *How well do the protocols or questionnaires elicit the desired information?* As much thought and effort should go into the design of questionnaires and protocols for case studies as goes into these instruments when administered to a national sample. Careful attention also needs to be paid to who will be interviewed as part of the case study. Similarly, other sources of information such as observations by researchers, examination of administrative records, and other documentation are often important components of the case study.
- *Is interviewer bias and interpretation being minimized?* If the data are only qualitative, they may be subject to interviewer interpretation and bias. To the greatest extent possible, guides for onsite visits and data collection as well as instructions for coding and analyzing the data should be developed beforehand. Visiting one or two sites for a pretest is also highly recommended because the actual site visits can reveal the complexity and difficulty of analyzing case study data.
- *How will the data be analyzed?* Some analysis usually needs to take place in real time at the site to resolve discrepancies or take advantage of the multiple perspectives offered by the different investigators on site. Objective data that have been gathered should be quantified and displayed with basic descriptive statistics. It is unlikely that inferential statistics or hypothesis testing could be used unless sample sizes are adequate and sites were selected appropriately to generalize.
- *What is the relevant comparison group?* Case study research may include comparisons between a program site and a “comparable” site that did not have the program. Unless very strict controls are in place, it is difficult to have a true comparison site for most case studies due to the unknown influences of other factors that could affect observed differences between the sites. The differences between sites that cannot be controlled can, however, often be articulated in advance and need to be carefully considered as limitations to comparisons; nonetheless, having a comparison group may provide more information than if there is no comparison group. Alternatively, case study designs may include baseline pre-program measurements of persons and post-program measurements of the same individuals.

26. When should agencies consider using experimental and quasi-experimental designs?

When agency research questions involve trying to determine whether there is a causal relationship between two variables or whether a program caused a change for participants, then agencies will need to employ an experimental or quasi-experimental design or demonstrate how their study design will allow them to determine causality. Even well conducted experimental and quasi-experimental designs may have limitations or alternative explanations for the results that the agency will need to carefully consider in designing the study and drawing conclusions from the results.

For example, it can often be difficult to identify appropriate comparison groups to evaluate the impact of Federal programs or interventions. If an agency wishes to evaluate a new education program that provided some districts or schools with competitive grants to implement the

program, it can be difficult to identify similar districts that are an appropriate comparison to gauge the effects of the program on student outcomes. If the comparison schools or districts differed systematically from those that received program funding then it is not clear whether any differences in student outcomes are due to the program or to the preexisting differences such as urbanicity or poverty level. In addition, sometimes, schools or districts that win (or even apply for) competitive grants may be more interested, motivated, or have greater capabilities for improving student outcomes than schools or districts that don't win (or apply) for the program grants, and the student outcomes may reflect the underlying motivation or capabilities rather than anything about the program itself. Thus, the agency needs to consider appropriate methods to select comparison schools or districts that will rule out or minimize alternative explanations for differences in student outcomes in order to maximize the value of the program evaluation.

One of the key characteristics of experimental designs is random assignment of persons or entities to treatment (or experimental) and control (or comparison) conditions. For example, participants in the treatment condition may receive benefits or services from a Federal program, while participants in the control condition do not. This random assignment of persons to conditions acts to equalize preexisting differences between the two groups so that differences observed between the groups can be attributed to the differences in the Federal program. If random assignment is not strictly possible, then quasi-experimental designs can be employed. These designs rely on identifying appropriate comparison groups and frequently take measurements at two or more points in time in order to rule out or reduce threats to the validity of the conclusions or alternative explanations for differences between the experimental and comparison groups.

Different kinds of experimental designs may be used by an agency depending on the research questions or the types of decisions the agency intends to make based on the results. Sometimes the goal may be simply to assess whether a new demonstration program is having the intended effect, before investing additional resources to expand the program and study it further. In this case, it may be possible for an agency to justify using a sample that is not nationally representative or even representative of potential program members. In other cases, the agency may want to estimate the size of the effect a specific intervention would have if implemented throughout the country or evaluate the effectiveness of an ongoing program. In these circumstances, the agency would need a representative sample of program participants in order to accurately describe the population and generalize the results to the rest of the country or to all program sites (see section on Sampling).

Agencies need to consider the difficulties of implementing experimental designs and guard against potential threats to the internal validity of the design through choice of appropriate comparison groups and/or conducting multiple measurements over time. It is key that agencies design and implement programs in ways that they can be meaningfully evaluated. For example, conducting an experimental study requires advance planning so that participants can be assigned to conditions. Agencies should justify that the design they have chosen is practical to conduct and will provide the information they need to answer the agency's research questions. Agencies also need to acknowledge the limitations of their design and to identify clearly how they intend to generalize the results of experimental studies, especially if a representative sample is not proposed.

Useful Resources

Experimental and Quasi-Experimental Designs

Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Hedrick, T. E., Bickman, L. & Rog, D. J. (1993). *Applied Research Design: A Practical Guide*. Newbury Park, CA: Sage.

Program Evaluation Designs

Rossi, P.H. & Freeman, H.E (1993). *Evaluation: A Systematic Approach*. Newbury Park, CA: Sage.

U.S. General Accounting Office (March, 1991). *Designing Evaluations*. GAO/PEMD-10.1.4. Washington, DC: Government Printing Office.

U.S. Office of Management and Budget (2004). What constitutes strong evidence of program effectiveness? (http://www.whitehouse.gov/omb/part/2004_program_eval.pdf)

Qualitative and Case Studies

U.S. General Accounting Office (November, 1990). *Case Study Evaluations*. GAO/PEMD-10.1.9. Washington, DC: Government Printing Office.

Yin, Robert K. (1989). *Case Study Research: Design and Methods*. Beverly Hills, CA: Sage.

Surveys

Biemer, P. P. and Lyberg, L. E. (2003). *Introduction to Survey Quality*. New York: Wiley.

Fowler, F. J. (1988). *Survey Research Methods*. Newbury Park, CA: Sage.

Groves, R. M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. Hoboken, NJ:Wiley.

SAMPLING

This section is intended as a broad overview of some key issues in survey sampling. Designing an appropriate sample for a given purpose and target population requires considerable technical expertise, and agencies will need to consult with statisticians and experts in survey sampling in designing their studies. The purpose of this section is to provide a general non-technical introduction to some of the concepts of survey sampling that agencies will need to describe and justify in their Information Collection Requests (ICRs) when proposing to do studies whose purpose is statistical in nature.

27. What is the difference between a census and a sample survey and when is each appropriate?

A study where all target population members are asked to participate is often called a universe survey or a census. In contrast, a *sample survey* is a survey where only a portion of the population of interest is included in the study; that is, only a selected number of households (or businesses) are asked to participate rather than including all members of the population. Furthermore, the members of the target population must be selected with a known probability of selection from a sampling frame that contains all (or nearly all) of the members of the target population.

When the target population is small and each unit is unique, a census is likely to be preferred over a sample survey. For example, when an agency evaluates a Federal program that is implemented by the states (each one perhaps somewhat differently), a census of state program directors may provide higher quality information with little cost difference from a sample survey of a slightly smaller number of states. In this case, there may also be concerns about missing practices of some states that were not included in the sample if a census were not conducted.

Sample surveys are useful when it is not possible or desirable to collect data from every single member of the population of interest due to reasons such as respondent burden, cost, and operational feasibility. Often it would be simply too burdensome, expensive, or logistically impractical to collect data from every single unit of the target population. Agencies should consider collecting data from a sample and trying to ensure a high response rate from the sampled units. For a given budget, an agency can devote more resources to quality control activities such as callbacks to nonrespondents and data editing for a sample survey than would be possible with a census, and the results from the sample survey should be more representative and provide less biased estimates of the population than a poorly conducted census.

Agencies should carefully consider the benefits and costs of conducting a sample survey versus a census. When the data must be representative of the target population, carefully designed samples can be used to ensure data quality in a way that is often more economical and efficient than a census. Agencies need to justify in Part B of their ICRs their decision to conduct a census instead of a sample survey.

28. What is a sampling frame and what is the coverage of the sampling frame?

A sampling frame is a list or set of procedures for identifying all elements of a target population. In theory, the sampling frame should include everyone in the target population as well as other information that will be used in the sampling process or can be used to assess the representativeness of the sample. There are different types of sampling frames, e.g., area and list frames. In an area frame, geographic areas are defined, listed, and then sampled. Often, lists of elements (e.g., housing units) are constructed within the sampled areas and then elements are selected from the lists. In a list frame, a list of all the population elements is used to select the sample directly. Sampling frames also may include information on characteristics of the elements, such as employment levels for a business or enrollment for schools.

Sampling frames should be up to date and accurate. The coverage of the sampling frame refers to how well the frame matches the target population. For example, approximately 97 percent of U.S. households have land-based telephone lines; therefore, a frame of all residential telephone numbers would have a national coverage rate of 97 percent.⁹ However, there are systematic differences between households with and without telephones, so that telephone coverage rates for some target populations such as the poor, young adults, and racial or ethnic minorities are often much lower and may not be adequate for some purposes. When those subgroups differ from others on key survey variables, coverage error in the survey estimates can result.

The coverage of a sampling frame can change over time and, therefore, it needs to be kept current and accurate. A list of business establishments that is two or three years old will not include any new businesses formed in the past two to three years but will include establishments that have gone out of business, and also may have incorrect contact information for those that have relocated. The availability and accuracy of contact information for sample units within the frame may affect the agency's choice of mode of data collection. In addition, the availability and accuracy of information for stratification is also an important consideration for choosing a frame.

Agencies need to consider the adequacy of potential sampling frames for their target population and should justify in Part B of their ICRs the frame they have chosen for their collection, its coverage, the mechanism for updating, how recently it has been updated, and what is done to assess or adjust for potential coverage errors.

29. Is a list of Internet subscribers available and acceptable for use as a sampling frame?

There currently are no unduplicated lists of Internet users from which to draw a probability sample (see question #30). In other words, there is no sampling frame available for Internet users or those with Internet access. Furthermore, unlike telephone numbers, there is no set format for e-mail addresses that could be used to generate meaningful addresses to construct a sampling frame for those addresses. Currently, lists of e-mail addresses that are commercially available tend to have unknown coverage for most target populations or consist of persons who

⁹ Blumberg, S., Cynamon, M., Lake, J., & Frankel, M. (2006). Recent trends in household telephone coverage in the United States. Paper presented at the Second International Conference on Telephone Survey Methodology, Miami, Florida.

have self-selected or volunteered to participate in studies; thus, these lists represent convenience samples (see question #35).

Recent estimates are that more than 50 percent of households have Internet access at home.¹⁰ Despite the increasing rate of Internet access in the U.S., there remain systematic differences in socio-demographic characteristics between those who have access to the Internet at home and those who do not. Thus, there are significant coverage errors in any sampling frame composed only of those who have access to the Internet, which could lead to biased estimates when generalizing to the national population.¹¹

In some cases, an agency may have e-mail addresses from its list frame for the target population that could be used for a census or sample survey. For example, administrative records of program participants may include a variety of means of contacting participants including their e-mail addresses. In this case, the coverage of the sampling frame is based on the characteristics of the frame the agency has and the specific target population; it does not use or require an Internet sampling frame.

The limitations for coverage and sampling of current lists of Internet users means that agencies should consider using any Internet sampling frame only for exploratory purposes, such as part of a pretest (if the main study will have a response option via the Internet), or in other instances where a convenience sample would be appropriate (see question #35). However, these limitations of the Internet for sampling do not imply that the Internet cannot be used as one mode of collecting survey data in a mixed-mode collection (see Modes of Collection), but rather that it is not suitable for drawing a probability sample that can be generalized to a target population. When used simply as a mode of collection, Internet surveys can provide a convenient means for respondents with Internet access to respond to a survey. Using the Internet simply as a mode of data collection, rather than as a sampling frame, is further addressed in question #43.

30. What is an appropriate sample design to ensure the sample drawn represents the population of interest?

When a subset of the population is chosen randomly such that each unit has a known nonzero probability of selection, the sample is called a probability sample. For the purpose of making estimates with measurable sampling error that represent a population, the sample must be selected using probability methods (however, also see question #31 for a discussion of cut-off samples that are able to measure estimation error). These methods require that each case in the population has some known nonzero probability of being included in the sample. For example, an agency can randomly select a sample of 500 customers from a complete list of 10,000 customers by drawing their names out of a hat. This is commonly referred to as a simple random sample (SRS). In a simple random sample, every case in the population (i.e., each of the 10,000

10 National Telecommunications and Information Administration (NTIA) (2004). *A Nation Online: Entering the Broadband Age*. Washington, DC. This and earlier reports available online at www.ntia.doc.gov/reports/anol/index.html.

11 These coverage problems do not necessarily apply to panels or other studies that use some other sampling frame (such as RDD) to recruit panel members and then provide them with Internet access, see question #34.

customers) has the same probability of being selected. Although SRS is rarely used in practice, there are other probability methods that may involve stratifying and/or clustering the sample or involve unequal probabilities of selection (e.g., a design that intentionally oversamples minorities or includes with certainty large businesses that account for a high volume) that are often used in the design of Federal surveys (see question #32). As long as there is a probability mechanism used in selecting the cases (and every unit is given a nonzero chance of selection), samples constructed in this manner can allow the agency to estimate the characteristics of the population from which they were drawn with a known level of sampling error. Non-probability samples do not have this property.

When selecting a sample design, agencies need to consider how the information will be used and what generalizations are intended, and agencies need to explain in their ICRs how they will generalize the results of a survey. Agencies must have a statistical basis for generalizing the results beyond the particular sample selected and need to consult a sampling statistician in designing their sample for their survey. Agencies conducting surveys that are intended to produce valid and reliable results that can be generalized to the universe of study, but are not based on probability methods, must clearly justify the statistical methodology (e.g., see question #31) in Part B of the ICR. Otherwise, OMB cannot approve the collection.¹²

31. Are probability samples always the best for surveys of establishments?

Although a probability sample drawn from the general population is the best way to represent a population of individuals or households, it can be more efficacious to employ other sampling methods, such as cut-off samples, when the target population is businesses or other highly skewed populations. *Cut-off samples* are selected by ordering the universe of potential respondents by some important characteristic and selecting the units with the greatest amount of the characteristic until some specified percentage of the universe is included in the sample. A rule of thumb often used for cut-off samples is that the sample should cover 80 percent of the population total. This method gives an achieved sample that provides the minimum mean square error estimate for the total value of the variable used to specify the coverage. For highly skewed populations, such as those found in some establishment surveys, this method also provides the smallest possible sample. For example, an agency conducting a study of capital expenditures of manufacturers may “cut off” when the survey has received data from establishments with more than 80 percent of the revenues of the universe. Since the cutoff rule is based generally on estimates from a prior time period, the success of the cutoff rule is dependent on the level of stability in the estimates over time. In conjunction with a ratio based on a recent census survey of the population, this method is efficient, reduces respondent burden, and works well for estimating totals. However, it can be misleading if detail is needed on the smaller units, because they are more likely to be excluded from the sample.

Cut-off or other model-based samples are used for some economic surveys conducted by Federal agencies. Designing and using these samples requires that agencies have considerable information about the target population and statistical expertise in order to achieve estimates with smaller errors and biases than would be possible with a probability sample of the same size.

¹² 5 C.F.R. § 1320.5(d)(2)(v).

When the goal of the collection is to make an estimate for a target population, agencies need to provide a statistical justification in the ICR for using cut-off or other model-based samples that demonstrates that estimates of precision can be calculated and that the error of the estimates and potential biases are acceptably small.

32. What information should agencies provide about their complex sample designs?

Simple random samples (where all units and all equal-numbered combinations of units have the same probabilities of selection) are rare in practice for a number of reasons. Often they are not practical for many information collections because the sheer size of a universe listing and subsequent random sampling may be cost prohibitive. For example, it may be impractical for an agency wishing to survey and administer tests to high school students to select a simple random sample of students because there is not a comprehensive listing of all students in the United States, and even if there were, the costs of administering tests across the many sites where students were sampled could be prohibitive. Thus, other probability-based methods that employ multiple stages of selection, and/or stratification, and/or clustering are used to draw more practical samples that can be generalized with known degrees of sampling error. These samples are referred to as complex sample designs. To properly design and analyze data from these kinds of samples, agencies will need to consult with trained survey statisticians to accurately reflect the statistical effects of the design on the survey estimates.

Agencies need to consider tradeoffs between the cost and efficiency of different sample designs for their purpose, and should demonstrate why the particular design they have chosen is appropriate for their research questions and planned uses of the information. In their ICRs agencies should provide a complete description of the proposed sample design including a description of each stage of selection, a description and definition of the strata, including estimates of the size of the universe and the proposed sample by strata. Any clustering in the sample should also be described.

33. How large should a sample be for a statistical survey?

There are a variety of factors that will affect the size of the probability sample that an agency will need for a particular collection in order to obtain the quality of information that is needed. The size of a sample needed for an information collection is affected by a number of different factors including:

- degree of precision required--the significance level and confidence levels required for the estimates, and the acceptable margin of error;
- variability of the overall population on the key variables being measured;
- approximate values of the statistics that will be estimated, especially for proportions;
- type of estimate;
- sample design, e.g., the stratification and clustering of the sample;
- whether overall national estimates are the primary focus or whether estimates will also be made for subgroups—each subgroup must have adequate sample sizes; and

Sampling

- size of the overall population that estimates will describe.

Agencies will need to consult with trained survey statisticians to ensure that the sample size is adequate for its intended purposes. In Part B, agencies need to provide their precision requirements for the estimates they intend to produce from the survey to justify the sample size and the resulting respondent burden. Although overall national estimates are often the focus of Federal surveys, in many cases what is of greater analytic interest to the agency is either sub-national estimates or estimates for subgroups, e.g., different industries in an establishment survey or different income or education groups for a demographic survey. The precision requirements for estimates of these subgroups often drive the overall sample size that is needed, and therefore should be clearly documented in Part B of the ICR.

For illustrative purposes, Table 1 below provides a very general guide on sample sizes in the special case of a simple random sample and a survey variable that can be expressed as a percentage of the sample. The table provides 95 percent confidence intervals for different estimated percentages from the survey (shown in the first column) with different sample sizes of a simple random sample (shown on the second row across the columns). The size of the 95 percent confidence interval for each combination of survey estimates and sample sizes is shown in the body of the table. For example, if an item on a survey is selected by 50 percent of the respondents and the sample size is 400 respondents, the 95 percent confidence interval for this estimate would be 50 percent plus or minus 5.0 percent, or 45 percent to 55 percent. Values in this table are based on a simple random sample; many complex sample designs (see question #32), especially those using natural clusters, will typically require larger overall sample sizes to achieve the same level of precision.

Table 1. Half-Width 95 percent Confidence Intervals for Estimated Values of Percentages as a Function of Sample Size (for simple random samples)

Survey Estimate	Sample Size								
	50	100	200	300	400	500	700	1000	2000
50	14.1	10.0	7.1	5.8	5.0	4.5	3.8	3.2	2.2
60	13.9	9.8	7.0	5.7	4.9	4.4	3.7	3.1	2.2
70	13.0	9.2	6.5	5.3	4.6	4.1	3.5	2.9	2.0
80	11.3	8.0	5.7	4.6	4.0	3.6	3.0	2.5	1.8
90	8.5	6.0	4.2	3.5	3.0	2.7	2.3	1.9	1.3
92	7.7	5.4	3.8	3.1	2.7	2.4	2.0	1.7	1.2
95	6.2	4.4	3.1	2.5	2.2	1.9	1.6	1.4	1.0
98	4.0	2.8	2.0	1.6	1.4	1.3	1.1	0.9	0.6

34. Can pre-existing survey panels, such as Internet panels, be used to obtain representative samples?

Recently, some private sector firms have developed pre-recruited panels of respondents who respond to surveys on the Internet. These pre-existing panels consist of lists of potential respondents that were recruited from a variety of sources and are maintained for additional survey use. For market research, these panels have become an easy, quick, and inexpensive way to assess consumer preferences. However, use of these panels for Federal surveys that are seeking to generalize to a target population can be problematic. Often, respondents in these panels are not recruited using probability methods (see question #30), and the panels are typically simply convenience samples of persons interested in taking part in surveys on the Internet (see question #35). Because the sample is not a probability sample where each member of the target population had a known nonzero chance of selection, the results cannot be generalized to any target population using traditional statistical criteria.

Some Internet panels have been recruited from a probability-based sampling frame such as a Random Digit Dialing (RDD) sample of telephone numbers, and panel members are given Internet access as part of their participation. In this case, the Internet simply serves as the mode of data collection, not the sampling frame (see question #43). The issues of coverage and quality of the frame apply to whatever frame was used (e.g., RDD), not the Internet. However, there are also concerns about potential self-selection of respondents and low response rates in these panels (see question #72). These panels work well when samples of persons interested in taking part in surveys are needed, and the objective is not to generalize to a specific target population (e.g., pilot studies).

Agencies planning to use a pre-existing panel or Internet-based sampling frame need to justify its appropriateness for the intended use of the data in the ICR (see question #72).

35. What are some common nonprobability samples, and why are they used?

Under some circumstances, agencies may consider using nonprobability or purposive samples. It is not possible to calculate a probability of selection for these kinds of samples; therefore, their use is typically limited to research or exploratory purposes. Agencies need to understand the limitations of these samples and how those limitations will affect the use of data resulting from these samples. Agencies should justify in their ICRs the rationale for choosing a particular nonprobability sample and state how they will use the data. Agencies conducting surveys that were not designed to produce valid and reliable results that can be generalized to the universe of study must clearly explain how the collection is necessary to satisfy a statutory requirement or other substantial need. Otherwise, OMB cannot approve the collection.¹³

Convenience samples are mostly drawn from units of the population of interest that are close at hand or willing to participate. In convenience samples, there is little to no effort made to ensure that the samples are representative of the population. Consequently, they are relatively inexpensive, easy to plan, and take a minimal amount of time to draw. Though results cannot be

¹³ 5 C.F.R. § 1320.5(d)(2)(v).

Sampling

generalized to a target population, convenience samples can be useful for pilot research studies, testing of questionnaires, and some customer satisfaction surveys. Examples of convenience samples include shoppers at a mall, truck drivers visiting a weigh station, attendees at a conference, or visitors at a web site.

Quota samples are samples where units are selected nonrandomly based on a quota. The quota may be defined such that the final numbers of participating units with given characteristics have the same proportion as corresponding units have in the population. While the resulting quota sample may appear to be representative of the population for a set of characteristics, there is still an element of convenience—only those units that were the most ‘available’ become part of the sample. Also, there is no controlling for additional nonrepresentativeness that may exist in the sample for variables not used to define the quotas.

Expert choice samples are purposive samples in which an “expert” specifically chooses sample elements with certain characteristics to mimic ‘typical’ or ‘representative’ members of the population. In addition to the inability to determine the probability of selection associated with the sampled cases, this method can also produce entirely different types of samples depending on the opinions of the experts used.

Snowball samples are traditionally used to sample rare populations or populations that are hard to locate. A frame or sample for the rare population is created or identified by beginning with a set of units belonging to the target population, and asking this initial set to provide information on other members of this population. These units are then contacted for information that they may have on others in the population. This method of sampling is excellent for building a frame or creating a sample based on informal social networks and is often used for research or investigative purposes. For example, testing new questions on race with individuals of a particular background (e.g., Hmong) might be accomplished by finding some initial participants at a community center providing services to Hmong patrons and then asking them to refer others with the same background. However, there is no good way to evaluate the coverage of the frame constructed in this manner, and duplications in the frame are not always evident.

Cut-off samples are selected by ordering the universe of potential respondents by some important characteristic and selecting the units with the greatest amount of the characteristics until some specified percentage of the universe is included in the sample. Cut-off samples are used for some economic surveys conducted by Federal agencies. See question #31 for further information about justifying the use of cutoff samples.

Useful Resources

Best, S.J., Krueger, B.S. (2004). *Internet Data Collection*. Sage University Series on Quantitative Applications in the Social Sciences, 07-141. Thousand Oaks, CA: Sage.

Sampling

- Cochran, W. G. (1963). *Sampling Techniques*. New York: Wiley. (more technical)
- Cohen, J. (1988). *Statistical Power Analysis of the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Hansen, M.H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample Survey Methods and Theory*. New York: Wiley. (more technical)
- Kalton, G. (1976). *Introduction to Survey Sampling*. Sage University Series on Quantitative Applications in the Social Sciences, 07-035. London: Sage.
- Lee, E. S., Forthofer, R. N., and Lorimor, R.J. (1989). *Analyzing Complex Survey Data*. Sage University Series on Quantitative Applications in the Social Sciences, 07-071. London: Sage.
- Levy, P. S. and Lemeshow, S. (1999). *Sampling of Populations: Methods and Applications*. New York: Wiley.
- United States General Accounting Office (1992). *Using Statistical Sampling*, GAO/PEMD-10.1.6. Washington, DC.

MODES OF DATA COLLECTION

The purpose of this section is to provide an overview of the different modes for survey data collection and some of the strengths and limitations of each mode. Because the choice of mode affects and is affected by many other aspects of the survey design, the choice of mode or modes should be carefully considered by agencies, and they should consult with trained survey methodologists in selecting the appropriate data collection mode or modes given the survey's purpose.

36. What are the different modes of survey data collection?

The mode of data collection includes the way in which respondents are contacted and how their responses are obtained. The most commonly used data collection modes are in-person (or face-to-face), telephone, mail, and web (including e-mail). In-person and telephone surveys are typically interviewer-administered, while mail and web surveys are self-administered, though technology is creating new hybrids, such as self-administered telephone surveys using touchtone data entry (TDE) or interactive voice response (IVR). Although mail, telephone, and in-person surveys were traditionally conducted with a paper and pencil questionnaire, many Federal surveys now use some form of computer-assisted interviewing (CAI). Each mode of administration may be computer assisted: Computer-Assisted Telephone Interviewing (CATI), Computer-Assisted Personal Interviewing (CAPI), and Computer-Assisted Self-Interviewing (CASI).

There are a number of advantages of CAI. It allows for more complex questionnaire designs because CAI instruments can use answers from several questions to direct the interview through different series of questions, skip particular questions, and change question wording based on previous responses. Data quality may also be improved by including range and consistency checks into the instrument to help ensure that the correct information is being entered by the respondent or the interviewer. Furthermore, for longitudinal surveys, information provided previously by the respondent may be available to reduce respondent burden and/or improve data quality. Finally, data are usually available quickly and with fewer data entry errors than data from paper instruments that require clerical check-in and keying.

There are also disadvantages associated with CAI. CAI instruments often take longer to develop and program than paper instruments and may be costly to pretest and revise. Agencies need to schedule lead time to draft specifications and author the instrument, as well as test and debug the instrument. The time and costs involved in these efforts can be considerable for complex instruments. In addition, interviewers will need training in using the instrument. Data quality may also be affected by the usability and design of the instrument, so agencies are encouraged to include usability testing as part of their pretesting (see question #22 and #49).

37. What mode of data collection is appropriate for a given survey?

Each mode of data collection has inherent advantages and disadvantages, and there is no one best data collection mode for all situations. Selection of the mode requires consideration of many factors.

Specifically, agencies should consider the following statistical and nonstatistical issues when selecting a mode of collection:

Statistical issues in mode selection

- *Coverage*: Who in the target population may be missed in the sampling frame used for the mode? Because the mode of data collection is often intertwined with the selection or availability of a sampling frame, it has implications for how well the target population will be covered (see question #28). For example, a telephone survey would be inappropriate in terms of coverage for a study of poverty or means-tested programs where a significant portion of the target population does not have a phone or often has phone service interruptions due to nonpayment of bills.
- *Nonresponse bias*: How different are respondents expected to be from nonrespondents? Data collection modes can also affect response rates and nonresponse bias. For example, in-person surveys on average show the highest response rates, and for household surveys, telephone survey response rates have tended to be higher than mail surveys, although recent declines in telephone response rates may remove this advantage. There is also evidence that mail surveys yield higher response rates than web or e-mail surveys. Different data collection modes also have different implications for investigating potential nonresponse bias. For example, the interviewers for in-person surveys can record characteristics of the neighborhood for respondents and nonrespondents allowing a comparison of differences between these groups.
- *Measurement error*: What factors may affect the quality and completeness of responses? The choice of mode can affect the completeness of data that are collected and the extent to which there are other response effects such as social desirability bias, and response order effects. For example, the presence of an interviewer has been shown to affect reporting of sensitive behaviors such as illicit drug use.

Nonstatistical issues in mode selection

- *Timeliness*
- *Cost*

The following questions and answers on each mode provide some advantages and disadvantages for each mode of collection in terms of both statistical and nonstatistical issues. For clarity and simplicity, the information in the following questions and answers describes the advantages of each mode when it is the only one used. In practice, multiple modes are frequently used by agencies to overcome specific disadvantages associated with a single mode of collection (see question #38).

In considering which mode (or modes) of survey data collection is appropriate for their information collection, agencies will need to balance the advantages and disadvantages of each

mode for the given purpose of the survey, the use of the data, the characteristics of the respondent population, and available resources. Agencies must justify their choice of mode of data collection in their ICRs and provide details of their data collection methods in Part B of their supporting statements.

38. When should agencies consider a mixed-mode approach?

The two main reasons to consider using more than one mode of collection simultaneously are cost and response rates. The typical mixed mode approach is to use a less costly method for initial contact and a more costly mode for follow-up with nonrespondents, such as using a mail survey with telephone nonresponse follow-up or a telephone survey with an in-person nonresponse follow-up.

Using multiple modes often yields a higher response rate by offering alternative means to respond, so respondents may choose the mode that is most convenient for them; for example, some businesses may prefer to respond via the Internet rather than complete a paper questionnaire. A multimode survey can often be conducted at a lower cost than doing the entire survey using the more expensive mode. However, switching modes for a nonrandom subset of the sample (those who initially did not respond) may also introduce additional variability or bias due to mode effects. Using a mixed mode approach is best applied in situations where the trade-off in the reduction of nonresponse error compensates for any increase in response error related to mode effects. Agencies planning mixed mode collections should carefully design their survey questionnaires to minimize potential mode effects or they should consider conducting experimental studies to assess potential mode effects.

39. How does GPEA affect choice of modes for survey data collection?

The Government Paperwork Elimination Act (GPEA) required agencies by October 21, 2003 to provide for the option of electronic reporting when practicable, and OMB has issued implementation guidance on this law.¹⁴ Agencies are increasingly offering either computer-assisted interviewing, fax reporting, or options for filling out a survey on the web. Because many households do not have access to the Internet or computers, electronic reporting options for many household survey information collections will likely lead to more multi-mode surveys rather than replacing telephone or mail surveys completely with electronic collections; however, many business establishments may prefer an electronic option.

An electronic option should be considered for every data collection, and if electronic reporting is not offered for a survey, agencies should explain why it is not practicable for them to offer an electronic response option.

¹⁴ <http://www.whitehouse.gov/omb/fedreg/gpea2.html>.

40. What are the advantages and disadvantages of mail surveys?

Mail surveys have been particularly useful for mandatory household and establishment surveys. They are most appropriate when there is a good address list for a sampling frame. For example, both the Census Bureau and the Bureau of Labor Statistics have list frames of business establishments (the Census Bureau also has the master address file for households). Other Federal agencies may have administrative records of their program participants that serve as their sampling frames. Mail surveys have relatively low cost, and self-administration of the questionnaire improves response to sensitive questions, minimizing social desirability and interviewer biases. There is also evidence that question order effects are reduced in mail survey questionnaires. In addition, visual aids can be used with this mode.

There are a number of disadvantages of mail surveys. Mail surveys frequently can suffer from low response rates, especially for household surveys; therefore, they are often used in mixed mode surveys (see question #38) with follow-ups done via telephone or in-person. Furthermore, there may be more nonresponse bias in mail surveys because the respondent can look over the entire questionnaire before deciding whether to respond or not, increasing the likelihood that the decision to respond is based on his or her values on the key variables of the survey. Mail surveys require accurate mailing addresses, a longer data collection phase than other modes (usually eight weeks or more are required from the initial mailing), and greater literacy skills on the part of respondents. Household mail surveys should generally have short, less complex questions; however, more complex questions are frequently used in mail surveys of business establishments. It is important that mail questionnaires have clear instructions, easy to follow layouts, and clear question wording because there is no interviewer present to clarify the agency's intent. The items on mail survey questionnaires are more likely to be incomplete than those on surveys that employ other modes (higher item nonresponse). Mail surveys may be completed by someone other than the sampled person, which can be problematic for household surveys that seek to randomly select one adult from among the household members. In mail surveys of business establishments, gatekeepers can prevent respondents from receiving questionnaires. For example, a secretary or administrative assistant may open all the mail for an office and may routinely throw away voluntary surveys rather than passing them on to the appropriate respondent.

41. What are the advantages and disadvantages of telephone interviewing?

Telephone interviewing can provide lower costs and shorter data collection times compared to in-person interviews. Telephone coverage rates in the United States have been over 90 percent for several decades, and random-digit dialing (RDD) surveys have become very popular for a wide range of social, political, and marketing research. RDD telephone surveys have been very useful for screening large numbers of households to find rare populations. Often telephone surveys are conducted using computer assisted telephone interviewing (CATI) in large, centralized facilities that include careful supervision and monitoring of interviewers. Thus, in CATI calling centers, interviewer variance is likely to be less than for in-person interviews. Social desirability concerns may also be less than is the case for in-person surveys. Some

technology is also being used to reduce further the role of the interviewer and allow self-administered telephone surveys using touchtone data entry (TDE) or interactive voice response (IVR). For example, short, simple surveys that require numeric responses that can be entered on the telephone key pad may be done through touch-tone data entry (TDE), and may not require an interviewer at all. TDE has worked well with ongoing, simple establishment surveys when respondents have been trained to answer this way.

There are also disadvantages to telephone interviewing. Although telephone coverage in the United States is very high, some groups, such as those with low incomes, are more likely to be excluded from the sampling frame, thereby introducing bias. Cell phones currently are not covered by existing frames, and there are some legal restrictions on calling these phones (i.e., automated dialers cannot be used). Current estimates of households with only cell phones are still small; for example, a 2004 supplement to the Current Population Survey showed that approximately 6 percent of households had cell phones but no landline phone.¹⁵ However, the number of households with only mobile phones is growing and may present further coverage problems for telephone surveys. Increasing use of technologies including answering machines, voice-mail, and caller ID is making it harder to reach households, and RDD telephone survey response rates have been decreasing more rapidly than those of other modes in recent years (see question #70). RDD sampling frames have no information other than the phone number, so vendors are often used to match phone numbers to addresses and other area-level information to make it possible to mail advance letters, do in-person follow-up, or conduct non-response bias analyses (see questions #38, #70, and #71). However, match rates for phone numbers to addresses are frequently 50 percent or less when using a single vendor.

Conducting an interview over the telephone also imposes more constraints on the length of the questionnaire and complexity of the questions compared to some other modes. For example, the design of telephone surveys usually includes short questions, with a minimum number of response categories, and a relatively short interview length. Some response effects, such as question order and response order, are more likely to occur in telephone surveys than self-administered surveys. For example, respondents are more likely to select the last response option (recency effect) in a telephone survey than a mail survey. There is no ability to use a visual communication medium in telephone surveys (unless materials are mailed in advance), and it is difficult to search records or look up information during a telephone interview. In business establishments and some households, gatekeepers can prevent the interviewer from reaching the desired respondent. For example, a secretary or administrative assistant may screen all calls and not allow access to the appropriate respondent. Similarly, a spouse or parent may prevent direct access to the sample person.

Recent changes in the law related to the portability of telephone numbers are likely to have an increasing impact on telephone surveys and make it more difficult and expensive to conduct them in the short term. The longer term implications are not yet clear.

15 Tucker, N. C., Brick, J. M., & Meekins, B. (2005). Household telephone service and usage patterns in the U.S. in 2004: Implications for telephone samples. Paper presented at the 98th Meeting of the Committee on National Statistics, National Academy of Sciences, Washington, DC.

42. What are the advantages and disadvantages of in-person interviewing?

Area probability sampling and in-person interviewing provide the best coverage for household surveys, and are often considered the gold standard. Given adequate time and numbers of contact attempts, in-person interviews typically have the highest response rates. In-person interviews also allow the use of visual tools such as flash cards and calendars, and permit longer, more complex interviews to take place. In addition, the interviewer can make observations about the neighborhood of the household, or in establishment surveys, facility characteristics can be documented.

The primary disadvantage of in-person interviews is the high cost associated with sending an interviewer to households or business establishments to collect the data. Also, the data collection phase of the survey may take longer (or require a larger interviewing force) compared to other modes of collection. In-person interviewers may also face barriers in completing their assignments because some sampled addresses may be less accessible to interviewers, e.g., high rise buildings and gated communities, or be in high crime areas with greater risk to interviewer safety. Because in-person interviewers typically operate alone with much less supervision and control than is possible in more centralized telephone facilities, there may be greater interviewer variance in in-person surveys, and there are also greater opportunities for interviewer falsification of some survey items or entire interviews. Survey organizations typically conduct some reinterviews or independent verification of interviewers' work to detect and minimize falsification.

In-person interviews may not be the best mode of data collection when respondents in a business must refer to detailed records that take time to find, or when there are concerns about potential interviewer or social desirability bias. However, technology has been used to address concerns about social desirability with surveys on sensitive topics, such as illicit drug use. Portions of the in-person interview can be completed privately by respondents using an audio computer assisted self-interview (ACASI) instrument that "reads" the question to the respondent through headphones to increase privacy as well as reduce bias due to literacy or English proficiency problems. Respondents can listen to questions and look at the instrument on a computer screen at the same time and respond in private. Research has demonstrated that respondents report more incidents of drug use, sexual activities, and other sensitive behaviors using this mode of collection, which is presumed to reflect more accurate reporting.¹⁶

43. What are the advantages and disadvantages of using Internet surveys?

To comply with the Government Paperwork Elimination Act (GPEA), agencies are increasingly offering the opportunity to respond to surveys, especially surveys of business establishments, via electronic means, including the Internet (see question #39). When e-mail addresses are

¹⁶ See Turner, C., Forsyth, B., O'Reilly, J., Cooley, P., Smith, T., Rogers, S., and Miller, H. (1998). "Automated self-interviewing and the survey measurement of sensitive behaviors," in Couper, M., Baker, R., Bethlehem, J., Clark, C., Martin, J., Nicholls II, W., and O'Reilly, J. (eds.), *Computer assisted survey information collection*, pp. 455-473, New York: Wiley.

available, using the Internet can be a very inexpensive way to contact and remind respondents about completing the survey.

Similar to mail surveys, simple Internet surveys can be low cost, but data collection can be faster since reminders and responses can be sent and received without delay. The Internet offers the potential for presenting visual aids or even multi-media presentations of information to respondents, and self-administration of the questionnaire increases response to sensitive questions, while minimizing social desirability and interviewer biases. Like other modes that use computer administration, data processing time and cost may be reduced compared to paper and pencil surveys because data can be uploaded or captured directly into databases without additional keying. Data quality may also be higher because the instrument can contain built-in edits and prompts. However, more complex instruments may be costly to pretest and revise.

There are a number of disadvantages to Internet surveys. As noted in question #29, a key problem is that there is no sampling frame of persons or establishments with Internet access or means of randomly generating e-mail addresses to obtain a random sample of users. Low household coverage rates for Internet access as well as systematic differences between households with access and those without access means that using the Internet as the sole mode for population surveys is problematic. If an agency has a high quality sampling frame for its target population and knows the members of the target population have Internet access, then the agency could consider an Internet-only collection.

There are other disadvantages or limitations to Internet surveys. Mail surveys (on paper) typically achieve higher response rates than web surveys or e-mail surveys. Respondents need to be “computer literate” and have access to the Internet. Respondents may also have a variety of hardware and software configurations that may cause differences in how they see and interact with the survey. Therefore, usability testing should be an important part of the agency’s development and testing of the questionnaire. Respondents may have concerns about confidentiality and, therefore, be reluctant to provide some information over the Internet. Finally, there is little control or knowledge about whether the selected respondent is the actual survey respondent.

44. How does the data collection mode affect questionnaire design?

Each mode of data collection has implications for different issues in questionnaire design, including skip patterns, response options, and question wording. Studies that use multiple modes of collection also need to ensure that the questionnaire can be effectively administered in each mode and any response effects due to the mode of administration are minimized.

For example, skip patterns can be programmed easily into a computer assisted interview (CAI) instrument, but only limited skip patterns should be used in paper instruments because errors are more likely to occur when respondents or interviewers are asked to follow skip patterns. Self-administered interviews can have more response options than interviewer-administered interviews (especially telephone surveys) because the respondent can easily go back over the list before selecting a response and can more easily change a response if an inaccurate response has

been entered. Also, visual aids cannot be used to help the respondent understand the question or the response options in a telephone interview. Complex and long sentences should generally be avoided in survey questions, but they are particularly difficult to understand over the telephone.

It is important that agencies test their survey questionnaires in all modes that they plan to use to collect information for the full-scale survey (see section on Questionnaire Design). Usability testing of computer survey instruments should also be included as part of questionnaire pretesting to identify problems either interviewers or respondents may have with the instrument (see question #48).

Useful Resources

Couper, M. P., Baker, R., Bethlehem, J., Clark, C. Z. F., Martin, J., Nicholls II, W. L., and O'Reilly, J. M. (1998). *Computer Assisted Survey Information Collection*. New York: Wiley.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method (2nd edition)*. New York: Wiley.

Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

QUESTIONNAIRE DESIGN AND DEVELOPMENT

The focus of this section is to provide a brief overview of the methods for developing and testing questionnaire items and clarify the requirements for conducting these activities and obtaining OMB approval under the PRA. Agencies should consult with survey methodologists and cognitive psychologists trained in these methods to help design and test questionnaires prior to survey administration. Questionnaire Evaluation Methods are described in Statistical Policy Working Paper #47, *Evaluating Survey Questions: An Inventory of Methods*, published by the Federal Committee on Statistical Methodology.

45. What should agencies do when developing new questionnaires or questionnaire items?

Agencies need to use questions that will elicit the appropriate information from respondents to fill the agencies' data needs. Agencies should determine the different topics, estimated survey length, and mode of survey administration prior to drafting the actual survey questions. Questionnaire designers should inspect other instruments that have collected data on similar topics and must also adhere to OMB classifications and standards for particular data elements, such as data on race and ethnicity, and industry and occupation (see question #47, #52, and #55). If new questions are needed, agencies should use techniques to test the questions that will ensure that the questions they develop provide the information they need and have adequate statistical reliability (see question #48).

Agencies should clearly document in their ICRs the source for questions that were taken from other surveys and identify new questions the agency has developed and tested on its own. For ongoing surveys, any changes to the questionnaire should be clearly noted and described. The plan for testing or the results from the testing should also be described in Part B of the ICR.

46. Why should agencies consider using questions previously used by other agencies or researchers?

There are many surveys conducted by government agencies or private sector entities that may include questions that will effectively capture some of the data needed by an agency. Questions used in other major surveys are more likely to have known characteristics of reliability and validity, allowing the agency to take advantage of research data collected and analyzed by other agencies. However, the questions may have been asked in a different context or be more detailed than needed. Agencies will need to consider these factors when looking at other agencies' questions.

Agencies with data needs outside their particular subject matter specialty should consult with the statistical agencies that are most likely to collect the type of data needed as well as with agencies with expertise in the area. For example, if a health survey needs employment data, the agency should collaborate with the Bureau of Labor Statistics or the Census Bureau, or if an education survey needs disability questions, the agency should consult with the National Center for Health Statistics, the Census Bureau, and the Social Security Administration. Many Federal agencies currently post survey questionnaires on their web sites, which facilitates the sharing of questions.

Interagency groups are occasionally formed to develop standardized questions on subjects that cut across many agencies. For example, an Interagency Committee on Measures of Educational Attainment has reviewed and recommended a set of standard categories for educational attainment.¹⁷ More recently, an interagency committee on the American Community Survey has coordinated across agencies to share needs for information and to test alternative questions.¹⁸

On the other hand, asking previously used questions does not mean that the survey requires no pretesting. There is substantial evidence that the context of the question affects its performance; hence, pretesting is always needed.

47. When is it acceptable to duplicate questions used on other surveys?

In designing their information collections, agencies are expected to review existing studies to determine whether the information the agency needs exists elsewhere. Agencies are to describe the existing information in their ICRs and show specifically why the information already available cannot be used or modified for use. If the existing information will not fulfill the agency's needs, the agency should take advantage of the developmental work from existing collections to inform the design of its information collection.

Using questions from well-established national data collections such as the Current Population Survey, the Decennial Census, the Medical Expenditure Panel Survey, the National Crime Victimization Survey, or the National Health Interview Survey helps assure comparability of results. Reasons for using the same questions include benchmarking the responses of one survey to another, or obtaining comparable information from a different population or from the same population at a different time period. In their ICRs, agencies should clearly document the source for questions that were taken from other surveys.

48. What techniques can be used to develop new questions?¹⁹

Developing effective new questions is often more difficult than most people anticipate, especially if the questionnaire designer is not experienced with survey measurement. Agencies immersed in a topic are often surprised how little respondents know and care about the topic. Agencies may assume knowledge and points of view that respondents may not have. For example, respondents need to know what the intention of the question is so that they can answer it appropriately, and they may have difficulty understanding questions that use technical or unfamiliar terminology. Questions need to be developed so that respondents can answer the question and provide useful data for the agency. The following survey research methods can be used to develop and pretest new survey questions:

17 Federal Interagency Committee on Measures of Educational Attainment (2000). Federal measures of educational attainment: Report and Recommendations.

18 See *Statistical Programs of the U.S. Government FY 2005* at www.whitehouse.gov/omb; Go to Statistical Programs and Standards.

19 This section is based on *Census Bureau Standards: Pretesting Questionnaires and Related Materials for Surveys and Censuses*, U.S. Bureau of the Census, July 25, 2003.

Focus Groups

Focus groups are often a useful first step in questionnaire development. Typically, a moderator will guide participants in a focus group discussion on the topics related to the subject area of the survey. Participants are encouraged to talk using their own terms and experiences and react to what others have said. In fact, it is often the interaction among participants that provides the most useful insights. Agencies can learn the language that respondents use when discussing the topic and integrate more common terms and phrases into the design of survey questions.

Focus groups often serve as a way to test the ease of completing a self-administered questionnaire. After completing the questionnaire individually, the group discusses the experience with overall direction from the moderator. This provides information about the appearance and formatting of the questionnaire in addition to content problems.

Finally, focus groups can be very effective in the ultimate design of surveys that ask about sensitive topics. Asking sensitive questions in a survey environment can be especially awkward, and discussions among focus group participants can provide useful information on appropriate wording, terms, and phrases that respondents will not find offensive.

Pre-Survey Design Visits for Establishment Surveys

Visiting a respondent's place of business to review plans for a new survey or major changes in an existing survey can be very useful in improving the final design of a questionnaire. These visits generally involve discussions with a potential respondent on the following topics:

- Does the respondent keep the data that the agency wants?
- How closely does the establishment's record keeping correspond to the required survey data?
- How compatible are these record keeping systems with the agency's collection instruments?
- How difficult will it be to provide the data in the time period needed by the agency?

These visits can help in the preliminary stages of survey development to ensure that the data collectors will design a survey that respondents can complete and that will obtain useful and usable information that corresponds to the agency's data needs.

Cognitive Interviews

The goal of cognitive interviews is to gain insight into how respondents think about and interpret the survey questions. In classical cognitive interviews, this is done by asking respondents to think aloud as they answer questions (concurrent think aloud) and to identify anything that confuses them. Respondents are often asked to paraphrase a question so that researchers learn whether a respondent understands the question and interprets it as intended. If the same paraphrased wording is used by several respondents, it might suggest a better wording of the question. Cognitive interviews can also include in-depth retrospective debriefings during which the interviewer asks a series of probes after the completion of the survey.

A key benefit of using cognitive interview methods is that researchers can quickly diagnose problems, revise question wording to solve problems and conduct additional interviews to see if the new questions are less problematic. The most effective process is to plan iterative rounds, first identifying problems, then making changes to items, and then trying out those changes in successive rounds of cognitive interviews. OMB has issued an addendum to *Statistical Policy Directive No. 2, Standards and Guidelines for Cognitive Interviews*, available at www.WhiteHouse.gov/sites/default/files/omb/inforeg/directive2/final_addendum_to_stat_policy_dir_2.pdf.

49. What role does pretesting play in questionnaire development?

When an agency has developed new survey questions or is pulling questions from different sources into a new questionnaire, it is important to test how respondents will react to the individual items and the questionnaire as a whole, so a variety of methods are often used to test the questionnaire. In a pilot test, the survey (or some portion of the survey) is administered to a sample of respondents similar to those in the main study using procedures planned for the full survey. Although this is often the only type of testing done prior to data collection, this type of pretest is not beneficial for question development unless there is a respondent and/or interviewer debriefing as part of the process or the data from the pretest are reviewed for questions with high item nonresponse.

Pretests may aid in question development by using one or more of the following methods:

- respondent debriefing
- interviewer debriefing
- split panel designs
- behavior coding
- data reviews

Respondent Debriefing

Respondent debriefing typically consists of follow-up questions at the end of an interview that are designed to obtain quantitative information about respondents' interpretations of survey questions. These questions help researchers determine whether concepts and questions were understood by respondents in the same way that the survey designers intended. In an interviewer-administered survey, the debriefing questions may be followed by a discussion between respondent and interviewer, to further probe the respondent's reaction to and comprehension of the questions in the survey instrument.

Interviewer Debriefing

Evaluating pilot tests of demographic surveys conducted by personal interview has often centered on structured debriefing of field interviewers at the end of the test. Interviewers are trained prior to the survey and are asked to carefully record problems they encounter during the interview. Typically, interviewers know in advance that they will be participating in debriefing discussions at the end of the pilot test.

Although some valuable insights can be obtained from interviewers, it is important to recognize that they may not always be accurate reporters of certain types of questionnaire problems or may be conveying their opinions more than respondents' problems. For example, when interviewers report a problem, researchers cannot assess whether it was troublesome for one respondent or for

many, or whether the problem reflects the interviewer's own preference or understanding of the question rather than respondent confusion. In addition, experienced interviewers sometimes change the wording of problem questions as a matter of course to make them work, and may not even realize they have done so.

Split Panel Designs

In a split panel field test, respondents are randomly assigned into different groups to receive different versions of the questions. This is a very useful method for comparing two (or more) different versions of the same question or testing question-order effects because the responses can be compared between the different panels to examine the potential impact of the change on survey estimates.

Behavior Coding

Behavior coding focuses on the overt behavior of interviewers and respondents as they interact during the survey interview. Although behavior coding can be done by an evaluator in real time during the interaction between the interviewer and respondent, frequently the interaction is recorded and then coded by one or more evaluators. There are a variety of coding systems that reflect errors made by the interviewer and difficulties the respondent has with the questions. Because it is a quantitative method, a relatively large number of interviews need to be coded and statistical summaries created to identify problematic questions.

Data Review

A data review of the pilot test results is conducted to identify questions that have higher than expected or desired levels of non-response (either don't know or refusals). High item nonresponse in a pilot test could indicate poor question wording, generally unavailable data, or non-applicability of the question to a significant subset of respondents. Because data review involves examination of quantitative results from the pilot test, larger numbers of respondents may be needed with more complex instruments to ensure that an adequate number of respondents are asked each question.

50. What do agencies need to do to obtain clearance for pretesting activities?

Pretesting activities, including cognitive interviews and focus groups, must comply with requirements of the PRA, which are detailed in 5 C.F.R. § 1320. Although agencies do not need OMB approval to test draft questionnaires when they are administered to fewer than 10 persons, agencies must obtain approval to conduct iterative testing of the same questions even with minor modifications, on a total of 10 or more persons.²⁰ Thus, it is not acceptable for an agency to test a questionnaire on seven individuals, make minor format and grammar changes, and test those revised questions on another seven people without OMB approval. Focus groups are also subject to the PRA (see question #7). Because most meaningful pretesting, especially iterative rounds of testing on different versions of questions, will require more than nine persons, agencies will need to seek clearance to conduct their pretesting.

²⁰ Note, however, that if fewer than 10 persons or entities make up a substantial proportion of the entire population, e.g., car manufacturers, the collection may also be subject to the PRA.

Agencies that plan to do pretesting activities, including cognitive interviews and focus groups, can obtain OMB approval in one of two ways. First, the pretesting activities can be described and submitted as part of the ICR for the final survey. When this approach is used, OMB approval usually includes a term of clearance that the agency must report to OMB the results of the pretesting and any changes to the survey instrument that were made based on the findings. Alternatively, the agency can submit a separate ICR just for the pretesting activities, and later submit an ICR for the final survey that reflects the results of the pretest. Agencies usually do the latter when the pretest involves a design that is complex, includes large numbers of respondents, or has a relatively high response burden. Agencies also should submit the pretest separately from the full-scale collection when little has been decided about the design of the final survey when the pretesting is planned.

51. What is a generic clearance for pretesting activities?

Agencies that regularly do pretesting and development work for multiple surveys have found it beneficial to obtain a generic clearance specifically for these kinds of studies. Once the overall generic clearance is obtained on the pretesting activities and methods that will be used (e.g., cognitive interviews, focus groups, respondent debriefings, etc.) through the normal clearance process, agencies can submit abbreviated collection requests on the specific questions to be tested and obtain expedited OMB review (often within 10 working days) of the specific study, which can greatly facilitate ongoing and iterative rounds of testing. For example, cognitive laboratories at the Bureau of Labor Statistics, the Bureau of the Census, and the National Center for Health Statistics have these clearances.

The primary justification for having a generic clearance for pretesting is that agencies know in advance that methodological research is needed, but they cannot anticipate the specific kinds of tests or methods that will be used. Generic clearances provide a mechanism for agencies to quickly test and implement new survey questions that often arise to address policy issues or emerging programmatic needs. The generic clearance should only be used in a well-defined and structured context, such as methodological testing. It is not appropriate for an agency to use a generic clearance as a means to bypass the requirements of the PRA to conduct a variety of information collections. Agencies are encouraged to consult with their OMB desk officers before submitting a generic clearance to determine whether their plans are appropriate for this type of clearance (see also question #8).

Useful Resources

Bradburn, N.M., Sudman, S., & Wansink, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design -- For Market Research, Political Polls, and Social and Health Questionnaires*, Revised Edition. San Francisco: Jossey-Bass.

Converse, J. & Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, CA: Sage.

- Federal Committee on Statistical Methodology (2016). *Statistical Policy Working Paper 47, Evaluating Survey Questions: An Inventory of Methods*. Washington, DC: U.S. Office of Management and Budget.
- Presser, S., Rothgeb, J., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., & Singer, E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: Wiley.
- Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M., and Tourangeau, R. (1999). *Cognition and Survey Research*. New York: Wiley.
- U.S. Bureau of the Census (July 25, 2003). *Census Bureau Standards: Pretesting Questionnaires and Related Materials for Surveys and Censuses*. Washington, DC: U.S. Bureau of the Census.
- Willis, G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage.

STATISTICAL STANDARDS

The purpose of this section is to provide an introduction to the statistical standards that OMB has issued and that agencies must utilize if the standards apply to the information the agency is collecting. In section A.7 of the supporting statement, agencies certify in their ICRs that they are not using a statistical classification not approved by OMB, or they must request a waiver of the applicable OMB standard with a justification for not using the approved classification.

52. What are OMB statistical classifications, definitions, and data sources?

Under the PRA, OMB is charged with developing and overseeing the implementation of government-wide policies, principles, standards, and guidelines concerning statistical collection procedures and methods. Statistical classifications, definitions, and data sources encourage uniformity in data collection, analysis, and dissemination. They are designed and managed to support the full range of research and analytical objectives in a specific subject matter area rather than the needs of a specific program or a specific study. The general criteria OMB has for evaluating the standards have been relevancy, accuracy, currency, efficiency, minimization of burden, and stability ("continuity" and/or "comparability"). There is a clear trade-off between currency and stability; typically, revisions to these standards have been no more frequent than once every five years or longer. However, annual updates of statistical areas are issued based on Census Bureau population estimates.

OMB currently has a number of different statistical classifications for demographic, economic, and geographic data, including data on race and ethnicity, industries, occupations, and statistical areas described in more detail in the following questions. In addition, there are some standard definitions of economic concepts for statistical purposes, and standard sources for Federal data for some demographic and economic statistics.

53. What statistical classifications have been adopted by OMB?

Standard reporting categories are necessary to ensure comparability across Federal Government statistical data. The statistical classifications are issued pursuant to OMB's authority to promulgate standards and guidelines for Federal statistics. These standards apply to all data collected for statistical use by Federal agencies and their contractors. Some standards also apply to data collected for administrative use.

There are currently six statistical classifications adopted by OMB:

- Federal Administrative Regions
- Metropolitan and Micropolitan Statistical Areas
- North American Industry Classification System (formerly the Standard Industrial Classification of Establishments)
- Standard Occupational Classification
- Data on Race and Ethnicity
- Fields of Science and Engineering (R&D)

Classifications that have been updated after 1980 are available at www.whitehouse.gov/omb (Go to Statistical Programs and Standards). Tables 5 and 6 provide a brief summary of the statistical and administrative uses of these standards, and conditions under which there may be exceptions to their use. Agencies must justify in their ICRs the use of statistical classifications that differ from those approved by OMB.²¹

Table 5. Brief Summary of Statistical Classifications’ Required Uses and Exceptions

Classification	Required for Statistical Use	Required for Administrative Use	Exceptions	Last Updated ⁶
Federal Administrative Regions	Recommended	Yes	1,2	1978
Metropolitan and Micropolitan areas	Yes	No	3,2	December 27, 2000 ²²
North American Industry Classification System	Yes	No	3	April 20, 2000 ²³
Standard Occupational Classification	Yes	No	3	September 30, 1999 ²⁴
Data on Race and Ethnicity	Yes	Yes, 4		October 30, 1997 ²⁵
Fields of Science and Engineering	Yes	No	5	1978

Notes:

1. Exceptions have been recognized for regions defined and widely used prior to the standard, such as Census regions.
2. There are other stable, widely-used geographic classifications such as Census regions and districts, USPS ZIP code areas, and political (state and county) boundaries.
3. A notice and comment process consistent with the Administrative Procedure Act is usually required if an agency proposes using or modifying the statistical definitions for program administrative purposes.
4. Required for administrative reporting and record keeping.
5. Compatible classifications of educational curricula are permitted.
6. Standards that have not been updated were last issued in the U.S. Department of Commerce *Statistical Policy Handbook* (1978), when the statistical policy authority was the responsibility of that Department (1977-1981).

²¹ 5 C.F.R. § 1320.5(d)(2)(vi).

²² *Federal Register* 65:82228-82238.

²³ *Federal Register* 65:21242-21282.

²⁴ *Federal Register* 64:53135-53163.

²⁵ *Federal Register* 62:58781-58790.

54. What standard definitions and data sources have been adopted by OMB?

Statistical definitions are provided for two economic concepts:

- Poverty (used to monitor changes in the number of persons and families in poverty and their characteristics over time), and
- Payroll Periods for Employment Reports (used to standardize reference periods).

Standard sources for Federal data are provided for some economic and demographic statistics. There are currently standard statistical data sources for:

- Labor Force and Unemployment Data, (Bureau of Labor Statistics), and
- Population Data (the Decennial Censuses and the Census Bureau's intercensal estimates)

Table 6. Brief Summary of Standard Definitions and Data Sources' Required Uses and Exceptions.

Standard	Required for Statistical Use	Required for Administrative Use	Last Updated ¹
Definition of Poverty	Yes	No	1978 ²
Definition of Payroll Periods for Employment Reports	Yes	No	1978
Labor Force and Unemployment Data	Yes	No	1978
Population Data	Yes	Yes ³	1978

Notes:

1. Standards that have not been updated were last issued in the U.S. Department of Commerce *Statistical Policy Handbook* (1978), when the statistical policy authority was the responsibility of that Department (1977-1981).
2. Although the official definition has not been changed, several experimental measures are being developed and tracked over time.
3. This standard has been incorporated into several statutes.

55. What are the requirements for collecting individual data on race and ethnicity?

The most commonly used OMB statistical classification for population-based surveys concerns data on race and ethnicity. The OMB standards provide *how* agencies must collect data on race and ethnicity **if** they are collecting this information—the standards do not require agencies to gather data on race and ethnicity. Most, if not all, of the population-based surveys or censuses have now implemented the 1997 standards for data on race and ethnicity.

The OMB standards for data on race and ethnicity provide a minimum set of two categories for data on ethnicity:

- Hispanic or Latino and
- Not Hispanic or Latino,

and five categories for data on race collected from individuals:

- American Indian or Alaska Native,
- Asian,
- Black or African American,
- Native Hawaiian or Other Pacific Islander, and
- White.

Note: “other race” is not a response category.

Respondents are to be offered the option of selecting one or more racial designations. Based on research findings, the recommended forms for the instruction are *Mark one or more*, *Select one or more*, or *Choose one or more* (not check all that apply).

The mode of administration should be taken into account when designing the exact wording of the question. For example, face-to-face surveys permit the use of flashcards with a listing of the racial categories, whereas a telephone administration must rely on the interviewer reading each of the categories. Examples of questions for different modes are provided in the *Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity*.²⁶

The standards permit the collection of greater detail; however, the additional categories must be organized in such a way that they can be aggregated into these minimum categories for data on race and ethnicity.

Self-reporting or self-identification using separate questions (the two-question format) for race and ethnicity is the preferred method for collecting the data; note that the question on ethnicity should precede the question on race.

If self-reporting is not practicable or feasible, for example, when identification is done by funeral personnel, observer identification may be used. The use of the two-question format is strongly encouraged even when observer identification is used.

All information collections that include data on race and ethnicity were to be in compliance with the 1997 standards by no later than January 1, 2003. If an agency believes the standard categories are inappropriate, the agency must request a specific variance from OMB. Further information is available on the OMB web site, www.whitehouse.gov/omb/ under “Statistical Programs and Standards.”

Useful Resources

Links to copies of the *Federal Register* notices for the updated standards are available on the OMB web site, www.whitehouse.gov/omb/, Go to Statistical Programs and Standards.

²⁶ Available on the OMB website, www.whitehouse.gov/omb/; Go to “Statistical Programs and Standards”

INFORMING RESPONDENTS ABOUT THEIR PARTICIPATION AND THE CONFIDENTIALITY OF THEIR DATA

The purpose of this section is to provide a brief overview of the requirements for informing respondents about their participation in Federal surveys. One piece of information that can be very important to respondents is whether the Federal agency will keep their information confidential and use it only for statistical purposes. The statutory authority for such promises is also covered, as well as the requirements for documenting this authority in agency ICRs.

56. What should respondents be told about their participation in an information collection?

The Paperwork Reduction Act (PRA) requires that agencies provide certain information to respondents to help them understand why they are being asked to respond, how they are supposed to respond, and the effects the collection of information may have on them.²⁷ Within an agency, the Chief Information Officer or other designated official is responsible for ensuring that each collection of information informs and provides reasonable notice to respondents about the purpose of the study. Assuming that the basic information called for is provided, an agency can adjust the amount of detail provided depending on the scope, importance, and nature of the collection of information. For example, a brief telephone survey may call for less detail than a highly burdensome or personally intrusive written questionnaire. The following basic information must be provided to respondents:²⁸

- The reasons the information is to be collected;²⁹
- The way the information will be used to further agency purposes and serve agency needs;³⁰
- An estimate of the average burden of the collection and whom to contact about the estimate;³¹
- Whether responses to the collection of information are voluntary or mandatory, or required to obtain a benefit;³²
- The nature and extent of confidentiality to be provided, if any;³³
- The duration of respondents' expected involvement (e.g., if this is a longitudinal survey, they should be informed that they will be contacted in the future); and
- If the agency is collecting "sensitive information," respondents should be informed about what type(s) of sensitive information will be requested.

27 44 U.S.C. § 3506(c)(1)(B)(iii).

28 5 C.F.R. § 1320.8(b)(3).

29 44 U.S.C. § 3506(c)(1)(B)(iii)(I); 5 C.F.R. § 1320.8(b)(3)(i).

30 44 U.S.C. § 3506(c)(1)(B)(iii)(II); 5 C.F.R. § 1320.8(b)(3)(ii).

31 44 U.S.C. § 3506(c)(1)(B)(iii)(III); 5 C.F.R. § 1320.8(b)(3)(iii).

32 44 U.S.C. § 3506(c)(1)(B)(iii)(IV); 5 C.F.R. § 1320.8(b)(3)(iv).

33 5 C.F.R. § 1320.8(b)(3)(v). This provision was included in the regulation as a necessary component of telling the respondent of "the way such information is to be used" (44 U.S.C. § 3506(c)(1)(B)(iii)(II); see 5 C.F.R. § 1320.8(b)(3)(ii)).

Agencies that conduct research studies involving human subjects may also be required by Institutional Review Boards (IRBs) to provide additional information such as informed consent statements that are signed by the respondent. Typically, statistical surveys do not require formal consent forms.

57. What is a pledge of confidentiality and how should a pledge of confidentiality be made to respondents?

In the context of collecting data for statistical and research purposes,³⁴ an agency pledge of confidentiality "refers broadly to a quality or condition accorded to information as an obligation not to transmit that information to an unauthorized party."³⁵ Most important is that the identity of respondents not be revealed, either deliberately or inadvertently, as part of data processing and dissemination. Respondents are more likely to provide information (and in the case of "sensitive topics," the correct information) when they know the data that they provide will be kept confidential by the collecting agency. However, confidentiality is only meaningful when the agency is able to deliver the promised protection to the respondent, that is, "the data gatherer must have the will, technical ability, and moral and legal authority to protect the data."³⁶

Respondents may be given information on confidentiality in a number of different formats, depending on the mode of data collection. For a mail survey, the information is provided either in a cover letter or in a statement printed on the questionnaire. In telephone surveys, interviewers typically include a few summary sentences to potential respondents, and may refer to an advance letter that was sent. For surveys conducted by in-person interviewers, an introductory letter is usually mailed in advance or presented to the respondent. Also, an agency might provide its interviewers with a fact sheet containing answers to "frequently asked questions" or a reference to a web site or toll free number. In short, the mode of data collection determines how best to communicate the pledge of confidentiality.

Informed consent and pledges of confidentiality should be accurate and use words that are easy for the respondents to understand, taking into account their level of education. For example, a consent form for a survey of adults who have not completed high school should be composed at a basic reading level. To help ensure that respondents will understand a consent statement or confidentiality pledge, agencies should take several steps before sending a survey into the field. For example, an agency should pretest its forms, cover letters, consent statements, etc. using methods similar to those for developing and testing the survey questionnaire (see question #48).

All information collection materials such as consent forms, brochures explaining the purpose of the study and the use of the data, and so forth must be included in the ICR package submitted to

34 Confidentiality means different things and depends on the context. For example, in the classification of national security information, "confidential" is one of three classification levels, the other two being "secret" and "top secret". In such a context, "confidential is applied to information, the unauthorized disclosure of which reasonably could be expected to cause damage to the national security that the original classification authority is able to identify or describe" (White House: Executive Order 12958, Part 1, Section 1.3(3); April 17, 1995). The discussion in this document relates to confidential *statistical* information.

35 *Private Lives and Public Policies*, p. 22.

36 *Private Lives and Public Policies*, p. 23.

OMB. If an agency pledges confidentiality to respondents, it must also cite the statutory authority it has to protect the confidentiality of the information in its ICR (see question #58).

58. What legal authority does an agency have to protect the confidentiality of information it is collecting?

Before making a pledge of confidentiality, an agency must know whether or not it can protect the information. Some statistical agencies have specific legal authority to protect the confidentiality of the data they collect (e.g., the Bureau of the Census, the Bureau of Economic Analysis, the National Center for Health Statistics, and the Science Resources Statistics Division of the National Science Foundation). When agencies with statutory protection pledge confidentiality, the data cannot be used for nonstatistical "administrative purposes." For example, data collected by the Bureau of the Census are immune from legal process and cannot be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.³⁷

For surveys conducted by contractors, agencies may also be able to protect the confidentiality of responses by including such protection in the terms of the contract signed by the vendor (see question #60).

Agencies need to include in their ICRs all statements and pledges of confidentiality they are making to respondents, and they need to cite the statutory authority they have for those pledges and statements. Agencies cannot make a promise of confidentiality that they do not have statutory authority to make.

59. What is the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)?

Recent legislation has provided broad protection to information gathered solely for statistical purposes under a pledge of confidentiality. The Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)³⁸ provides uniform protection to data gathered under a pledge of confidentiality that will be used exclusively for statistical purposes. A statistical purpose is defined as the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups.³⁹ Provided that the requirements for CIPSEA are met, this law can be used by any Federal agency to protect the statistical data it collects under a pledge of confidentiality alone or in addition to the agency's existing statutory authority. This law prohibits disclosure of confidential statistical data and any nonstatistical uses of the data. Penalties for violations are a class E felony, punishable by up to five years in prison or a fine of \$250,000 or both.

³⁷ 13 U.S.C. § 9(a)(3).

³⁸ Pub. L. No. 107-347, title V.

³⁹ Pub. L. No. 107-347, Section 502(9)(A).

CIPSEA imposes strict requirements on agencies to fulfill the pledge of confidentiality. Agencies planning to use CIPSEA should consult with OMB to obtain guidance on all of the requirements, including the CIPSEA pledge, data security, use of agents, etc.⁴⁰

60. If an agency does not collect data under CIPSEA, how can it protect the confidentiality of the data?

CIPSEA cannot be used to protect data if an agency plans to use the data for nonstatistical purposes, which include the use of information in identifiable form for anything other than a statistical purpose, such as any administrative, regulatory, law enforcement, adjudicative, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent.⁴¹ However, the agency may be able to use other legal authority to protect the confidentiality of the data it has gathered. Other general Federal Government statutes that affect the confidentiality of information include the Privacy Act of 1974 and the Freedom of Information Act (FOIA). The Privacy Act can be useful in helping to ensure the confidentiality of information collected about private individuals.

The Freedom of Information Act establishes the public's right of access to Federal records. However, FOIA does have nine exemptions allowing agencies to withhold certain types of information from release. A key FOIA exemption (b)(4)⁴² allows an agency to withhold information when public release would cause substantial competitive harm. This exemption is useful when collecting proprietary information from businesses or other organizations that might be harmed if the information were publicly released. Agencies have also relied upon the Privacy Act and FOIA in some circumstances to prevent the release of information that was collected primarily for statistical and research purposes.

Agencies have also used contracts with data collection contractors to protect the confidentiality of their data. Agencies can specify in contracts that only aggregate results from the survey can be given to the sponsoring agency, and that the agency does not own and cannot receive identifiable microdata. This kind of third-party collection may also increase participation from respondents who might be hesitant to provide some kinds of information directly to an agency. For example, prior to CIPSEA, the Energy Information Administration used this kind of arrangement for the household survey on Residential Energy Consumption because the agency had no statutory authority to protect this information from release. This kind of arrangement can limit the kinds of analyses the agency can do, but may be necessary to protect the confidentiality of respondent data.

40 Please contact the Statistical and Science Policy Branch at 202-395-3093.

41 Pub. L. No. 107-347, Section 502(5).

42 5 U.S.C. § 552 (b)(4).

61. What must be done to protect data that are gathered under a pledge of confidentiality?

Agencies need to employ administrative, operational, and technical procedures to protect any data collected under a pledge of confidentiality. Administrative procedures include keeping the data in a secure environment with access limited to approved individuals. Operational procedures may include the administration of a survey in a secluded area, protection of survey forms in the possession of an interviewer, and so forth. Technical procedures are also required to ensure that data or results released do not reveal individually identifiable data.

These technical procedures are often referred to as statistical disclosure limitation (SDL) methods. SDL methods are applied to tables or microdata prior to release and include withholding release of selected data items as well as various manipulations to make data less identifiable. Data protection methods are described in Statistical Policy Working Paper #22, *Report on Statistical Disclosure Limitation Methodology* published by the Federal Committee on Statistical Methodology (FCSM). Many agencies have also found a checklist developed by the FCSM's Confidentiality and Data Access Committee (CDAC) to be very useful in assessing disclosure risks in their tables and microdata.⁴³

Although agencies must take reasonable steps to protect the confidentiality of the data they collect under a pledge of confidentiality, it is impossible to guarantee that there will be no breach of confidentiality or zero risk of disclosure.

Useful Resources

Duncan, G. T., Jabine, T. B. and de Wolf, V. A. (Eds.) (1993). *Private Lives and Public Policies*. Washington, DC: National Academy Press.

Federal Committee on Statistical Methodology (1995). *Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology*. Washington, DC: Statistical Policy Office, U.S. Office of Management and Budget. Available at <http://www.fcsm.gov/reports/>.

Interagency Confidentiality and Data Access Group (1999). "Checklist on Disclosure Potential of Proposed Data Releases." Washington, DC: Statistical Policy Office, U.S. Office of Management and Budget. Available at <http://www.fcsm.gov/committees/cdac/cdac.html>.

⁴³ These are available online at www.fcsm.gov. Go to Methodology Reports for SPWP#22, and go to Committees and then to CDAC for the Checklist.

RESPONSE RATES AND INCENTIVES

The focus in this section is on unit nonresponse or the failure to obtain any information from a selected sample member. Item nonresponse, or the failure of a respondent to respond to a specific survey item, is also discussed briefly. Nonresponse affects all surveys to varying degrees, and agencies need to consider the potential impact of nonresponse on the quality of information obtained from the survey. This section provides guidance on improving response rates and assessing potential nonresponse bias. Agencies should consult with trained survey methodologists in designing their surveys to minimize nonresponse bias.

62. Why are response rates important?

A survey's response rate is a valuable data quality and field performance indicator, and is probably the most widely cited single number associated with the generalizability of a survey's results. A high response rate increases the likelihood that the survey results reflect the views and characteristics of the target population. Conversely, a low response rate can be an indicator of potential nonresponse bias, which would be detrimental to the accuracy of the results of a study in a variety of ways, including:

- Survey estimates may be biased if those who choose to participate (respondents) differ substantially and systematically in some way from those who choose not to participate (nonrespondents). If these differences are related to critical information from the survey or the census, the results may be misleading or even erroneous.
- The standard errors of the survey estimates may also be biased because an incomplete sample may fail to capture the true variability that would be observed in a complete sample.

Nonresponse can occur for a variety of reasons, such as refusals, failure to contact the respondent, or the respondent's inability to respond due to language barriers, illness, etc. Often these different reasons for nonresponse reflect different causes, and thus, have different implications for reducing nonresponse and the potential for nonresponse bias. For example, in a household survey, noncontact may be due to respondents spending less time at home and may require more attempts by interviewers to reach them. Noncontacts may spend their time quite differently from people who are at home more, and therefore, their absence may lead to bias in survey estimates related to activities away from home. In contrast, a respondent who refuses may not be interested in the topic of the survey and may need greater persuasion as to the importance of the survey or an incentive to participate. The absence of data from the cohort of refusals may lead to bias in survey estimates of the prevalence or attitudes in the population about the main survey topic.

Agencies need to carefully consider the intended uses of the survey results and the potential impact of nonresponse bias on their data (see questions #18, #19, and #20). Agencies need to provide their best estimate for expected response rates in their ICRs and the basis for those estimates, e.g., prior surveys conducted by the agency, or similar survey methods used on similar populations by other organizations. Although response rates do not provide a clear indication of

nonresponse bias (because such bias is also a function of the differences between respondents and nonrespondents on specific survey estimates), response rates can be a useful indicator of the *risk* of nonresponse bias and should be computed and used by agencies to inform decisions on making efforts to improve cooperation and assessing potential nonresponse bias.

63. How should response rates be calculated?

Response rates have been calculated in a wide variety of ways, making comparisons across different surveys difficult. Recently, there have been attempts to standardize the calculation of response rates to provide a common basis for comparison. For example, the American Association for Public Opinion Research (AAPOR) has provided a set of six *standard definitions* of response rates as well as other formulas for calculating cooperation rates, refusal rates, and contact rates.⁴⁴ The variations in response rate calculations depend on how partial responses are considered and how cases of unknown eligibility are handled. Agencies are encouraged to use the AAPOR standard formulas in calculating and reporting response rates in their ICRs; however, agencies may use other formulas as long as the method used to calculate response rates is documented in the ICR.

At their most basic level, response rates can be viewed simply as the result of dividing the number of completed interviews/questionnaires by the number of eligible respondents who were selected to participate. Potential respondents may be split into the following categories:

1. Eligible and interview completed (c).
2. Eligible and not interviewed (e).
3. Ineligible (e.g., out of scope) (i).
4. Unable to determine eligibility (u).

Potential respondents that are eligible and not interviewed (e) may include refusals, non-contacts, non-interview due to incapacity, language difficulties, or other reasons for nonresponse. The response rate formula discussed below includes in the denominator an estimate for the proportion of cases of unknown eligibility that are actually eligible, which can be an important component for some surveys, like Random Digit Dialing (RDD) surveys, that often have many phone numbers that are never answered.

Sometimes only partial interviews are obtained due to a respondent's breaking off an interview or completing only part of a mailed questionnaire. For these cases, agencies need to set thresholds for completion of a proportion of the questionnaire or certain key items in order for the case to be counted as a completed interview. Thus, these cases would be treated either as eligible and interview completed (c) if all required items are completed or as eligible and not interviewed (e) if any required items are missing.

The total number of participants selected to be in the survey (n) is the sum of eligible and completed (c), eligible and not interviewed (e), ineligible (i), and unable to determine eligibility (u). That is $n = c + e + i + u$. Among those with unknown eligibility (u), there is a proportion (x)

44 See www.aapor.org; Go to Standards.

that is eligible. This proportion may be estimated as part of the collection process, with the most common estimate of x being $(c + e) / (c + e + i)$. The response rate is defined as

$$\text{Response rate} = c / (c + e + x u).$$

In the above formula,

- the denominator includes all original survey units that were identified as being eligible, including units with pending responses with no data received, post office returns because of “undeliverable as addressed,” and new eligible units added to the survey. The denominator does not include units deemed out-of-business, out-of-scope, or duplicates.
- the numerator includes all survey units that have submitted all the required items for the report period.

The response rate formula above is unweighted because every case is treated equally. An unweighted response rate is used to measure the proportion of the sample that resulted in useable information for analysis, and it is a useful indicator of field performance. A weighted response rate can be defined as the proportion of the survey population for which useable information is available. In some instances, the two response rates may result in identical values (if a census is taken or if a sample is selected with equal probability (see question #64).

64. When should weighted response rates be reported?

As noted in question #63, unweighted and weighted response rates may result in different values if a sample is selected with different probabilities of selection as the result of oversampling or undersampling specific subpopulations. Oversampling or undersampling of specific subpopulations occurs when the sample size for a specific subpopulation is increased (relative to the remainder of the population) to support analytic objectives and goals. For example, the analytic objectives for a study may require a sampling design with oversampling of persons in minority subpopulations or in rural areas to permit sufficiently precise estimates for these subpopulations. The oversampling of specific subpopulations will assign a higher probability of selection for units in the oversampled subpopulation than for units in the undersampled subpopulations or in the remainder of the full population. Many Federal studies use oversampling of specific subpopulations to support analyses in a cost and statistically efficient fashion.

The weighted response rate takes into account the oversampling and undersampling of the subpopulation by using the sampling weights (which are computed from the inverse of the selection probabilities). By using the sampling weight, this weighted response rate is an unbiased estimate of the proportion of the target population for which useable data are available.

Weighted response rates are often used differently in establishment surveys to take into account the relative importance assigned to different reporting units (rather than probability of selection as is done in household surveys). For example, it is common that a few very large businesses dominate an industry in terms of their production or sales with many smaller firms accounting for only a small percentage of the total production or sales within the United States. Thus, nonresponse by one or two very large businesses could jeopardize a survey estimate whereas

nonresponse by dozens of small firms may have almost no impact. In this case, weighted response rates may be constructed as the ratio of the total weighted quantity for responding units to the total weighted quantity for all eligible units to obtain a rate that reflects the proportion of the quantity being estimated that is being covered by the survey respondents.

Because unweighted and weighted response rates can provide different and useful information, agencies should generally report both in their ICRs. Whenever there are complex sample designs or the probability of selection is not equal for all cases, it is essential that weighted response rates be reported. Similarly, agencies should always report weighted response rates for establishment surveys in their ICRs and describe what is used for the weight.

65. What are typical response rates for Federal Government statistical surveys?

National surveys conducted by and for Federal statistical agencies to provide official Federal statistics generally have much larger samples, invest more resources, and achieve higher response rates than surveys sponsored by academic or commercial organizations. While some Federal surveys are mandatory, the vast majority are voluntary. For example, the Current Population Survey that provides the monthly unemployment rate is a voluntary monthly survey of over 50,000 households and has a response rate of 93 percent at the household level

The Paperwork Reduction Act does not specify a minimum response rate. In the 1980's and 1990's, many Federal surveys achieved response rates above 90 percent. Such high performance levels were well known in major household surveys, and OMB research in the 1980's showed equally high performance in many important business surveys, with a median response rate of about 90 percent across all business surveys conducted as "small censuses" or "probability samples" by major statistical agencies. In the 1990's, due to changing social and business environments, many business and household surveys saw a slippage in their response rates.

In 2001, OMB examined 199 general statistical survey information collections that were approved in 1998. OMB requested detailed information from agencies on the actual response rates achieved. These collections included mandatory and voluntary surveys, household and establishment surveys, and surveys conducted by both statistical and non-statistical agencies using a variety of data collection modes. The mean response rate was 82.2 percent (unweighted) and the median response rate was 84.7 percent. The distribution of response rates showed that about two-thirds of surveys achieved response rates above 80 percent and eighty percent of surveys achieved response rates above 70 percent. Although one might expect there to be large differences between household and establishment surveys or voluntary versus mandatory surveys, average response rates for these different types of surveys were in fact very similar. There were also small overall differences in survey response rates by mode, though it should be noted that most of the surveys were multi-mode (further information about the methods and results of this study can be found in Lu (2002)).⁴⁵

45 Lu, R. (2002). Response Rates Achieved in Government Surveys: Results from an OMB Study. *Federal Committee on Statistical Methodology Working Paper #35*. (Available at www.fcsm.gov/reports.)

The studies noted above reflect a snapshot of response rates at a particular point in time. More recent, but less systematic observations suggest that response rates have been decreasing in many ongoing surveys in the past few years. Some evidence suggests these declines have occurred more rapidly for some data collection modes (such as RDD telephone surveys) and are more pronounced for non-government surveys than Federal Government surveys. Generally, these declines have occurred despite increasing efforts and resources that have been expended to maintain or bolster response rates. It is likely that agencies will need to increase attention to their survey methods and expand innovations to continue to ensure that information gathered through Federal statistical surveys yields high quality, useful information.

The next few questions and answers are intended to help agencies evaluate their response rates, improve survey methods through the sharing of best practices, and assess potential nonresponse bias using a variety of methodologies.

66. What are acceptable response rates for different kinds of survey collections?

The 2001 OMB study of information collections described in question #65 clearly shows that the majority of Federal statistical surveys achieve good response rates. Response rates are an important indicator of the potential for non-response bias (see question #62). Clearly, the lower the response rate, the greater the caution or risk that bias can occur. Therefore, agencies should strive to obtain the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data collection costs. Agencies should also plan additional efforts to study non-response bias if projected response rates suggest the potential for bias to occur.

An agency's justification for a survey response rate should reflect, at least in part, the intended use of the data. For example, surveys collecting influential information or information that will otherwise have a substantial impact on an agency's programs or policies should be designed to minimize all sources of survey error (see question #20), including nonresponse bias. As defined in OMB and agency Information Quality Guidelines, "influential" means that "an agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions." The Information Quality Guidelines require that agencies hold the information they designate as "influential" to a higher standard of reproducibility and transparency than information that is not defined as influential under the Information Quality Guidelines (see also question #18). Agencies need to document in their ICRs the importance and use of the information and the methods they will use to achieve acceptable response rates for their collections.

In their ICRs, agencies need to report expected response rates for their surveys, which should reflect the overall unit response rate as calculated in questions #63 and #64. For ongoing surveys, the most recent actual achieved response rates should also be reported. As noted in question #62, agencies should use expected response rates as an indicator of potential risk for nonresponse bias. Agencies are encouraged to carefully consider how they can use current and new methodological tools to maximize data quality and minimize nonresponse bias. ICRs for surveys with expected response rates of 80 percent or higher need complete descriptions of the

basis of the estimated response rate and a detailed description of steps that will be taken to achieve the expected response rate (see question #69). ICRs for surveys with expected response rates lower than 80 percent need complete descriptions of how the expected response rate was determined, a detailed description of steps that will be taken to maximize the response rate (see question #69), and a description of plans to evaluate nonresponse bias (see question #71). Agencies also need a clear justification as to why the expected response rate is adequate based on the purpose of the study and the type of information that will be collected (whether influential or not). This discussion may include past experience with response rates when studying this population, prior investigations of nonresponse bias, plans to evaluate nonresponse bias, and plans to use survey methods that follow best practices that are demonstrated to achieve good response rates (see question #69). The ICR should also include a discussion of the selection of the mode of data collection and its impact on the expected response rate.

ICRs with lower response rates are often justified by agencies in cases when they are seeking to gather information that is planned for internal use only, is exploratory, or is not intended to be generalized to a target population. Examples for these kinds of collections may include some customer satisfaction and web site user surveys and other qualitative or anecdotal collections.

While the focus is often on the overall unit response rate, agencies should also pay attention to response rates for specific subgroups or levels (e.g., regions or states) for which the agency produces estimates. For example, if each state collects data from establishments within the state or if the agency produces estimates for each state, then response rates for each state should be examined. In this case, an agency may achieve an acceptable response rate at the national level, but could have substantial problems in some states that should not be overlooked.

Oftentimes, OMB may require in the terms of clearance that the agency report to OMB the actual response rate achieved at the completion of the data collection and the results of any nonresponse bias analyses or investigations. Even after the approval and fielding of an information collection, agencies should be prepared to provide detailed response rate information to OMB upon request.

67. Do longitudinal and multi-stage surveys need to achieve the same levels of response rates as other surveys?

In multi-stage and longitudinal surveys, the response rate for the last stage or latest wave is only one component of the overall response rate. While each stage or wave may have a high response rate, it is the overall unit response rate that is the most comprehensive indicator of potential nonresponse bias. Agencies that submit ICRs with multi-stage sampling plans should provide expected response rates for each stage of the sampling process, and the total response rate, taking into account all stages or prior waves. The final (or cumulative) response rate should be calculated by multiplying each stage's response rate together and should be considered an indicator for the risk of nonresponse bias and used accordingly, as noted in question #66. For these types of surveys, agencies may chose to focus their nonresponse bias analyses on a particular stage or wave that appears to be the greatest contributor towards nonresponse or take into account all stages/waves. For example, in longitudinal surveys, the response rate for each wave after the initial wave is often high, and the major contributor to the response rate may be the initial recruitment into the study. In such a case an agency may want to compare respondents

and nonrespondents to the first wave but wait to examine bias due to attrition until later waves when response rates have dropped to 80 percent or less from the first wave (see question #71).

68. Are different response rates acceptable for different modes of data collection?

Different modes of data collection typically yield differences in response rates, depending on the target population and specific methods used. For example, while a 60 percent response rate to a mail survey or RDD telephone survey may be considered quite good for some populations, such a response rate would not be considered as good for a personal visit survey. However, there are not established differences in risk for nonresponse bias by mode that would suggest that a 60 percent response rate in one mode carries a higher risk of nonresponse bias than another mode. Therefore, OMB has not set different thresholds in question #66 for different data collection modes.

Agencies need to consider how the choice of data collection mode will affect their response rates, potential for nonresponse bias, and the information that will be available to assess potential nonresponse bias (see question #71) and weigh these factors along with the other advantages and disadvantages of the modes or modes of collection they are considering. Agencies need to justify in their ICRs their choice of mode given the advantages and disadvantages of that mode (see question #37).

69. How can response rates be improved?

Regardless of the type of information collection, widely-acknowledged procedures can have a major effect on the number of respondents who complete the information request. Agencies should consult with professional survey methodologists in designing their information collections and consider answers to the following questions to maximize response rates:

- Is the agency sending an advance letter to respondents? Even for RDD surveys, agencies can obtain addresses for a large proportion of the phone numbers they have in their sample through vendors that offer reverse matching. Sending a letter in advance to inform respondents about the survey can lead to improved response rates. The letter should:
 - be signed by a senior agency official;
 - be personally addressed to the respondent if possible;
 - provide meaningful motivation for the respondent to participate;
 - answer questions of who, what, when, why, and how;
 - address how long the survey will take and whether participation is voluntary, mandatory, or required to obtain benefits (if the survey is mandatory some agencies opt to state this on the outside envelope rather than in the letter);
 - contain a contact number (toll-free phone if possible) for respondents to verify the legitimacy of the survey or ask questions; and
 - include any confidentiality pledge or assurance of anonymity (which should also be provided with the questionnaire if it is a mail survey).

- Has the agency considered ways to promote awareness of the survey? Agencies should utilize their websites and consider obtaining the endorsement of stakeholders, interest groups, and community leaders. The agency may want to conduct outreach sessions with presentations in several cities or provide news releases to trade journals, state associations, and other interested parties. It may also be possible to engage the local media for localized surveys.
- What mode of administration is being used? Has the agency carefully considered the use of more than one collection mode, e.g., following up nonrespondents to a mail survey with telephone calls, to improve response rates?
- Is the questionnaire well-designed with user-friendly formatting? Is it as brief as possible? Are the questions, instructions, and definitions easy to understand? Is the content of the survey relevant to the respondent?
- Has the agency identified strategies for contacting hard-to-reach populations?
- Does the survey allow for proxy responses? Some household surveys allow one household member to report for another member; however, whether this is practical or would have larger implications for data quality depends on the kind of information the agency is gathering.
- Has the agency investigated various survey introductions to minimize break-offs and maximize participation?
- For longitudinal surveys, has the agency considered creative ways to maintain contact with respondents between waves? For example, some agencies send newsletters or birthday cards to respondents or provide postcards that respondents can use to notify agencies of address changes. Is information from prior waves used to determine the best time to call? Are records kept of concerns respondents raise that can be addressed in future contacts?
- For longitudinal or panel surveys, does the agency go back to nonrespondents from prior waves? Some prior nonrespondents can be reached and brought back into the study in later panels or waves. Sometimes data from the missing wave can also be “filled in” based on subsequent interviews with the respondent. For example, some longitudinal studies will ask respondents about life events such as marriages, births of children, jobs, etc., that occurred since the last interview, and this information may be used to fill in questions asked in a prior missed wave.
- For personal visit and telephone surveys, has the agency clearly described the number and timing of contact attempts? More contact attempts spaced across days and times of the day offer greater potential for reaching respondents at home.
- For mail surveys, has the agency planned to conduct a follow-up contact after the first mailout with a second mailing, phone call, or a fax (if surveying businesses)? Is the agency planning to mail the survey using priority mail or a courier delivery service to distinguish it from regular mail? Is there a plan to send reminder/thank you cards and replacement questionnaires as part of nonresponse follow-up? Are there plans to allow respondents to complete the survey on the web or via phone?
- For Internet surveys, does the agency plan to use e-mail for advance notification, reminders, and follow-ups? Are respondents allowed to complete the survey on a hardcopy (to mail in) or via phone? Do nonresponse follow-up efforts include phone contact (or fax, if a business)?

- Are respondents hesitant to provide this kind of information directly to the government or the particular agency? Agencies should ensure that they can adequately protect the confidentiality of the information and communicate this to respondents (see questions #56 to #61). To provide a clearer barrier between the agency and the respondent, agencies can also consider using another agency or an external contractor to collect the data and specify in the contract that only aggregate results from the survey and no individually identifiable data can be given to the sponsoring agency.
- Does the survey allow for an increase in the length of the field period to improve low response rates?
- Have interviewers received adequate training about the survey and about interacting with respondents? Does the interviewer training include “refusal conversion” and other techniques to maximize response rates?
- For RDD surveys, has the agency purchased a higher “grade” RDD sample that removes out-of-scope numbers to minimize interviewer time spent on non-productive cases, so more time is available to spend on potential respondents?
- If the above efforts have been attempted, and there are still problems with the response rate, have incentives been tried in experiments to improve response rates (see questions #74-76 below)?

70. Given that Random Digit Dialing (RDD) telephone survey response rates have been declining, will OMB approve ICRs with this methodology?

Some recent evidence suggests that response rates to RDD surveys have been declining more rapidly than those for other modes of data collection in the past few years. RDD surveys do have some advantages for certain types of studies (see question #41); however, agencies need to carefully consider their total survey design, weighing the expected response rates they are likely to achieve using RDD methodology against its other advantages for their particular survey.

OMB has approved ICRs for RDD studies when agencies provide a clear justification in their ICRs that this is the most appropriate methodology for their study, and agencies are using appropriate methods to maximize the response rate and assess and adjust for potential nonresponse bias and coverage error.

71. How can agencies examine potential nonresponse bias?

Nonresponse bias associated with a survey statistic may be considered to have two components: the nonresponse rate and differences between respondents and nonrespondents. The lower the response rates are and the greater the differences between respondents and nonrespondents, the greater the nonresponse bias. Another way of looking at nonresponse bias is that it occurs when there is a correlation between the likelihood of participation in the survey and the survey variable(s) being measured. This view highlights the fact that some survey estimates may have nonresponse bias (because they are correlated with the likelihood of participation) while others do not.

Agencies should plan to evaluate potential nonresponse bias if they expect response rates may fall below the levels noted in question #66; these plans should be described in their ICRs. When agencies are gathering influential information (under OMB information quality guidelines) or other information with a substantial impact on programs and policies that requires high precision, agencies should consider examining potential nonresponse bias even when normally acceptable response rates are achieved.

Because nonresponse bias is particular to each survey estimate, it is possible that some survey estimates are unbiased while others have a great deal of bias. Therefore, it is important that agencies attempt to assess nonresponse bias on key survey estimates. For example, a survey on willingness to pay for some environmental improvements should assess bias on the key estimate of willingness to pay (or something highly related to it); it is not sufficient for an agency to simply look at the demographic composition of the sample compared to, for example, the latest official Census figures and, if similar, conclude there is no nonresponse bias. Similarly, agencies cannot simply assume that because the demographic composition of their achieved sample was close to the composition of the decennial census before adjustment, that there is no bias on the other substantive survey variables or that making the weighting adjustments to the demographic composition of the sample will eliminate nonresponse bias in the other variables.

Agencies should consult with professional statisticians and survey methodologists to ensure that potential nonresponse bias is addressed in the design of the study as options are far more limited after the collection has occurred. Although assessing bias and potentially adjusting the data to account for nonresponse can be complicated and time-consuming, there are a number of methods that can be used. These methods vary in the amount and kind of information that is available on respondents and nonrespondents.

At a minimum, agencies should plan to compare respondents and nonrespondents on information available from the sampling frame. Sampling frames that include data on various attributes of the population unit are helpful in examining whether response rates vary on those attributes or whether the characteristics of respondents and nonrespondents differ on these characteristics. For example, response rates from large companies versus small can be compared for establishment surveys.

In addition, agencies should seek out other available external information that they may be able to match to their sampling frame that would provide some insights into nonresponse bias. For example, agencies that survey their program participants may have other administrative data that can be matched at the individual level to compare respondents and nonrespondents more directly. If this kind of information is not available, there are other possibilities to consider, such as mapping telephone exchanges in an RDD survey to census tracts or zip codes, and then matching with aggregated data from the Census long form, permitting comparison of respondents and nonrespondents at an area level (as opposed to the specific household).

Another source of information in longitudinal surveys is to compare respondents and nonrespondents on characteristics gathered at prior waves. For some multi-stage surveys, agencies should consider including items at a screener stage that may be useful in comparing respondents and nonrespondents to the later extended interview.

When there are no good sources of information about respondents and nonrespondents on the substantive variables of interest, agencies can also use additional follow-up procedures with an abbreviated questionnaire to estimate the characteristics of nonrespondents on some key variables of interest. Sometimes these follow-up studies are done by selecting a probability sample of nonrespondents for extensive and more expensive efforts on a smaller sample that are then used to estimate the characteristics of all nonrespondents and compare to respondents.

Agencies can also assess potential nonresponse bias by analyzing differences between respondents and initial refusals (who were later “converted”) or conduct analyses of key estimates by levels of effort to obtain the response (e.g., the number of reminders sent for a mail survey or the number of calls made in a telephone survey).

Finally, agencies can also evaluate and compare different methods of nonresponse weighting adjustments using additional variables and information noted above to see what impact these have on the key survey estimates.

All of the above methods have varying strengths and weaknesses in providing useful information on nonresponse bias. Thus, agencies should attempt to use a variety of methods whenever possible.

72. What response rate issues are involved with using samples derived from pre-existing multipurpose panels, such as Internet or consumer panels?

Multi-purpose consumer and Internet survey panels are similar to multi-stage surveys or longitudinal studies in that there are several stages of agreement and participation by respondents over some period of time before they become members of the panel (see question #67). Panel members are also typically expected to participate in the panel for some fixed period of time and complete some number of surveys during that time. Often the only response rate reported for studies using these panels is based on the number of panel members who completed the specific survey and those who did not; however, this provides a very incomplete picture, because each prior stage of selection or participation, including dropping out of the panel before a respondent’s scheduled time was completed, potentially affects the representativeness of the panel, may introduce nonresponse bias, and must be taken into account in calculating the overall response rate.

For example, one vendor who conducts Internet panel surveys has documented that in 2002, a 36 percent response rate for households agreeing to join the panel was achieved. However, this was only the first stage; of those households who agreed, only 67 percent actually installed the equipment for Internet access, and only 47 percent of installed households had an adult who completed the profile and was an active panel participant available to complete a survey. Although individual survey response rates averaged 75 to 80 percent of these active members, the cumulative response rate taking into account all stages ($.36 \times .67 \times .47 \times .80$) was about 9 percent.

Because of the multiple stages of initiation of a prospective panel member and the resulting opportunities for nonresponse, different biases due to nonresponse may enter into the panel at different stages. For example, those who agree to become part of the panel may be systematically different from those who do not agree to join the panel, and those who do not install the equipment or complete the profile may be different from those who remain in the panel. Panel members also often depart from the panel before their “term” is completed, introducing further potential nonresponse bias due to attrition.

In their ICRs, agencies proposing to use multipurpose survey panels should provide a justification for their use, provide expected response rates in detail, and devote careful attention to potential nonresponse bias as warranted (see questions #66, #67, and #71). Although these panels have been used as a convenience sample and/or for pilot studies, there is some recent research that examines the quality of estimates from these panels.⁴⁶ OMB will continue to monitor this research area and evaluate results from agency studies on nonresponse bias.

Agencies should carefully consider the response rates that they are likely to achieve and the quality of the information that they will be able to obtain from pre-existing multi-purpose survey panels, taking into account the utility of the data for its intended use. While there may appear to be cost and time advantages to using a pre-existing panel, the quality of estimates obtained using this method will require careful scrutiny to ensure it is sufficient to meet its intended purposes. In their ICRs, agencies need to justify use of a multipurpose survey panel and describe how they will attempt to assess and address the potential nonresponse bias and other limitations of these panels (see question #71).

73. What should agencies do to assess and deal with nonresponse bias due to item nonresponse?

The focus of this section has been on unit nonresponse, the failure to obtain any participation from the respondent. However, even when respondents agree to participate in a survey, they do not necessarily provide all of the information that the agency requests. Thus, agencies also need to examine nonresponse to questionnaire items to see what impact this has on their results.

Agencies should note in their ICRs if substantial item nonresponse is expected for any key or sensitive items, and how this will be handled. Similar to unit nonresponse, agencies need to consider the risk of nonresponse bias at the item level. At a minimum, agencies should plan to conduct nonresponse bias analyses (see question #71) if an item missing rate exceeds 30 percent, but agencies should consider lower thresholds for key variables. Because respondents have provided information to other items on the questionnaire, there is generally a great deal of

⁴⁶ For example, see Cameron, T.A. & DeShazo, J.R. (November, 2005). Comprehensive selectivity assessment for a major consumer panel: Attitudes toward government regulation of environment, health, and safety risks. Unpublished manuscript. Krosnick et al. (May, 2005). Comparing the results of probability and nonprobability sample surveys. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Miami, Florida. Viscusi, W. K., Huber, J., & Bell, J. (2004). The value of regional water quality improvements. Available at www.law.harvard.edu/programs/olin_center/.

information from the survey itself that can be used to assess potential bias due to item nonresponse.

For key survey estimates, many large statistical surveys use a variety of statistical methods to impute values for the missing items. These imputation methods include identifying “donor” records that are similar to the case with the missing item on a variety of other variables and replacing the missing value with the value from the donor case. Other methods use regression or other statistical models to predict values for the missing variable based on complete cases and then generate a value for the missing case from this model.

Agencies should consult with trained survey statisticians on the appropriate ways to handle missing item data in their surveys. Agencies need to specify how they will handle missing item data and assess or control potential nonresponse bias, including whether the information will be imputed. If an agency uses imputation, the method that will be used should be described in the ICR.

74. What are incentives?

An incentive is defined as a positive motivational influence; something that induces action or motivates effort. Incentives are often used in market research, and sometimes used in survey research, to encourage participation. They may be monetary or non-monetary, such as phone cards, books, calculators, etc. Incentives are often unconditional; that is, they are paid prior to and regardless of a respondent’s decision to participate in the study. Research has consistently shown that giving an unconditional incentive when first contacting the respondent is more effective in obtaining cooperation than the promise of an incentive after completion of the survey.

Incentives are most appropriately used in Federal statistical surveys with hard-to-find populations or respondents whose failure to participate would jeopardize the quality of the survey data (e.g., in panel surveys experiencing high attrition), or in studies that impose exceptional burden on respondents, such as those asking highly sensitive questions, or requiring medical examinations (see question # 76).

Incentives are also often used in studies used to develop surveys. For example, research subjects who participate in cognitive research protocols and focus groups are typically paid an incentive for their participation.

Distinctions are sometimes made between an honorarium and an incentive. An honorarium is a payment given to professional individuals or institutions for services for which fees are not legally or traditionally required in order to secure their participation. Thus, this term is more appropriately used for payments to physicians, accountants, school administrators, teachers, and so forth. An honorarium is usually paid on the condition of a respondent’s participation as a token of appreciation.

75. Why must agencies provide a justification to give incentives to respondents?

While incentives have been used in the private sector without much controversy, most Federal Government surveys do not provide incentives to respondents, and the use of incentives by Federal agencies has raised a variety of concerns about their cost, the use of taxpayer funds, impact on survey responses, and implications for the “social contract” between the Federal Government and citizens. The regulations implementing the Paperwork Reduction Act (PRA) of 1980 prohibited the use of incentives for respondents to Federal surveys unless agencies could demonstrate a substantial need. The regulations implementing the 1995 reauthorization of the PRA require agencies to justify any payments to respondents.

In keeping with these concerns, OMB’s guidelines on providing incentives to respondents follow a general conceptual framework that seeks to avoid the use of incentives except when the agency has clearly justified the need for the incentive and has demonstrated positive impacts on response and data quality by using an incentive (see question #76).

76. What factors should agencies address in their justification to give incentives to respondents?

Research has consistently shown that monetary incentives are more effective in increasing survey response than nonmonetary incentives. However, agencies should still consider appropriate nonmonetary incentives, especially if they are related to the survey (or the agency more generally) and are likely to be of interest to respondents. For example, respondents to a business survey on wages may be very interested in the results to see how they compare to the industry average, and schools may find curriculum materials or books for the library an effective incentive. Other examples of nonmonetary incentives that agencies sometimes use include items directly related to the data collection, such as a folder for receipts or a calculator for respondents in a survey on expenditures. In lieu of a relevant nonmonetary incentive, agencies should consider appropriate monetary incentives (or debit cards with a PIN provided) instead of phone cards or gift certificates, as research has generally shown cash to be more effective.

In justifying their proposed use of incentives, agencies should consider the following principles, many of which overlap:

- *Data quality:* One possible justification for requesting use of an incentive is improvement in data quality. For example, agencies may be able to provide evidence that, because of an increase in response rates, an incentive will significantly improve validity and reliability to an extent beyond that possible through other means.
- *Burden on the respondent:* An agency can justify an incentive if it can demonstrate that there is a need to pay a respondent for exerting unusual effort or having an unusual reporting burden in responding to a collection of information. This type of effort can be seen in data collections that require respondents to keep daily logs for an extended period of time, participate in a medical examination, abstract information from a significant number of records, coordinate study team visits, and so forth.

- *Complex study design:* Some studies require ongoing participation of various respondents, each of whom is important to the achievement of study goals. For example, there may be a panel study over multiple years that requires participation by the same schools, teachers, parents, and students.
- *Past experience:* Agencies may be able to justify the use of incentives by relating past survey experience, results from pretests or pilot tests, or findings from similar studies. This is especially true where there is evidence of attrition and/or poor prior response rates.
- *Improved coverage of specialized respondents, rare groups, or minority populations:* A survey may have as its target population a highly selective group. Offering incentives to this population can be justified by describing the importance and difficulty of obtaining their participation in the study. For example, a study on the health of the hearing-impaired may propose providing an incentive to obtain enough respondents with particular impairments to participate. Similarly, a justification to pay an incentive for a study that involves recruiting highly specialized physicians may be considered.
- *Reduced survey costs:* If prior or similar surveys have devoted considerable resources to nonresponse follow-up, it may be possible to demonstrate that the cost of incentives will be less than the costs of extensive follow-up. While some personal visit surveys have found that the cost of incentives has been made up in reduced field interviewer time and travel costs, this is rarely true for other data collection modes. Thus, agencies should not assume that incentives will pay for themselves.
- *Equity:* Agencies should treat all respondents equally with regard to incentives. OMB generally does not approve agency plans to give incentives solely to convert refusals, or treat specific subgroups differently, unless the plan is part of an experimental design for further investigation into the effects of incentives.
- *Research into the effects of incentives:* Because OMB has expressed interest over the years in encouraging research into the effects of incentives, proposals that include experimental designs that provide insight into incentive effects are often approved. Agencies should plan to examine not only the impact on overall response rates by the use of an incentive, but also the effects on key estimates (with a similar purpose to that addressed in question #71).

OMB desk officers carefully review the justification of incentives. Agencies should cite the research literature and demonstrate how their study particularly merits use of an incentive by its similarity to specific studies on similar populations using similar methods that exist in the literature, or propose a field test or experiment to evaluate the effects of the incentive.

Agencies should either propose, or OMB may request in the terms of clearance, that results of the use of incentives will be reported to OMB.

Useful Resources

- American Association for Public Opinion Research (2004). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (available at www.aapor.org).
- Groves, R.M. and Brick, J. M. (2005). Practical Tools for Nonresponse Bias Studies. *Joint Program in Survey Methodology Short Course* (see www.jpsm.umd.edu).
- Groves, R.M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (2002). *Survey Nonresponse*. New York: Wiley.
- Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little (Eds.) *Survey Nonresponse*. New York: Wiley.

ANALYSIS AND REPORTING

The focus of this section is on the documentation that agencies need to provide in their ICRs on their plans for analyzing and reporting the information they will collect in their survey.

77. What information should agencies include in their analysis plans?

In their ICRs agencies need to provide information on their plans for analyzing and publishing the information they are collecting. The analysis plans should include a description of the statistical methods as well as any other relevant model or analytic plan that will be used to address the research questions or purposes for which the information was collected.

With respect to statistical methods, agencies should specify the estimation methods they will use, including any use of weighting. Agencies should clearly describe how weights will be derived and any adjustments that will be made to the weights to minimize potential nonresponse or coverage errors. When analyzing data from a complex survey design, agencies must ensure that appropriate statistical methods and software are used so that accurate estimates and associated variances or standard errors of those estimates are reported. For complex sample designs, specialized software is necessary that takes into account the sample design in estimating the variances. The statistical methods and software should be clearly identified in the ICR.

Often, research questions involve comparisons between groups or subgroups. Agencies should specify what statistical tests will be used to assess potential differences between the groups. The information collection should be designed with an appropriate sample size so that planned comparisons between groups or subgroups have adequate statistical power to statistically detect the differences between the groups or subgroups that are likely to exist (see question #33). Agencies should provide a power analysis in their ICRs to justify the sample size when key analyses involve comparisons among groups or subgroups (this may be included in Part B of the ICR in the justification for sample size).

When possible, agencies should include table shells or actual results from prior collections to show how the information will be presented. If detailed estimates by subgroups are planned, agencies should also describe criteria that are used to determine the amount of detail that is published in a table or figure. For example, agencies should consider criteria such as a minimum sample size, precision of the estimate, or potential disclosure risk (see question #61) in publishing estimates in tables.

Sometimes, agencies use the data gathered from a survey as inputs to models (e.g., economic forecasting models, biomechanical models) to conduct analyses. For recurring collections, agencies should submit information on the relevant models with sufficient details to allow OMB to assess the practical utility of the data being collected. For one-time collections, agencies should submit as much information as possible on the tentative models and analytic plans.

78. What predissemination review do agencies need to do for reports based on surveys or statistical collections?

Agencies are responsible for the quality of the information that they disseminate and must institute appropriate review procedures to comply with OMB and agency Information Quality Guidelines. Agencies proposing information products that involve reporting results from surveys or other statistical collections should include as part of their review process a statistical and methodological review to ensure that appropriate statistical methods are used and reported. The reviewer should have appropriate expertise in the methodology that is used, and should be provided with sufficient technical documentation to evaluate the information in the report (See OMB Peer Review Bulletin).⁴⁷

The statistical and methodological review should include an evaluation of the suitability of the statistical methods used, the accuracy of the assumptions and limitations of the data, and the appropriateness of the conclusions and technical recommendations (from a statistical perspective). The statistical and methodological review should also include examination of presentations of data in tables or figures as well as examination of any public use datasets that are released. Agencies must ensure that appropriate statistical disclosure limitation methods and procedures have been followed in keeping with the confidentiality pledge made to the respondents (see questions #57 and #61).

Useful Resources

Groves, R.M., Dillman, D. A., Eltinge, J. L. and Little, R. J. A. (2002). *Survey Nonresponse*. New York: Wiley.

Kalton, G. (1981). *Compensating for Missing Survey Data*. Ann Arbor: Institute for Social Research.

U.S. Office of Management and Budget (2005). Final Information Quality Bulletin on Peer Review. *Federal Register* 70: 2664-2677. (available at http://www.whitehouse.gov/omb/fedreg/2005/011405_peer.pdf)

⁴⁷ See www.whitehouse.gov/omb/fedreg/2005/011405_peer.pdf.

STUDIES USING STATED PREFERENCE METHODS

The focus of this section is on surveys that use stated preference methods, which are frequently used in regulatory analyses by Federal agencies. Generally, the same requirements described throughout this guidance for surveys collecting influential information apply to these surveys; however, some of these considerations are highlighted in this section with particular illustrations for this type of survey.

79. What are stated preference methods?

Stated Preference Methods (SPM) have been developed and used in the peer-reviewed literature to estimate both “use” and “non-use” values of goods and services. They have also been widely used in regulatory analyses by Federal agencies, in part because these methods can be creatively employed to address a wide variety of goods and services that are not easy to study through revealed preference methods.

The distinguishing feature of these methods is that questions about the use or non-use value of a good are posed to survey respondents in order to obtain willingness-to-pay estimates relevant to benefit or cost estimation. Some examples of SPM include contingent valuation, conjoint analysis, and risk-tradeoff analysis. The surveys used to obtain the health-utility values used in cost effectiveness analysis are similar to stated preference surveys but do not entail monetary measurement of value. Nevertheless, the principles governing quality stated preference research, with some obvious exceptions involving monetization, are also relevant in designing quality health-utility research.

80. What should agencies consider when designing questions for stated preference studies?

Stated Preference Methods have been developed and used to estimate both “use” and “non-use” values of goods and services. Because these methods pose hypothetical questions about use or non-use values to survey respondents in order to obtain willingness-to-pay estimates relevant to benefit or cost estimation, the following principles should be considered when designing these questions:

- the good or service being evaluated should be explained to the respondent in a clear, complete and objective fashion, and the survey instrument should be pre-tested;
- willingness-to-pay questions should be designed to focus the respondent on the reality of budgetary limitations and on the availability of substitute goods and alternative expenditure options; and
- the survey instrument should be designed to probe beyond general attitudes (e.g., a “warm glow” effect for a particular use or non-use value) and focus on the magnitude of the respondent’s economic valuation.

In addition, the results from the questions should be consistent with economic theory using both “internal” (within respondent) and “external” (between respondent) scope tests such as the willingness to pay is larger (smaller) when more (less) of a good is provided.

The challenge in designing quality stated preference studies is arguably greater for non-use values and unfamiliar use values than for familiar goods or services that are traded (directly or indirectly) in market transactions. The good being valued may have little meaning to respondents, and respondents may be forming their valuations for the first time in response to the questions posed. Since these values are effectively constructed by the respondent during the elicitation, the instrument and mode of administration should be rigorously pre-tested to make sure that responses are not simply an artifact of specific features of instrument design and/or mode of administration.

81. What factors should be considered when designing or evaluating studies using stated preference methods?

When designing or evaluating a stated preference study, agencies need to consider all aspects of the survey design, including coverage of the target population, mode of data collection, sampling, questionnaire design, and response rates that are covered in more detail in the preceding sections (also see question #20).

Since stated preference methods generate data from respondents in a hypothetical setting, often on complex and unfamiliar goods, special care is demanded in the design and execution of surveys, analysis of the results, and characterization of the uncertainties. A stated preference study may be the only way to obtain quantitative information about non-use values, though a number based on a poor quality study is not necessarily superior to no number at all.

There is no simple formula that can be used to determine whether a particular study is of sufficient quality to justify its use in regulatory analysis. However, OMB and agency Information Quality Guidelines require that agencies hold the information they designate as “influential” to a higher standard of reproducibility and transparency than information that is not defined as influential under the Information Quality Guidelines. As defined in OMB and agency Information Quality Guidelines, “influential” means that “an agency can reasonably determine that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions.” If SPM are used to support a Regulatory Impact Analysis (RIA), weaknesses in the study design should be acknowledged, including any resulting biases or uncertainties that are suspected. If a study has too many weaknesses with unknown consequences for the quality of the data, the study should not be used for an RIA. In employing SPM for regulatory analyses agencies should consider the following:

- documentation should be provided about the target population, the sampling frame used and its coverage of the target population, the design of the sample including any stratification or clustering, the cumulative response rate (including response rate at each stage of selection if applicable), the item non-response rate for critical questions, the exact wording and sequence of questions and other information provided to respondents, and the training of interviewers and techniques they employed (as appropriate);
- the analytic results should be consistent with economic theory using both "internal" (within respondent) and "external" (between respondents) scope tests, such as the willingness to pay is larger (smaller) when more (less) of a good is provided; and

Stated Preference Methods

- the statistical and econometric methods used to analyze the collected data should be transparent, well suited for the analysis, and applied with rigor and care.

More detailed guidance on regulatory analysis is provided in OMB Circular No. A-4.⁴⁸

⁴⁸ See www.whitehouse.gov/omb/circulars/a004/a-4.pdf.

Glossary of Abbreviations

ACASI	audio computer assisted self interview
AAPOR	American Association for Public Opinion Research
CAI	computer assisted interview
CAPI	computer assisted personal interview
CASI	computer assisted self interview
CATI	computer assisted telephone interview
CDAC	Confidentiality and Data Access Committee
CFR	Code of Federal Regulations
CIPSEA	Confidential Information Protection and Statistical Efficiency Act
CPS	Current Population Survey
FCSM	Federal Committee on Statistical Methodology
FOIA	Freedom of Information Act
GPEA	Government Paperwork Elimination Act
ICR	Information Collection Request
IQG	Information Quality Guidelines
IVR	interactive voice response
MSA	Metropolitan Statistical Area
NAICS	North American Industry Classification System
PRA	Paperwork Reduction Act
OMB	Office of Management and Budget
RDD	random digit dialing
RIA	regulatory impact analysis
SDL	statistical disclosure limitation
SIC	Standard Industrial Classification
SIPP	Survey of Income and Program Participation
SOC	Standard Occupational Classification
SPM	stated preference method
SRS	simple random sample
TDE	touchtone data entry
USC	United States Code

Glossary of Terms

-A-

Administrative records are information kept by business establishments, institutions, and governments primarily for their own purposes in running their business or program. Respondents may need to refer to these records in order to answer questions on Federal surveys.

-B-

Behavior coding is a technique used for pretesting that involves monitoring the interaction between interviewers and respondents (often through reviewing tape recordings) and coding certain behaviors, such as the interviewer misreading the question or a respondent asking for clarification, in order to identify problem questions.

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.

-C-

A **case study** is a method for learning about a complex instance, based on a comprehensive understanding of that instance obtained by extensive description and analysis of that instance taken as a whole and in its context.

A **census** survey is a survey of the entire universe or target population that is of interest.

Clustering refers to a sample design in which geographic groups are formed (clusters) for purposes of sampling in order to reduce the costs of interviewer travel.

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

Cognitive interviews are used to test and refine questionnaires. In a cognitive interview, respondents are required to report aloud everything they are thinking as they attempt to answer a survey question.

A **complex sample design** is one that involves multiple stages, stratification, unequal probabilities of selection or clustering.

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

A **convenience sample** is a nonprobability sample that is drawn from units of the population of interest that are close at hand or willing to participate.

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 **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 percent cut-off sample consists of the largest units accounting for at least 90 percent 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.

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.

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

-E-

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

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

An **establishment survey** is a survey of a business establishment.

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.

An **experimental design** is a type of research design in which the researcher controls and manipulates conditions in order to assess the effect on some outcome of interest. Experiments are conducted when researchers want to be able to infer causality.

An **expert choice sample** is a nonprobability sample in which an “expert” specifically chooses sample elements with certain characteristics to mimic ‘typical’ or ‘representative’ members of the 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.

Forecasting involves 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.

-G-

A **gatekeeper** is a person who is between the interviewer and the respondent and may prevent the interviewer from gaining access to the respondent. In an establishment survey, secretaries or administrative assistants may control what mail and telephone calls reach a respondent and, thus, act as gatekeepers.

-H-

Honoraria are payments given to professional individuals or institutions for services for which fees are not legally or traditionally required in order to secure their participation. Thus, this term is more appropriately used for payments to physicians, CPAs, schools, administrators, teachers, and so forth. An honorarium is usually paid on the condition of a respondent's participation as a token of appreciation.

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.

An **incentive** is a positive motivational influence; something that induces action or motivates effort. Incentives are sometimes used in surveys to increase cooperation.

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.

Interactive voice response refers to a method for data collection in which a computer "reads" the question to respondents over the phone, and respondents reply by using the keypad or saying their answers aloud.

Internal validity refers to the soundness of an experimental design so that the results reflect only the differences the researcher intended by manipulating the conditions, and the absence of alternative explanations for the results from the experiment.

An **Internet panel** consists of a large sample of respondents who have volunteered or been recruited to participate in many surveys on the Internet over some period of time.

Interviewer bias refers to effects that interviewers may have that bias the survey results. One bias that may occur is that respondents will report fewer socially undesirable attitudes or behaviors to an interviewer.

Interviewer debriefing is a method used in pretesting in which interviewers are asked to report any problems that they perceived that respondents had with the questions in the survey instrument.

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).

-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.

A **mixed mode** survey is one that uses more than one mode for data collection, for example, a mail survey is initially sent to respondents, but nonrespondents are called on the telephone to complete the survey.

Mode of data collection refers to whether the information is gathered from respondents in a face-to-face interview, over the telephone, from a mail survey, or via a web survey.

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

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

A **multi-stage design** for a sample is a complex design in which the sample is selected in stages because a comprehensive listing of sample elements is not available. One example of a multi-stage sample design for an area sample is first states may be selected, then counties, then a

census tract or block group within the county. The sample elements are then listed and finally selected for inclusion into the sample.

-N-

Nonprobabilistic methods are methods for selecting a sample that do not select sample elements such that each one has some known nonzero probability of being selected into the sample.

A **nonprobability sample** is a sample in which sample elements are not chosen so that each one has some known nonzero probability of being selected into the sample. Common examples of nonprobability samples are convenience sample, quota samples, and expert choice samples.

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.

An **oversample** refers to selecting some sample elements with probabilities greater than their proportion in the population usually in order to have a large enough sample to make separate estimates for a subgroup with greater precision than would be possible if the group was selected in proportion to its representation in the population.

-P-

A **panel survey** is a survey that is repeated with the same respondents over some period of time. 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.”

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.

A **probability sample** is a sample that is selected so that each sample member has some known nonzero probability of being selected into the sample.

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.

Proxy responses are responses provided by anyone other than the respondent who is reporting for the respondent or providing information about the respondent. For example, parents often report for young children in surveys.

A **public-use data file or public-use microdata file** 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-

A **qualitative study** is one done with the goal of obtaining rich information and insight to describe or explore a phenomena through a variety of methods such as focus groups, case studies, in-depth interviews, and observations.

A **quasi-experimental design** is one in which the researcher is unable to assign persons randomly to conditions but is still able to have some control over the conditions.

A **quota sample** samples are samples where units are selected nonrandomly based on a quota. The quota may be defined such that the final numbers of participating units with given characteristics have the same proportion as corresponding units have in the population.

-R-

A **RDD survey** refers to Random Digit Dialing, which is a means of generating telephone numbers to select a random sample.

Refusal conversion refers to the practice of having interviewers call back (either on the telephone or in-person) a respondent who has refused and attempt to gain his or her cooperation in the survey. Often the most effective interviewers in an organization are selected to do refusal conversions.

Reinterview refers to asking usually only a small subset of respondents to take the survey again or answer a subset of the survey questions again. Reinterviews may be conducted to assess interviewers or statistical properties of questions in the survey.

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.

Respondent debriefing is a pretesting method in which respondents are asked questions about the survey questions after they have completed the survey. Respondents may be asked to report if they had any trouble understanding any of the questions or found any questions unclear or confusing.

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

Response rates: see weighted response rate and unweighted response rate.

-S-

A **sample survey** selects respondents from only a portion of the total target population using probability methods in order to make an inference about the target population.

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.

A **sampling frame** is a list or set of procedures for identifying all elements of a target population from which one can draw a sample.

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.

Sampling weights are the inverse of the probability of selection for a sample element.

A **simple random sample** is one in which every member of the population has an equal probability of being selected, and all samples of a given size have the same probability of being selected.

Skip patterns are used in questionnaires to indicate when the next question should be skipped because of the respondent's answer to the current question. On paper questionnaires, skip patterns may be indicated by arrows or instructions to go to a specific item. Computer-assisted interviews have the skip patterns programmed into the instrument.

A **snowball sample** is a sample that is built by asking a respondent to provide the name of someone that he or she knows for the researcher to contact. It can be a useful technique to build a sample of individuals who have a rare condition and know one another.

Social desirability is a bias that occurs when respondents answer questions in a manner designed to present themselves favorably.

In a **split panel design**, respondents are randomly assigned into different groups to receive different versions of the questions to measure any differences between the different question wordings.

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.

Statistical disclosure limitation methods refer to a variety of techniques and rules for reducing the amount of information that is presented that could be used identify a respondent from data in tables or in microdata files.

Statistical methods include the use of sampling, estimation, imputation, or techniques for modeling or data analysis.

Strata are created by partitioning the frame and are generally defined to include relatively homogeneous units within strata.

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.

Survey panel: see panel survey.

-T-

The **target population** is any group of potential sample units or persons, businesses, or other entities of interest.

Touchtone data entry refers to using the touchtone key pad to enter numeric information in response to a survey question.

-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).

The **unweighted response rate** is the response rate calculated using the direct counts of the number of completed interviews, noncontacts, refusals, etc. not taking into account any differential probabilities of selection.

Usability testing involves some assessment of how well a survey instrument can be used in practice by an interviewer or a respondent. Some Federal agencies have usability laboratories where they can observe and record respondent's behavior interacting with a computerized survey instrument.

-V-

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.

A **weighted response rate** is the response rate calculated using the counts of the number of completed interviews, noncontacts, refusals, etc. taking into account the probabilities of selection for each case to measure the proportion of the sampling frame that is represented by the responding units.

Appendices

PAPERWORK REDUCTION ACT SUBMISSION

Please read the instructions before completing this form. For additional forms or assistance in completing this form, contact your agency's Paperwork Clearance Officer. Send two copies of this form, the collection instrument to be reviewed, the Supporting Statement, and any additional documentation to: **Office of Information and Regulatory Affairs, Office of Management and Budget, Docket Library, Room 10102, 725 17th Street NW, Washington, DC 20503.**

1. Agency/Subagency originating request	2. OMB control number b. ☐ None a. _____ - _____
3. Type of information collection (<i>check one</i>) a. ☐ New Collection b. ☐ Revision of a currently approved collection c. ☐ Extension of a currently approved collection d. ☐ Reinstatement, without change , of a previously approved collection for which approval has expired e. ☐ Reinstatement, with change , of a previously approved collection for which approval has expired f. ☐ Existing collection in use without an OMB control number <i>For b-f, note Item A2 of Supporting Statement instructions</i>	4. Type of review requested (<i>check one</i>) a. ☐ Regular b. ☐ Emergency - Approval requested by: ____/____/____ c. ☐ Delegated 5. Small entities Will this information collection have a significant economic impact on a substantial number of small entities? ☐ Yes ☐ No
7. Title	6. Requested expiration date a. ☐ Three years from the approval date b. ☐ Other: ____/____
8. Agency form number(s) (<i>if applicable</i>)	
9. Keywords	
10. Abstract	
11. Affected public (<i>Mark primary with "P" and all others with "X"</i>) a. ☐ Individuals or households d. ☐ Farms b. ☐ Business or other for-profit e. ☐ Federal Government c. ☐ Not-for-profit institutions f. ☐ State, Local, or Tribal Government	12. Obligation to respond (<i>Mark primary with "P" and all others that apply with "X"</i>) a. ☐ Voluntary b. ☐ Required to obtain or retain benefits c. ☐ Mandatory
13. Annual reporting and recordkeeping hour burden a. Number of respondents _____ b. Total annual responses _____ 1. Percentage of these responses collected electronically _____ % c. Total annual hours requested _____ d. Current OMB inventory _____ e. Difference _____ f. Explanation of difference 1. Program change _____ 2. Adjustment _____	14. Annual reporting and recordkeeping cost burden (<i>in thousands of dollars</i>) a. Total annualized capital/startup costs _____ b. Total annual costs (O&M) _____ c. Total annualized cost requested _____ d. Current OMB inventory _____ e. Difference _____ f. Explanation of difference 1. Program change _____ 2. Adjustment _____
15. Purpose of information collection (<i>Mark primary with "P" and all others that apply with "X"</i>) a. ☐ Application for benefits e. ☐ Program planning or management b. ☐ Program evaluation f. ☐ Research c. ☐ General purpose statistics g. ☐ Regulatory or compliance d. ☐ Audit	16. Frequency of recordkeeping or reporting (<i>check all that apply</i>) a. ☐ Recordkeeping b. ☐ Third party disclosure c. ☐ Reporting: 1. ☐ On occasion 2. ☐ Weekly 3. ☐ Monthly 4. ☐ Quarterly 5. ☐ Semi-annually 6. ☐ Annually 7. ☐ Biennially 8. ☐ Other (describe) _____
17. Statistical methods Does this information collection employ statistical methods? ☐ Yes ☐ No	18. Agency contact (<i>person who can best answer questions regarding the content of this submission</i>) Name: _____ Phone: _____

19. Certification for Paperwork Reduction Act Submissions

On behalf of this Federal agency, I certify that the collection of information encompassed by this request complies with 5 CFR 1320.9.

NOTE: The text of 5 CFR 1320.9, and the related provisions of 5 CFR 1320.8(b)(3), appear at the end of the instructions. *The certification is to be made with reference to those regulatory provisions as set forth in the instructions.*

The following is a summary of the topics, regarding the proposed collection of information, that the certification covers:

- (a) It is necessary for the proper performance of agency functions;
- (b) It avoids unnecessary duplication;
- (c) It reduces burden on small entities;
- (d) It uses plain, coherent, and unambiguous language that is understandable to respondents;
- (e) Its implementation will be consistent and compatible with current reporting and recordkeeping practices;
- (f) It indicates the retention periods for recordkeeping requirements;
- (g) It informs respondents of the information called for under 5 CFR 1320.8 (b)(3) about:
 - (i) Why the information is being collected;
 - (ii) Use of information;
 - (iii) Burden estimate;
 - (iv) Nature of response (voluntary, required for a benefit, or mandatory);
 - (v) Nature and extent of confidentiality; and
 - (vi) Need to display currently valid OMB control number;
- (h) It was developed by an office that has planned and allocated resources for the efficient and effective management and use of the information to be collected (see note in Item 19 of the instructions);
- (i) It uses effective and efficient statistical survey methodology (if applicable); and
- (j) It makes appropriate use of information technology.

If you are unable to certify compliance with any of these provisions, identify the item below and explain the reason in Item 18 of the Supporting Statement.

Signature of Senior Official or designee

Date

Instructions For Completing OMB Form 83-I

Please answer all questions and have the Senior Official or designee sign the form. These instructions should be used in conjunction with 5 CFR 1320, which provides information on coverage, definitions, and other matters of procedure and interpretation under the Paperwork Reduction Act of 1995.

1. Agency/Subagency originating request

Provide the name of the agency or subagency originating the request. For most cabinet-level agencies, a subagency designation is also necessary. For non-cabinet agencies, the subagency designation is generally unnecessary.

2. OMB control number

- a. If the information collection in this request has previously received or now has an OMB control or comment number, enter the number.
- b. Check "None" if the information collection in this request has not previously received an OMB control number. Enter the four digit agency code for your agency.

3. Type of information collection (check one)

- a. Check "New collection" when the collection has not previously been used or sponsored by the agency.
- b. Check "Revision" when the collection is currently approved by OMB, and the agency request includes a material change to the collection instrument, instructions, its frequency of collection, or the use to which the information is to be put.
- c. Check "Extension" when the collection is currently approved by OMB, and the agency wishes only to extend the approval past the current expiration date without making any material change in the collection instrument, instructions, frequency of collection, or the use to which the information is to be put.
- d. Check "Reinstatement without change" when the collection previously had OMB approval, but the approval has expired or was withdrawn before this submission was made, and there is no change to the collection.
- e. Check "Reinstatement with change" when the collection previously had OMB approval, but the approval has expired or was withdrawn before this submission was made, and there is change to the collection.
- f. Check "Existing collection in use without OMB control number" when the collection is currently in use but does not have a currently valid OMB control number.

4. Type of review requested (check one)

- a. Check "Regular" when the collection is submitted under 5 CFR 1320.10, 1320.11, or 1320.12 with a standard 60 day review schedule.
- b. Check "Emergency" when the agency is submitting the request under 5 CFR 1320.13 for emergency processing and provides the required supporting material. Provide the date by which the agency requests approval.
- c. Check "Delegated" when the agency is submitting the collection under the conditions OMB has granted the agency delegated authority.

5. Small entities

Indicate whether this information collection will have a significant impact on a substantial number of small entities. A small entity may be (1) a small business which is deemed to be one that is independently owned and operated and that is not dominant in its field of operation; (2) a small organization that is any not-for-profit enterprise that is independently owned and operated and is not dominant in its field; or (3) a small government jurisdiction which is a government of a city, county, town, township, school district, or special district with a population of less than 50,000.

6. Requested expiration date

- a. Check "Three years" if the agency requests a three year approval for the collection.
- b. Check "Other" if the agency requests approval for less than three years. Specify the month and year of the requested expiration date.

7. Title

Provide the official title of the information collection. If an official title does not exist, provide a description which will distinguish this collection from others.

8. Agency form number(s) (if applicable)

Provide any form number the agency has assigned to this collection of information. Separate each form number with a comma.

9. Keywords

Select and list at least two keywords (descriptors) from the "Federal Register Thesaurus of Indexing Terms" that describe the subject area(s) of the information collection. Other terms may be used but should be listed after those selected from the thesaurus. Separate keywords with commas. Keywords should not exceed two lines of text.

10. Abstract

Provide a statement, limited to five lines of text, covering the agency's need for the information, uses to which it will be put, and a brief description of the respondents.

11. Affected public

Mark all categories that apply, denoting the primary public with a "P" and all others that apply with "X."

12. Obligation to respond

Mark all categories that apply, denoting the primary obligation with a "P" and all others that apply with "X."

- a. Mark "Voluntary" when the response is entirely discretionary and has no direct effect on any benefit or privilege for the respondent.
- b. Mark "Required to obtain or retain benefits" when the response is elective, but is required to obtain or retain a benefit.
- c. Mark "Mandatory" when the respondent must reply or face civil or criminal sanctions.

13. Annual reporting and recordkeeping hour burden

- a. Enter the number of respondents and/or recordkeepers. If a respondent is also a recordkeeper, report the respondent only once.
- b. Enter the number of responses provided annually. For recordkeeping as compared to reporting activity, the number of responses equals the number of recordkeepers.
 - b1. Enter the estimated percentage of responses that will be submitted/collected electronically using magnetic media (i.e., diskette), electronic mail, or electronic data interchange. Facsimile is **not** considered an electronic submission.
- c. Enter the total annual recordkeeping and reporting hour burden.
- d. Enter the burden hours currently approved by OMB for this collection of information. Enter zero (0) for any new submission or for any collection whose OMB approval has expired.
- e. Enter the difference by subtracting line d from line c. Record a negative number (d larger than c) within parentheses.
- f. Explain the difference. The difference in line e must be accounted for in lines f.1. and f.2.
 - f.1. "Program change" is the result of deliberate Federal government action. All new collections and any subsequent revision of existing collections (e.g., the addition or deletion of questions) are recorded as program changes.
 - f.2. "Adjustment" is a change that is not the result of a deliberate Federal government action. Changes resulting from new estimates or action not controllable by the Federal government are recorded as adjustments.

14. Annual reporting and recordkeeping cost burden (in thousands of dollars)

The costs identified in this item must exclude the cost of hour burden identified in Item 13.

- a. Enter the total dollar amount of annualized cost for all respondents of any associated capital or start-up costs.
- b. Enter recurring annual dollar amount of cost for all respondents associated with operating or maintaining systems or purchasing services.
- c. Enter total (14.a. + 14.b.) annual reporting and recordkeeping cost burden.
- d. Enter any cost burden currently approved by OMB for this collection of information. Enter zero (0) if this is the first submission after October 1, 1995.
- e. Enter the difference by subtracting line d from line c. Record a negative number (d larger than c) within parenthesis.
- f. Explain the difference. The difference in line e must be accounted for in lines f.1. and f.2.
 - f.1. "Program change" is the result of deliberate Federal government action. All new collections and any subsequent revisions or changes resulting in cost changes are recorded as program changes.

f.2. "Adjustment" is a change that is not the result of a deliberate Federal government action. Changes resulting from new estimations or actions not controllable by the Federal government are recorded as adjustments.

15. Purpose of information collection

Mark all categories that apply, denoting the primary purpose with a "P" and all others that apply with "X."

a. Mark "Application for benefits" when the purpose is to participate in, receive, or qualify for a grant, financial assistance, etc., from a Federal agency or program.

b. Mark "Program evaluation" when the purpose is a formal assessment, through objective measures and systematic analysis, of the manner and extent to which Federal programs achieve their objectives or produce other significant effects.

c. Mark "General purpose statistics" when the data is collected chiefly for use by the public or for general government use without primary reference to the policy or program operations of the agency collecting the data.

d. Mark "Audit" when the purpose is to verify the accuracy of accounts and records.

e. Mark "Program planning or management" when the purpose relates to progress reporting, financial reporting and grants management, procurement and quality control, or other administrative information that does not fit into any other category.

f. Mark "Research" when the purpose is to further the course of research, rather than for a specific program purpose.

g. Mark "Regulatory or compliance" when the purpose is to measure compliance with laws or regulations.

16. Frequency of recordkeeping or reporting

Check "Recordkeeping" if the collection of information explicitly includes a recordkeeping requirement.

Check "Third party disclosure" if a collection of information includes third-party disclosure requirements as defined by 1320.3(c).

Check "Reporting" for information collections that involve reporting and check the frequency of reporting that is requested or required of a respondent. If the reporting is on "an event" basis, check "On occasion."

17. Statistical methods

Check "Yes" if the information collection uses statistical methods such as sampling or imputation. Generally, check "No" for applications and audits (unless a random auditing scheme is used). Check "Yes" for statistical collections, most research collections, and program evaluations using scientific methods. For other types of data collection, the use of sampling, imputation, or other statistical estimation techniques should dictate the response for this item. Ensure that supporting documentation is provided in accordance with Section B of the Supporting Statement.

18. Agency contact

Provide the name and telephone number of the agency person best able to answer questions regarding the content of this submission.

19. Certification for Paperwork Reduction Act Submissions

The Senior Official or designee signing this statement certifies that the collection of information encompassed by the request complies with 5 CFR 1320.9. Provisions of this certification that the agency cannot comply with should be identified here and fully explained in item 18 of the attached Supporting Statement. NOTE: The Office that "develops" and "uses" the information to be collected is the office that "conducts or sponsors" the collection of information. (See 5 CFR 1320.3(d)).

Certification Requirement for Paperwork Reduction Act Submissions

5 CFR 1320.9 reads "As part of the agency submission to OMB of a proposed collection of information, the agency (through the head of the agency, the Senior Official, or their designee) shall certify (and provide a record supporting such certification) that the proposed collection of information--

"(a) is necessary for the proper performance of the functions of the agency, including that the information to be collected will have practical utility;

"(b) is not unnecessarily duplicative of information otherwise reasonably accessible to the agency;

"(c) reduces to the extent practicable and appropriate the burden on persons who shall provide information to or for the agency, including with respect to small entities, as defined in the Regulatory Flexibility Act (5 U.S.C. § 601(6)), the use of such techniques as:

"(1) establishing differing compliance or reporting requirements or timetables that take into account the resources available to those who are to respond;

"(2) the clarification, consolidation, or simplification of compliance and reporting requirements; or collections of information, or any part thereof;

"(3) an exemption from coverage of the collection of information, or any part thereof;

"(d) is written using plain, coherent, and unambiguous terminology and is understandable to those who are to respond;

"(e) is to be implemented in ways consistent and compatible, to the maximum extent practicable, with the existing reporting and recordkeeping practices of those who are to respond;

"(f) indicates for each recordkeeping requirement the length of time persons are required to maintain the records specified;

"(g) informs potential respondents of the information called for under §1320.8(b)(3); [see below]

"(h) has been developed by an office that has planned and allocated resources for the efficient and effective management and use of the information to be collected, including the processing of the information in a manner which shall enhance, where appropriate, the utility of the information to agencies and the public;

"(i) uses effective and efficient statistical survey methodology appropriate to the purpose for which the information is to be collected; and

"(j) to the maximum extent practicable, uses appropriate information technology to reduce burden and improve data quality, agency efficiency and responsiveness to the public."

NOTE: 5 CFR 1320.8(b)(3) requires that each collection of information:

"(3) informs and provides reasonable notice to the potential persons to whom the collection of information is addressed of:

"(i) the reasons the information is planned to be and/or has been collected;

"(ii) the way such information is planned to be and/or has been used to further the proper performance of the functions of the agency;

"(iii) an estimate, to the extent practicable, of the average burden of the collection (together with a request that the public direct to the agency any comments concerning the accuracy of this burden estimate and any suggestions for reducing this burden);

"(iv) whether responses to the collection of information are voluntary, require to obtain or retain a benefit (citing authority) or mandatory (citing authority);

"(v) the nature and extent of confidentiality to be provided, if any (citing authority); and

"(vi) the fact that an agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number."

Supporting Statement for Paperwork Reduction Act Submissions

General Instructions

A Supporting Statement, including the text of the notice to the public required by 5 CFR 1320.5(a)(i)(iv) and its actual or estimated date of publication in the Federal Register, must accompany each request for approval of a collection of information. The Supporting Statement must be prepared in the format described below, and must contain the information specified in Section A below. If an item is not applicable, provide a brief explanation. When Item 17 of the OMB Form 83-I is checked "Yes", Section B of the Supporting Statement must be completed. OMB reserves the right to require the submission of additional information with respect to any request for approval.

Specific Instructions

A. Justification

1. Explain the circumstances that make the collection of information necessary. Identify any legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.
2. Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.
3. Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, e.g., permitting electronic submission of responses, and the basis for the decision for adopting this means of collection. Also describe any consideration of using information technology to reduce burden.
4. Describe efforts to identify duplication. Show specifically why any similar information already available cannot be used or modified for use for the purposes described in Item 2 above.
5. If the collection of information impacts small businesses or other small entities (Item 5 of OMB Form 83-I), describe any methods used to minimize burden.
6. Describe the consequence to Federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.
7. Explain any special circumstances that would cause an information collection to be conducted in a manner:
 - * requiring respondents to report information to the agency more often than quarterly;
 - * requiring respondents to prepare a written response to a collection of information in fewer than 30 days after receipt of it;
 - * requiring respondents to submit more than an original and two copies of any document;

- * requiring respondents to retain records, other than health, medical, government contract, grant-in-aid, or tax records, for more than three years;
- * in connection with a statistical survey, that is not designed to produce valid and reliable results that can be generalized to the universe of study;
- * requiring the use of a statistical data classification that has not been reviewed and approved by OMB;
- * that includes a pledge of confidentiality that is not supported by authority established in statute or regulation, that is not supported by disclosure and data security policies that are consistent with the pledge, or which unnecessarily impedes sharing of data with other agencies for compatible confidential use; or
- * requiring respondents to submit proprietary trade secrets, or other confidential information unless the agency can demonstrate that it has instituted procedures to protect the information's confidentiality to the extent permitted by law.

8. If applicable, provide a copy and identify the date and page number of publication in the Federal Register of the agency's notice, required by 5 CFR 1320.8(d), soliciting comments on the information collection prior to submission to OMB. Summarize public comments received in response to that notice and describe actions taken by the agency in response to these comments. Specifically address comments received on cost and hour burden.

Describe efforts to consult with persons outside the agency to obtain their views on the availability of data, frequency of collection, the clarity of instructions and recordkeeping, disclosure, or reporting format (if any), and on the data elements to be recorded, disclosed, or reported.

Consultation with representatives of those from whom information is to be obtained or those who must compile records should occur at least once every 3 years - even if the collection of information activity is the same as in prior periods. There may be circumstances that may preclude consultation in a specific situation. These circumstances should be explained.

9. Explain any decision to provide any payment or gift to respondents, other than reenumeration of contractors or grantees.

10. Describe any assurance of confidentiality provided to respondents and the basis for the assurance in statute, regulation, or agency policy.

11. Provide additional justification for any questions of a sensitive nature, such as sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private. This justification should include the reasons why the agency considers the questions necessary, the specific uses to be made of the information, the explanation to be given to persons from whom the information

is requested, and any steps to be taken to obtain their consent.

12. Provide estimates of the hour burden of the collection of information. The statement should:
 - * Indicate the number of respondents, frequency of response, annual hour burden, and an explanation of how the burden was estimated. Unless directed to do so, agencies should not conduct special surveys to obtain information on which to base hour burden estimates. Consultation with a sample (fewer than 10) of potential respondents is desirable. If the hour burden on respondents is expected to vary widely because of differences in activity, size, or complexity, show the range of estimated hour burden, and explain the reasons for the variance. Generally, estimates should not include burden hours for customary and usual business practices.
 - * If this request for approval covers more than one form, provide separate hour burden estimates for each form and aggregate the hour burdens in Item 13 of OMB Form 83-I.
 - * Provide estimates of annualized cost to respondents for the hour burdens for collections of information, identifying and using appropriate wage rate categories. The cost of contracting out or paying outside parties for information collection activities should not be included here. Instead, this cost should be included in Item 13.

13. Provide an estimate for the total annual cost burden to respondents or recordkeepers resulting from the collection of information. (Do not include the cost of any hour burden shown in Items 12 and 14).

* The cost estimate should be split into two components: (a) a total capital and start-up cost component (annualized over its expected useful life) and (b) a total operation and maintenance and purchase of services component. The estimates should take into account costs associated with generating, maintaining, and disclosing or providing the information. Include descriptions of methods used to estimate major cost factors including system and technology acquisition, expected useful life of capital equipment, the discount rate(s), and the time period over which costs will be incurred. Capital and start-up costs include, among other items, preparations for collecting information such as purchasing computers and software; monitoring, sampling, drilling and testing equipment; and record storage facilities.

* If cost estimates are expected to vary widely, agencies should present ranges of cost burdens and explain the reasons for the variance. The cost of purchasing or contracting out information collections services should be a part of this cost burden estimate. In developing cost burden estimates, agencies may consult with a sample of respondents (fewer than 10), utilize the 60-day pre-OMB submission public comment process and use

existing economic or regulatory impact analysis associated with the rulemaking containing the information collection, as appropriate.

* Generally, estimates should not include purchases of equipment or services, or portions thereof, made: (1) prior to October 1, 1995, (2) to achieve regulatory compliance with requirements not associated with the information collection, (3) for reasons other than to provide information or keep records for the government, or (4) as part of customary and usual business or private practices.

14. Provide estimates of annualized costs to the Federal government. Also, provide a description of the method used to estimate cost, which should include quantification of hours, operational expenses (such as equipment, overhead, printing, and support staff), and any other expense that would not have been incurred without this collection of information. Agencies may also aggregate cost estimates from Items 12, 13, and 14 in a single table.

15. Explain the reasons for any program changes or adjustments reported in Items 13 or 14 of the OMB Form 83-I.

16. For collections of information whose results will be published, outline plans for tabulation and publication. Address any complex analytical techniques that will be used. Provide the time schedule for the entire project, including beginning and ending dates of the collection of information, completion of report, publication dates, and other actions.

17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons that display would be inappropriate.

18. Explain each exception to the certification statement identified in Item 19, "Certification for Paperwork Reduction Act Submissions," of OMB Form 83-I.

B. Collections of Information Employing Statistical Methods

The agency should be prepared to justify its decision not to use statistical methods in any case where such methods might reduce burden or improve accuracy of results. When Item 17 on the Form OMB 83-I is checked, "Yes," the following documentation should be included in the Supporting Statement to the extent that it applies to the methods proposed:

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

2. Describe the procedures for the collection of information including:

- * Statistical methodology for stratification and sample selection,
- * Estimation procedure,
- * Degree of accuracy needed for the purpose described in the justification,
- * Unusual problems requiring specialized sampling procedures, and
- * Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.



ADMINISTRATOR
OFFICE OF
INFORMATION AND
REGULATORY AFFAIR

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

January 20, 2006

MEMORANDUM FOR THE PRESIDENT'S MANAGEMENT COUNCIL

FROM: John D. Graham, Ph.D. *JDG*
Administrator

SUBJECT: Guidance on Agency Survey and Statistical Information Collections

The Paperwork Reduction Act of 1995 requires that Federal agency information collections employ effective and efficient survey and statistical methodologies appropriate to the purpose for which the information is to be collected. It further directs the Office of Management and Budget (OMB) to develop and oversee the implementation of Government-wide policies, principles, standards, and guidelines concerning statistical collection procedures and methods.

The attached guidance document, entitled "Questions and Answers When Designing Surveys for Information Collections" (Q&A), provides details about the OMB review process, assistance in strengthening supporting statements for information collection requests, and, most importantly, advice for improving information collection designs. The document was circulated for agency comment on December 14, 2004, and has been revised in response to comments from agencies and external peer reviewers.

The content of this document is focused on what agencies need to consider when designing information collections and preparing requests for OMB approval. The guidance addresses issues that frequently arise in OMB reviews, including topics ranging from basic procedural requirements to best practices for technical documentation of surveys. It has been written for a wide audience. We anticipate that the document will be updated and revised as developments warrant so that the guidance will remain current with professional practice and useful to the agencies. Ultimately, we hope the Q&A's will serve to improve the quality of Federal surveys and statistical information.

Please share the attached Q&A document with appropriate program managers and paperwork clearance officers in your agency.

Attachment

Revision History

Version	Date Published	Comments
2.0	October 2016	Incorporated references to updated standards and guidelines for cognitive interviewing in the Questionnaire Design and Development section
1.0	January 2006	Initial publication