The American Statistical Association (ASA) is pleased to respond to OMB's request for comments on the 2020 Census information collection request (ICR), as invited in the *Federal Register* of February 13, 2019 (84 FR 3748).

The ASA is the nation's oldest scientific professional society with a proud tradition of supporting the Census Bureau. Indeed, the first modernization of the census in time for the 1850 decennial count was due in large part to the urging and work of the ASA. We are the world's largest community of statisticians. Our members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare.

After reviewing Supporting Statement Part A and Supporting Statement Part B of the 2020 Census ICR, we write to indicate that these documents are non-compliant with OMB Paperwork Reduction Act (PRA) guidelines, Information Quality Act guidelines, and Title 13, Section 6 of the U.S. Code.

Consequently, we request that OMB ask the Census Bureau to correct the multiple deficiencies in Supporting Statement Parts A and B so that OMB may more thoroughly and accurately evaluate the ICR.

In summary, the American Statistical Association finds the following deficiencies, (sources and references for each are provided following this statement):

1) Data collection not authorized by law – Contrary to PRA guidelines, the Supporting Statement does not affirm and describe compliance of the 2020 Census ICR with all relevant legislation, particularly 13 USC 6(c). This law (attached) authorizes the Commerce Secretary to add new questions to surveys only when administrative records from other federal agencies are not available and suitable. However, the Census Bureau has failed to establish that administrative records from other federal agencies are not available and suitable. Indeed, Census experts have identified administrative data that is not only available and suitable, but also less costly and disruptive. For example, in the attached memo to Secretary Ross on January 19, 2018, Census Chief Scientist John Abowd indicated that obtaining administrative records on citizenship would result in a 2020 Census that was less costly and more accurate than would adding a citizenship question. Consequently, as Judges Furman

and Seeborg recently found, the inclusion of the citizenship question violates federal law.

2) Noncompliance with Information Quality Act guidelines – The Supporting Statement fails to: (a) note that the 2020 Census ICR concerns "influential" data; (b) describe how the Census Bureau's data collection will meet the higher transparency and reproducibility standards required for influential data; and (c) provide written assurance from the Census Bureau's Chief Information Officer that the 2020 Census ICR is in compliance with OMB Information Quality Act guidelines (attached) and the Census Bureau's Statistical Quality Standards (attached).

3) Failure to describe and address expected declines in data quality – As the Abowd memorandum, the attached Census Bureau working paper, and the expert comments provided in response to the 2018 Census Bureau Call for Comments—including ASA's—indicate, the inclusion of a new citizenship question is likely to lead to a decline in data quality. However, inconsistent with OMB PRA guidelines, the Supporting Statement Parts A and B do not admit to an expected decline in data quality nor identify how the Census Bureau plans to address this decline.

4) Failure to describe and address potential sources of error – The Abowd memorandum and the attached Census Bureau working paper describe the potential sources of error due to the inclusion of a new citizenship question. However, inconsistent with OMB PRA guidelines, the Supporting Statement Part B does not fully describe these potential sources of error nor how the Census Bureau plans to address them.

5) Failure to identify item non-response rates for citizenship question – The Abowd memorandum and the attached Census Bureau working paper indicate that the inclusion of a new citizenship question would lead to higher item non-response rate for that question. However, inconsistent with OMB PRA guidelines, the Supporting Statement Parts A and B do not identify expected item non-response nor how the Census Bureau plans to fully address this issue.

6) Failure to pretest – Contrary to OMB PRA guidelines and Census Bureau Statistical Quality Standards, the Census Bureau did not conduct a pretest of a new question on citizenship status before adding it to the final version of the questionnaire.

In consequence, we strongly encourage OMB to return the 2020 Census ICR to the Census Bureau for revision to address the above deficiencies so that implementation of the 2020 Census may be a success under your watch.

We also request that the Census Bureau add the Abowd memorandum and the Brown et al. working paper to the ICR's supplementary documents.

In addition to the procedural concerns above, we reiterate and append the concerns for the addition of the citizenship questions expressed in our August 2018 comments. Specifically, we

urge the question be removed because of the very strong potential the quality of the census will be undermined by its inclusion and because of the lack of scientific justification.[1]

We appreciate the opportunity to provide these comments and hope you find them of value.

Questions on this document can be directed to the ASA Director of Science Policy Steve Pierson, pierson@amstat.org.

[1] https://www.amstat.org/asa/files/pdfs/POL-2020CensusCallForComments.pdf

Deficiencies in 2020 Census ICR – Notes and Sources

1) Noncompliance with 13 USC 6

The Supporting Statement does not indicate that the Census Bureau has legal authority consistent with 13 USC 6 to ask a citizenship status question, particularly as the Census Bureau Chief Scientist determined that the use of administrative records was more appropriate than adding a new question.

John Abowd, "Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census," January 19, 2018.

- "We consider three alternatives in response to the request: (A) no change in data collection, (B) adding a citizenship question to the 2020 Census, and (C) obtaining citizenship status from administrative records for the whole 2020 Census population.
- We recommend either Alternative A or C. Alternative C best meets DoJ's stated uses, is comparatively far less costly than Alternative B, does not increase response burden, and does not harm the quality of the census count. Alternative A is not very costly and also does not harm the quality of the census count.
- Alternative B better addresses DoJ's stated uses than Alternative A. However, ***Alternative B is very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available from administrative sources.*** (emphasis added)

13 USC 6 - Information from other Federal departments and agencies; acquisition of reports from other governmental and private sources

a) The Secretary, whenever he considers it advisable, may call upon any other department, agency, or establishment of the Federal Government, or of the government of the District of Columbia, for information pertinent to the work provided for in this title.
b) The Secretary may acquire, by purchase or otherwise, from States, counties, cities, or other units of government, or their instrumentalities, or from private persons and agencies, such copies of records, reports, and other material as may be required for the efficient and economical conduct of the censuses and surveys provided for in this title.
c) ***To the maximum extent possible and consistent with the kind, timeliness, quality and scope of the statistics required, the Secretary shall acquire and use information available from any source referred to in subsection (a) or (b) of this section instead of conducting direct inquiries.*** (emphasis added)

2) Noncompliance with Information Quality Act

The Supporting Statement fails to:

- Note that the 2020 Census ICR concerns "influential" data, where we refer to this 2002 OMB definition of "influential": "the 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";
- Describe how the Census Bureau's 2020 Census data collection will meet the higher transparency and reproducibility standards required for influential data; and
- Provide assurance from the Census Bureau's Chief Information Officer that the 2020 Census ICR is in compliance with OMB Information Quality Act guidelines and the Census Bureau's Statistical Quality Standards.

OMB, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication," February 22, 2002

- If an agency is responsible for disseminating ***influential*** scientific, financial, or statistical information, agency guidelines shall include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties.
- "Influential", when used in the phrase "influential scientific, financial, or statistical information", means that the 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. Each agency is authorized to define "influential" in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.
- IV. Agency Reporting Requirements
    - Agencies must designate the Chief Information Officer or another official to be responsible for agency compliance with these guidelines.

Census Bureau, Information Quality Guidelines

- The Census Bureau considers its published statistical information to be "influential," and does not distinguish among the many data it releases annually those which are more or less influential. The original and supporting data it collects and develops to generate its statistical information products are covered by the same quality guidelines as the information it disseminates and are proprietary and confidential.

OMB PRA Guidelines

- 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 objectivity 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. (p. 9)

- ***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.*** (p. 14) (emphasis added)
- 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. (p. 65)

3) Failure to Describe and Address Decline in Data Quality

The Abowd memoranda and Brown et al. working paper indicate that the inclusion of a new citizenship question would lead to a decline in data quality. However, inconsistent with OMB PRA and IQA guidelines, the Supporting Statement Parts A and B do not admit to an expected decline in data quality nor how the Census Bureau plans to fully address this decline.

In "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census" (August 2018), the Census Bureau describes in substantial detail the nature of data quality issues that would result from the introduction of the citizenship question.

- "[A]dding a citizenship question to the 2020 Census would lead to lower self-response rates in households potentially containing noncitizens, resulting in more nonresponse follow-up (NRFU) fieldwork, more proxy responses, and ***a lower-quality population count***."[2] (emphasis added)

While the 2020 Census ICR does indicate an increase in non-response, it does not describe the paper's findings regarding decline in data quality nor how it will address this problem.

OMB PRA Guidelines

- Agencies are encouraged to carefully consider how they can use current and new methodological tools to maximize data quality and minimize nonresponse bias. (p. 60)

[2] J. David Brown, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi, "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census," CES 18-38, U.S. Census Bureau, August 2018, p. 54.

- OMB PRA guidelines indicate that high item non-response would result in data quality problems. (p. 17)

4) Failure to Describe and Address Potential Sources of Error

Abowd memoranda and the Brown et al. working paper describe the potential sources of error due to the inclusion of a new citizenship question. However, inconsistent with OMB PRA guidelines, the Supporting Statement Part B does not describe these potential sources of error.

OMB PRA Guidelines

- 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.
- 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.
- 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. (pp. 20-21)

5) Failure to Identify Item Non-Response Rates for Citizenship Question

The Abowd memoranda and Brown et al. working paper indicate that the inclusion of a new citizenship question would lead to higher item non-response rate for that question. However, inconsistent with OMB PRA guidelines, the Supporting Statement Parts A and B do not identify expected item non-response nor how the Census Bureau plans to fully address this issue.

OMB PRA Guidelines

- 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.
- 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. (pp. 67-68)

6) Failure to Pretest

The Census Bureau is not in compliance with the requirement to pretest a question on citizenship status, per OMB PRA guidelines and Census Bureau Statistical Quality Standards.

OMB PRA Guidelines

- "[A]sking 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.***" (emphasis added) (p. 41)
- "The plan for testing or the results from the testing should also be described in Part B of the ICR." (p. 40)

Census Bureau, Statistical Quality Standards

- Sub-Requirement A2-3.3 of the Statistical Quality Standards requires that "[d]ata collection instruments and supporting materials must be pretested with respondents to identify problems (e.g., problems related to content, order/context effects, skip instructions, formatting, navigation, and edits) and then refined, prior to implementation, based on the pretesting results."
- Under Sub-Requirements A2-3.3-1c and A2-3.3-1d, pretesting must be performed when "Review by cognitive experts reveals that adding pretested questions to an existing instrument may cause potential context effects" and when "An existing data collection instrument has substantive modifications (e.g., existing questions are revised or new questions added)."

Attachments

John Abowd, "Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census," January 19, 2018.

J. David Brown, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi, "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census," CES 18-38, U.S. Census Bureau, August 2018.

Census Bureau, "Statistical Quality Guidelines."

OMB, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication," *Federal Register*, February 22, 2002.

OMB, "Questions and Answers When Designing Surveys for Information Collections," October 2016.

ASA's comments in response to summer 2018 Call for Comments from the U.S. Census Bureau

The American Statistical Association Strongly Cautions against Addition of a Citizenship Question on the 2020 Census

August 2, 2018

The American Statistical Association (ASA) appreciates this opportunity to respond to the comment request from the U.S. Census Bureau on the 2020 census. As the oldest American scientific professional society we have a proud tradition of supporting the Census Bureau. Indeed, the first modernization of the census in time for the 1850 decennial count was due in large part to the urging and work of the ASA.[3]

The decennial census forms the backbone of the U.S. data infrastructure, supporting our economy, health, and well-being. Providing the foundation for evidence-based policymaking and data-driven decision-making, its impact on the public and private sectors is broad and profound. In these comments, we focus on the March 2018 addition of the citizenship question to the questionnaire, a decision made by Secretary of Commerce Wilbur Ross, apparently against the advice of career Census Bureau officials.[4]

We reiterate ASA President Lisa LaVange's message to Secretary Ross in January 2018,[5] "strongly caution[ing] against adding this or any other question at this late stage of the process unless there is convincing evidence that its impact will be minimal." We urge the question be removed because of the very strong potential the quality of the census will be undermined by its inclusion and because of the lack of scientific justification.

It is an established scientific standard that instrumentation in large-scale data collections should not be changed without adequate time to field test and evaluate the impact of the changes before the data collection takes place. The addition of a potentially sensitive question, such as this one about citizenship, is a change in instrumentation that has the potential to compromise the validity and integrity of results from the census.

---

[3] http://magazine.amstat.org/blog/2016/12/01/asa_census/

[4] http://thehill.com/homenews/administration/391527-census-bureau-official-warned-in-memo-against-adding-citizenship

[5] http://www.amstat.org/asa/files/pdfs/POL-CitzenshipQuestion.pdf

Adding a citizenship question at this late stage of the decennial census process would likely increase distrust or suspicion of the government among immigrants, many of whom are already anxious about government inquiries and activities. Indeed, the Census Bureau researchers reported in 2017 an increase "in respondents spontaneously expressing concerns to researchers and field staff about confidentiality and data access relating to immigration" and "observation of increased rates of unusual respondent behaviors during pretesting and production surveys (data falsification, item non-response, break-offs)."[6] The fact that adding a citizenship question goes beyond the constitutional intent of the decennial census could further exacerbate the undercount concerns for immigrant populations. For these reasons specific to the nature of the question, adding a citizenship question is very likely to undermine the census.

These concerns have been expressed by many prominent experts. Most recently, in his March 1 memo to Secretary Ross,[7] Census Chief Scientist and Associate Director for Research and Methodology John Abowd stated the addition of the citizenship question would have "negative cost and quality implications" and "would result in poorer quality citizenship data" than a much less costly and less burdensome alternative using administrative data. Further, six former census directors who served under Democrat and Republican administrations urged Secretary Ross not to add the question,[8] noting that it had not gone through the rigorous testing required for putting new questions on surveys and censuses. "It is highly risky to ask untested questions in the context of the complete 2020 Census design," they said, observing that the effect of adding such a question at such a late date "would put the accuracy of the enumeration and success of the census in all communities at grave risk."

Further, former Commerce Secretaries Penny Pritzker and Carlos Gutierrez expressed[9] their "grave concerns about the proposed addition of a citizenship question to the decennial census in 2020," saying the "question will put in jeopardy the accuracy of the data that the census collects, and increase costs." Acting Census Director Ron Jarmin wrote in a December 22, 2017 email to a DOJ official that the Census Bureau's preliminary analysis suggested the best way to obtain the desired DOJ data "would be through utilizing a linked file of administrative and survey data the Census Bureau already possesses," concluding this approach "would result in higher quality data produced at lower cost."[10]

More generally, adding a question at this late stage of the Census process does not allow time for adequate testing, particularly if the testing reveals substantial problems. The decennial census questionnaire development is a multi-year, highly technical process that goes through various rigorous quality and scientific checks of the questions and, as importantly if not more so, the accompanying materials, interviewer instructions, and the contact process. Much of this

---

[6] https://www2.census.gov/cac/nac/meetings/2017-11/Meyers-NAC-Confidentiality-Presentation.pdf
[7] http://www.osec.doc.gov/opog/FOIA/Documents/AR%20-%20FINAL%20FILED%20-%20ALL%20DOCS%20%5bCERTIFICATION-INDEX-DOCUMENTS%5d%206.8.18.pdf#page=1320
[8] https://www.washingtonpost.com/r/2010-2019/WashingtonPost/2018/03/27/Editorial-Opinion/Graphics/DOJ_census_ques_request_Former_Directors_ltr_to_Ross.pdf?tid=a_mcntx
[9] https://www.bloomberg.com/view/articles/2018-04-04/u-s-census-is-not-about-citizenship
[10] https://www.documentcloud.org/documents/4616787-December-22-2017-Email-From-Ron-Jarmin.html#document/p1/a441604

development and testing culminated in the full end-to-end dress rehearsal carried out in Providence, RI earlier this year.

Some have optimistically stated that the addition of the citizenship question for the 2020 census requires no testing. Indeed, Secretary Ross,[9] after noting prior decennial census surveys "consistently asked citizenship questions up until 1950" and the American Community Survey (ACS) has included a citizenship question since 2005, concludes, "[T]he citizenship question has been well tested." Not only does this conclusion fail to recognize the many changes that have occurred in the census-taking process over the past 70 years; it also neglects the important differences between the ACS and decennial census—including in the accompanying materials, interviewer instructions, and contact process.

Beyond the testing standards discussed above, we maintain considerable justification is required before adding any question to ensure the quality of the decennial census would not be undermined. The addition of the citizenship question seems to be on questionable scientific grounding. In his March memo,[11] Secretary Ross concluded adding a citizenship question to the census will provide the Department of Justice (DOJ) accurate block-level data, but his claim is undermined by his own words in the same memo that non-citizens responding to ACS questions on citizenship inaccurately mark "citizen" about 30 percent of the time. Because of this inaccuracy rate, Census Chief Scientist Abowd indicated in his March 1 memo to Secretary Ross that his decided-upon approach—"Alternative D"—would yield only a minor improvement over the administrative record approach—"Alternative C"—and the marginal improvement comes at large expense, at great risk to the quality of the decennial census, and with burdening 100% of respondents with an additional question.

We are also concerned by the lack of scientific justification provided by the Department of Justice for why it needs the additional data and how it would use the information. It is particularly perplexing that DOJ on two previous occasions, as part of the well-established decade-long process leading up to the decennial census, cited its satisfaction with the data provided them by the Census Bureau. In a June 25, 2014 letter[12] to Department of Commerce General Counsel Kelly Welsh, DOJ official Arthur Gary affirmed the questions asked in the American Community Survey (ACS) were relevant and did not cite any inadequacy of the ACS data for DOJ purposes. In 2016 correspondence to the Census Bureau, Gary stated that DOJ had no need to amend the current content and uses or to request new content in the ACS for the 2020 census. (Gary later wrote to request that the Census Bureau consider a new topic in the ACS relating to LGBT populations.[13])

We continue to urge the citizenship question not be included "unless there is convincing evidence its impact will be minimal."[3] In his memo announcing the decision however, Secretary Ross states no one "could document that the response rate would in fact decline materially" due to the addition of a citizenship question. While it is true we have little experience and testing on

---

[11] https://www.commerce.gov/sites/commerce.gov/files/2018-03-26_2.pdf

[12] http://www.osec.doc.gov/opog/FOIA/Documents/AR%20-%20FINAL%20FILED%20-%20ALL%20DOCS%20%5bCERTIFICATION-INDEX-DOCUMENTS%5d%206.8.18.pdf#page=290

[13] http://www.osec.doc.gov/opog/FOIA/Documents/AR%20-%20FINAL%20FILED%20-%20ALL%20DOCS%20%5bCERTIFICATION-INDEX-DOCUMENTS%5d%206.8.18.pdf#page=323

this specific change, Secretary Ross ignores the expert opinion of the broad scientific community involved with survey and questionnaire research, which includes government, industry, and academic scientists. As LaVange urged in January,[3] the burden of proof should be to show minimal impact of adding this question because of the risks of an undercount and the questionable justification for the question.

The ASA has strong confidence in the professionals at the U.S. Census Bureau, including in its most capable, talented, and strong leadership. Activities to modernize the Bureau and the census in this decade are commendable and will help to reduce respondent burden and save taxpayer dollars while setting the foundation for an effective decennial in 2020. Jeopardizing these accomplishments by adding a citizenship question at this late stage will likely result in a lower response rate, more non-participation, and ultimately a sharp increase in costs for non-response follow-up.

Questions on this document can be directed to the ASA Director of Science Policy Steve Pierson, pierson@amstat.org.

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

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 <u>are</u> 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 objectivity[6] 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 methdology 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.fcsm.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 methdological 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 competetive 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) competetive 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

- 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).

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.

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 questionnaires, 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 knows 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

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 <u>each</u> 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*.67*.47*.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

- 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 e**stablishment 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 sample 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 *(check one)*

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

For b-f, note Item A2 of Supporting Statement instructions

4. Type of review requested *(check one)*

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

6. Requested expiration date

a. ☐ Three years from the approval date b. ☐ Other: _____/_____

7. Title

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

9. Keywords

10. Abstract

11. Affected public *(Mark primary with "P" and all others with "X")*

a. __ Individuals or households
b. __ Business or other for-profit
c. __ Not-for-profit institutions
d. __ Farms
e. __ Federal Government
f. __ State, Local, or Tribal Government

12. Obligation to respond *(Mark primary with "P" and all others that apply with "X")*

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 *(in thousands of dollars)*

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 *(Mark primary with "P" and all others that apply with "X")*

a. __ Application for benefits
b. __ Program evaluation
c. __ General purpose statistics
d. __ Audit
e. __ Program planning or management
f. __ Research
g. __ Regulatory or compliance

16. Frequency of recordkeeping or reporting *(check all that apply)*

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 *(person who can best answer questions regarding the content of this submission)*

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 extend 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.

Cover Memorandum for Version Published January 2006

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.
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 professioanl 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

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

January 19, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census

The Department of Justice has requested block-level citizen voting-age population estimates by OMB-approved race and ethnicity categories from the 2020 Census of Population and Housing. These estimates are currently provided in two related data products: the PL94-171 redistricting data, produced by April 1st of the year following a decennial census under the authority of 13 U.S.C. Section 141, and the Citizen Voting Age Population by Race and Ethnicity (CVAP) tables produced every February from the most recent five-year American Community Survey data. The PL94-171 data are released at the census block level. The CVAP data are released at the census block group level.

We consider three alternatives in response to the request: (A) no change in data collection, (B) adding a citizenship question to the 2020 Census, and (C) obtaining citizenship status from administrative records for the whole 2020 Census population.

We recommend either Alternative A or C. Alternative C best meets DoJ's stated uses, is comparatively far less costly than Alternative B, does not increase response burden, and does not harm the quality of the census count. Alternative A is not very costly and also does not harm the quality of the census count. Alternative B better addresses DoJ's stated uses than Alternative A. However, Alternative B is very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available from administrative sources.

Summary of Alternatives			
	Alternative A	***Alternative B***	***Alternative C***
Description	No change in data collection	Add citizenship question to the 2020 Census (i.e., the DoJ request), all 2020 Census microdata remain within the Census Bureau	Leave 2020 Census questionnaire as designed and add citizenship from administrative records, all 2020 Census microdata and any linked citizenship data remain within the Census Bureau
Impact on 2020 Census	None	Major potential quality and cost disruptions	None
Quality of Citizen Voting-Age Population Data	Status quo	Block-level data improved, but with serious quality issues remaining	Best option for block-level citizenship data, quality much improved
Other Advantages	Lowest cost alternative	Direct measure of self-reported citizenship for the whole population	Administrative citizenship records more accurate than self-reports, incremental cost is very likely to be less than $2M, USCIS data would permit record linkage for many more legal resident noncitizens
Shortcomings	Citizen voting-age population data remain the same or are improved by using small-area modeling methods	Citizenship status is misreported at a very high rate for noncitizens, citizenship status is missing at a high rate for citizens and noncitizens due to reduced self-response and increased item nonresponse, nonresponse followup costs increase by at least $27.5M, erroneous enumerations increase, whole-person census imputations increase	Citizenship variable integrated into 2020 Census microdata outside the production system, Memorandum of Understanding with United States Citizen and Immigration Services required to acquire most up-to-date naturalization data

Approved: _____ Date: _____

John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

Detailed Analysis of Alternatives

The statistics in this memorandum have been released by the Census Bureau Disclosure Review Board with approval number CBDRB-2018-CDAR-014.

Alternative A: Make no changes

Under this alternative, we would not change the current 2020 Census questionnaire nor the planned publications from the 2020 Census and the American Community Survey (ACS). Under this alternative, the PL94-171 redistricting data and the citizen voting-age population (CVAP) data would be released on the current schedule and with the current specifications. The redistricting and CVAP data are used by the Department of Justice to enforce the Voting Rights Act. They are also used by state redistricting offices to draw congressional and legislative districts that conform to constitutional equal-population and Voting Rights Act nondiscrimination requirements. Because the block-group-level CVAP tables have associated margins of error, their use in combination with the much more precise block-level census counts in the redistricting data requires sophisticated modeling. For these purposes, most analysts and the DoJ use statistical modeling methods to produce the block-level eligible voter data that become one of the inputs to their processes.

If the DoJ requests the assistance of Census Bureau statistical experts in developing model-based statistical methods to better facilitate the DoJ's uses of these data in performing its Voting Rights Act duties, a small team of Census Bureau experts similar in size and capabilities to the teams used to provide the Voting Rights Act Section 203 language determinations would be deployed.

We estimate that this alternative would have no impact on the quality of the 2020 Census because there would be no change to any of the parameters underling the Secretary's revised life-cycle cost estimates. The estimated cost is about $350,000 because that is approximately the cost of resources that would be used to do the modeling for the DoJ.

Alternative B: Add the question on citizenship to the 2020 Census questionnaire

Under this alternative, we would add the ACS question on citizenship to the 2020 Census questionnaire and ISR instrument. We would then produce the block-level citizen voting-age population by race and ethnicity tables during the 2020 Census publication phase.

Since the question is already asked on the American Community Survey, we would accept the cognitive research and questionnaire testing from the ACS instead of independently retesting the citizenship question. This means that the cost of preparing the new question would be minimal. We did not prepare an estimate of the impact of adding the citizenship question on the cost of reprogramming the Internet Self-Response (ISR) instrument, revising the Census Questionnaire Assistance (CQA), or redesigning the printed questionnaire because those components will not be finalized until after the March 2018 submission of the final questions. Adding the citizenship question is similar in scope and cost to recasting the race and ethnicity questions again, should that become necessary, and would be done at the same time. After the 2020 Census ISR, CQA and printed questionnaire are in final form, adding the citizenship question would be much more expensive and would depend on exactly when the implementation decision was made during the production cycle.

For these reasons, we analyzed Alternative B in terms of its adverse impact on the rate of voluntary cooperation via self-response, the resulting increase in nonresponse followup (NRFU), and the consequent effects on the quality of the self-reported citizenship data. Three distinct analyses support the conclusion of an adverse impact on self-response and, as a result, on the accuracy and quality of the 2020 Census. We assess the costs of increased NRFU in light of the results of these analyses.

B.1. Quality of citizenship responses

We considered the quality of the citizenship responses on the ACS. In this analysis we estimated item nonresponse rates for the citizenship question on the ACS from 2013 through 2016. When item nonresponse occurs, the ACS edit and imputation modules are used to allocate an answer to replace the missing data item. This results in lower quality data because of the statistical errors in these allocation models. The analysis of the self-responses responses is done using ACS data from 2013-2016 because of operational changes in 2013, including the introduction of the ISR option and changes in the followup operations for mail-in questionnaires.

In the period from 2013 to 2016, item nonresponse rates for the citizenship question on the mail-in questionnaires for non-Hispanic whites (NHW) ranged from 6.0% to 6.3%, non-Hispanic blacks (NHB) ranged from 12.0% to 12.6%, and Hispanics ranged from 11.6 to 12.3%. In that same period, the ISR item nonresponse rates for citizenship were greater than those for mail-in questionnaires. In 2013, the item nonresponse rates for the citizenship variable on the ISR instrument were NHW: 6.2%, NHB: 12.3% and Hispanic: 13.0%. By 2016 the rates increased for NHB and especially Hispanics. They were NHW: 6.2%, NHB: 13.1%, and Hispanic: 15.5% (a 2.5 percentage point increase). Whether the response is by mail-in questionnaire or ISR instrument, item nonresponse rates for the citizenship question are much greater than the comparable rates for other demographic variables like sex, birthdate/age, and race/ethnicity (data not shown).

B.2. Self-response rate analyses

We directly compared the self-response rate in the 2000 Census for the short and long forms, separately for citizen and noncitizen households. In all cases, citizenship status of the individuals in the household was determined from administrative record sources, not from the response on the long form. A noncitizen household contains at least one noncitizen. Both citizen and noncitizen households have lower self-response rates on the long form compared to the short form; however, the decline in self-response for noncitizen households was 3.3 percentage points greater than the decline for citizen households. This analysis compared short and long form respondents, categories which were randomly assigned in the design of the 2000 Census.

We compared the self-response rates for the same household address on the 2010 Census and the 2010 American Community Survey, separately for citizen and noncitizen households. Again, all citizenship data were taken from administrative records, not the ACS, and noncitizen households contain at least one noncitizen resident. In this case, the randomization is over the selection of household addresses to receive the 2010 ACS. Because the ACS is an ongoing survey sampling fresh households each month, many of the residents of sampled households completed the 2010 ACS with the same reference address as they used for the 2010 Census. Once again, the self-response rates were lower in the ACS than in the 2010 Census for both citizen and noncitizen households. In this 2010 comparison, moreover, the decline in self-response was 5.1 percentage points greater for noncitizen households than for citizen households.

In both the 2000 and 2010 analyses, only the long-form or ACS questionnaire contained a citizenship question. Both the long form and the ACS questionnaires are more burdensome than the shortform. Survey methodologists consider burden to include both the direct time costs of responding and the indirect costs arising from nonresponse due to perceived sensitivity of the topic. There are, consequently, many explanations for the lower self-response rates among all household types on these longer questionnaires. However, the only difference between citizen and noncitizen households in our studies was the presence of at least one noncitizen in noncitizen households. It is therefore a reasonable inference that a question on citizenship would lead to some decline in overall self-response because it would make the 2020 Census modestly more burdensome in the direct sense, and potentially much more burdensome in the indirect sense that it would lead to a larger decline in self-response for noncitizen households.

B.3. Breakoff rate analysis

We examined the response breakoff paradata for the 2016 ACS. We looked at all breakoff screens on the ISR instrument, and specifically at the breakoffs that occurred on the screens with the citizenship and related questions like place of birth and year of entry to the U.S. Breakoff paradata isolate the point in answering the questionnaire where a respondent discontinues entering data—breaks off—rather than finishing. A breakoff is different from failure to self-respond. The respondent started the survey and was prepared to provide the data on the Internet Self-Response instrument, but changed his or her mind during the interview.

Hispanics and non-Hispanic non-whites (NHNW) have greater breakoff rates than non-Hispanic whites (NHW). In the 2016 ACS data, breakoffs were NHW: 9.5% of cases while NHNW: 14.1% and Hispanics: 17.6%. The paradata show the question on which the breakoff occurred. Only 0.04% of NHW broke off on the citizenship question, whereas NHNW broke off 0.27% and Hispanics broke off 0.36%. There are three related questions on immigrant status on the ACS: citizenship, place of birth, and year of entry to the United States. Considering all three questions Hispanics broke off on 1.6% of all ISR cases, NHNW: 1.2% and NHW: 0.5%. A breakoff on the ISR instrument can result in follow-up costs, imputation of missing data, or both. Because Hispanics and non-Hispanic non-whites breakoff much more often than non-Hispanic whites, especially on the citizenship-related questions, their survey response quality is differentially affected.

B.4. Cost analysis

Lower self-response rates would raise the cost of conducting the 2020 Census. We discuss those increased costs below. They also reduce the quality of the resulting data. Lower self-response rates degrade data quality because data obtained from NRFU have greater erroneous enumeration and whole-person imputation rates. An erroneous enumeration means a census person enumeration that should not have been counted for any of several reasons, such as, that the person (1) is a duplicate of a correct enumeration; (2) is inappropriate (e.g., the person died before Census Day); or (3) is enumerated in the wrong location for the relevant tabulation (https://www.census.gov/coverage_measurement/definitions/). A whole-person census imputation is a census microdata record for a person for which all characteristics are imputed.

Our analysis of the 2010 Census coverage errors (Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States, Memo G-01) contains the relevant data. That study found that when the 2010 Census obtained a valid self-response (219 million persons),

the correct enumeration rate was 97.3%, erroneous enumerations were 2.5%, and whole-person census imputations were 0.3%. All erroneous enumeration and whole-person imputation rates are much greater for responses collected in NRFU. The vast majority of NRFU responses to the 2010 Census (59 million persons) were collected in May. During that month, the rate of correct enumerations was only 90.2%, the rate of incorrect enumeration was 4.8%, and the rate of whole-person census imputations was 5.0%. June NRFU accounted for 15 million persons, of whom only 84.6% were correctly enumerated, with erroneous enumerations of 5.7%, and whole-person census imputations of 9.6%. (See Table 19 of 2010 Census Memorandum G-01. That table does not provide statistics for all NRFU cases in aggregate.)

One reason that the erroneous enumeration and whole-person imputation rates are so much greater during NRFU is that the data are much more likely to be collected from a proxy rather than a household member, and, when they do come from a household member, that person has less accurate information than self-responders. The correct enumeration rate for NRFU household member interviews is 93.4% (see Table 21 of 2010 Census Memorandum G-01), compared to 97.3% for non-NRFU households (see Table 19). The information for 21.0% of the persons whose data were collected during NRFU is based on proxy responses. For these 16 million persons, the correct enumeration rate is only 70.1%. Among proxy responses, erroneous enumerations are 6.7% and whole-person census imputations are 23.1% (see Table 21).

Using these data, we can develop a cautious estimate of the data quality consequences of adding the citizenship question. We assume that citizens are unaffected by the change and that an additional 5.1% of households with at least one noncitizen go into NRFU because they do not self-respond. We expect about 126 million occupied households in the 2020 Census. From the 2016 ACS, we estimate that 9.8% of all households contain at least one noncitizen. Combining these assumptions implies an additional 630,000 households in NRFU. If the NRFU data for those households have the same quality as the average NRFU data in the 2010 Census, then the result would be 139,000 fewer correct enumerations, of which 46,000 are additional erroneous enumerations and 93,000 are additional whole-person census imputations. This analysis assumes that, during the NRFU operations, a cooperative member of the household supplies data 79.0% of the time and 21.0% receive proxy responses. If all of these new NRFU cases go to proxy responses instead, the result would be 432,000 fewer correct enumerations, of which 67,000 are erroneous enumerations and 365,000 are whole-person census imputations.

For Alternative B, our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 630,000 households, or approximately 0.5 percentage points. We currently estimate that for each percentage point increase in NRFU, the cost of the 2020 Census increases by approximately $55 million. Accordingly, the addition of a question on citizenship could increase the cost of the 2020 Census by at least $27.5 million. It is worth stressing that this cost estimate is a lower bound. Our estimate of $55 million for each percentage point increase in NRFU is based on an average of three visits per household. We expect that many more of these noncitizen households would receive six NRFU visits.

We believe that $27.5 million is a conservative estimate because the other evidence cited in this report suggests that the differences between citizen and noncitizen response rates and data quality will be amplified during the 2020 Census compared to historical levels. Hence, the decrease in self-response for citizen households in 2020 could be much greater than the 5.1 percentage points we observed during the 2010 Census.

Alternative C: Use administrative data on citizenship instead of add the question to the 2020 Census

Under this alternative, we would add the capability to link an accurate, edited citizenship variable from administrative records to the final 2020 Census microdata files. We would then produce block-level tables of citizen voting age population by race and ethnicity during the publication phase of the 2020 Census using the enhanced 2020 Census microdata.

The Census Bureau has conducted tests of its ability to link administrative data to supplement the decennial census and the ACS since the 1990s. Administrative record studies were performed for the 1990, 2000 and 2010 Censuses. We discuss some of the implications of the 2010 study below. We have used administrative data extensively in the production of the economic censuses for decades. Administrative business data from multiple sources are a key component of the production Business Register, which provides the frames for the economic censuses, annual, quarterly, and monthly business surveys. Administrative business data are also directly tabulated in many of our products.

In support of the 2020 Census, we moved the administrative data linking facility for households and individuals from research to production. This means that the ability to integrate administrative data at the record level is already part of the 2020 Census production environment. In addition, we began regularly ingesting and loading administrative data from the Social Security Administration, Internal Revenue Service and other federal and state sources into the 2020 Census data systems. In assessing the expected quality and cost of Alternative C, we assume the availability of these record linkage systems and the associated administrative data during the 2020 Census production cycle.

C.1. Quality of administrate record versus self-report citizenship status

We performed a detailed study of the responses to the citizenship question compared to the administrative record citizenship variable for the 2000 Census, 2010 ACS and 2016 ACS. These analyses confirm that the vast majority of citizens, as determined by reliable federal administrative records that require proof of citizenship, correctly report their status when asked a survey question. These analyses also demonstrate that when the administrative record source indicates an individual is not a citizen, the self-report is "citizen" for no less than 23.8% of the cases, and often more than 30%.

For all of these analyses, we linked the Census Bureau's enhanced version of the SSA Numident data using the production individual record linkage system to append an administrative citizenship variable to the relevant census and ACS microdata. The Numident data contain information on every person who has ever been issued a Social Security Number or an Individual Taxpayer Identification Number. Since 1972, SSA has required proof of citizenship or legal resident alien status from applicants. We use this verified citizenship status as our administrative citizenship variable. Because noncitizens must interact with SSA if they become naturalized citizens, these data reflect current citizenship status albeit with a lag for some noncitizens.

For our analysis of the 2000 Census long-form data, we linked the 2002 version of the Census Numident data, which is the version closest to the April 1, 2000 Census date. For 92.3% of the 2000 Census long-form respondents, we successfully linked the administrative citizenship variable. The 7.7% of persons for whom the administrative data are missing is comparable to the item non-response for self-responders in the mail-in pre-ISR-option ACS. When the administrative data indicated that the 2000 Census respondent was a citizen, the self-response was citizen: 98.8%. For this same group, the long-form response was

noncitizen: 0.9% and missing: 0.3%. By contrast, when the administrative data indicated that the respondent was not a citizen, the self-report was citizen: 29.9%, noncitizen: 66.4%, and missing: 3.7%.

In the same analysis of 2000 Census data, we consider three categories of individuals: the reference person (the individual who completed the census form for the household), relatives of the reference person, and individuals unrelated to the reference person. When the administrative data show that the individual is a citizen, the reference person, relatives of the reference person, and nonrelatives of the reference person have self-reported citizenship status of 98.7%, 98.9% and 97.2%, respectively. On the other hand, when the administrative data report that the individual was a noncitizen, the long-form response was citizen for 32.9% of the reference persons; that is, reference persons who are not citizens according to the administrative data self-report that they are not citizens in only 63.3% of the long-form responses. When they are reporting for a relative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 28.6% of the long-form responses. When they are reporting for a nonrelative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 20.4% of the long-form responses.

We analyzed the 2010 and 2016 ACS citizenship responses using the same methodology. The 2010 ACS respondents were linked to the 2010 version of the Census Numident. The 2016 ACS respondents were linked to the 2016 Census Numident. In 2010, 8.5% of the respondents could not be linked, or had missing citizenship status on the administrative data. In 2016, 10.9% could not be linked or had missing administrative data. We reached the same conclusions using 2010 and 2016 ACS data with the following exceptions. When the administrative data report that the individual is a citizen, the self-response is citizen on 96.9% of the 2010 ACS questionnaires and 93.8% of the 2016 questionnaires. These lower self-reported citizenship rates are due to missing responses on the ACS, not misclassification. As we noted above, the item nonresponse rate for the citizenship question has been increasing. These item nonresponse data show that some citizens are not reporting their status on the ACS at all. In 2010 and 2016, individuals for whom the administrative data indicate noncitizen respond citizen in 32.7% and 34.7% of the ACS questionnaires, respectively. The rates of missing ACS citizenship response are also greater for individuals who are noncitizens in the administrative data (2010: 4.1%, 2016: 7.7%). The analysis of reference persons, relatives, and nonrelatives is qualitatively identical to the 2000 Census analysis.

In all three analyses, the results for racial and ethnic groups and for voting age individuals are similar to the results for the whole population with one important exception. If the administrative data indicate that the person is a citizen, the self-report is citizen at a very high rate with the remainder being predominately missing self-reports for all groups. If the administrative data indicate noncitizen, the self-report is citizen at a very high rate (never less than 23.8% for any racial, ethnic or voting age group in any year we studied). The exception is the missing data rate for Hispanics, who are missing administrative data about twice as often as non-Hispanic blacks and three times as often as non-Hispanic whites.

C.2. Analysis of coverage differences between administrative and survey citizenship data

Our analysis suggests that the ACS and 2000 long form survey data have more complete coverage of citizenship than administrative record data, but the relative advantage of the survey data is diminishing. Citizenship status is missing for 10.9 percent of persons in the 2016 administrative records, and it is missing for 6.3 percent of persons in the 2016 ACS. This 4.6 percentage point gap between administrative and survey missing data rates is smaller than the gap in 2000 (6.9 percentage points) and 2010 (5.6

percentage points). Incomplete (through November) pre-production ACS data indicate that citizenship item nonresponse has again increased in 2017.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than administrative records, albeit less so now than in 2000. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the predicted answers of the sampled non-respondents are statistically the same as those of respondents. Our analysis casts serious doubt on this assumption, suggesting that those who do not respond to either the entire ACS or the citizenship question on the ACS are not statistically similar to those who do; in particular, their responses to the citizenship question would not be well-predicted by the answers of those who did respond.

The consequences of missing citizenship data in the administrative records are asymmetric. In the Census Numident, citizenship data may be missing for older citizens who obtained SSNs before the 1972 requirement to verify citizenship, naturalized citizens who have not confirmed their naturalization to SSA, and noncitizens who do not have an SSN or ITIN. All three of these shortcomings are addressed by adding data from the United States Citizen and Immigration Services (USCIS). Those data would complement the Census Numident data for older citizens and update those data for naturalized citizens. A less obvious, but equally important benefit, is that they would permit record linkage for legal resident aliens by allowing the construction of a supplementary record linkage master list for such people, who are only in scope for the Numident if they apply for and receive an SSN or ITIN. Consequently, the administrative records citizenship data would most likely have both more accurate citizen status and fewer missing individuals than would be the case for any survey-based collection method. Finally, having two sources of administrative citizenship data permits a detailed verification of the accuracy of those sources as well.

C.3. Cost of administrative record data production

For Alternative C, we estimate that the incremental cost, except for new MOUs, is $450,000. This cost estimate includes the time to develop an MOU with USCIS, estimated ingestion and curation costs for USCIS data, incremental costs of other administrative data already in use in the 2020 Census but for which continued acquisition is now a requirement, and staff time to do the required statistical work for integration of the administrative-data citizenship status onto the 2020 Census microdata. This cost estimate is necessarily incomplete because we have not had adequate time to develop a draft MOU with USCIS, which is a requirement for getting a firm delivery cost estimate from the agency. Acquisition costs for other administrative data acquired or proposed for the 2020 Census varied from zero to $1.5M. Thus the realistic range of cost estimates, including the cost of USCIS data, is between $500,000 and $2.0M

Questions on the Jan 19 Draft Census Memo on the DoJ Citizenship Question Reinstatement Request

1. **With respect to Alternatives B and C, what is the difference, if any, between the time when the data collected under each alternative would be available to the public?**

   Since the collection of this data, whether from administrative records or from an enumerated question, occurs prior to the creation of the Microdata Detail File (MDF) from which all tabulations will be performed, there is no difference in the timing of when the data collected under either alternative B or C could be made available to the public. The exact date for completion of the MDF is still being determined as the 2020 Census schedule is matured. However, the 2020 Census is working towards publishing the first post-apportionment tabulation data products as early as the first week of February 2021.

2. **What is the "2020 Census publication phase" (page 1 of the Detailed Analysis for Alternative B) versus Alternative C? Would there be any difference?**

   The 2020 Census publication phase is a broad window stretching from the release of the apportionment counts by December 31, 2020 through the last data product or report published in FY 2023, the final year of decennial funding for the 2020 Census. However, as stated in the answer to question 1, these data could be made available to the public on the same schedule as any other post-apportionment tabulated data product regardless of whether alternative B or C is used in its collection.

3. **What is the non-response rate for: (A) each question on the 2000 and 2010 Decennial Census short form and (B) each question on the 2010 ACS and most recent ACS?**

   The table below shows the item non-response (INR) rate for each question on the 2000 and 2010 Decennial Census short form. This is the percentage of respondents who did not provide an answer to an item.

   Item Nonresponse Rates for 2000 and 2010 Short Form Person Questions

	Relationship	Sex	Age	Hispanic Origin	Race	Tenure
2010	1.5	1.5	3.5	3.9	3.3	4.5
2000	1.3	1.1	3.7	3.1	2.9	4.1

   Source: Rothhaas, Lestina and Hill (2012) Tables

   Notes and Soucre:
   Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, January 24, 2012.

From report:

The INR rate is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item (i.e., the respondent did not provide an answer to the item). For INR, missing values are included in the rates, but inconsistent responses (i.e., incompatible with other responses) are considered non-missing responses.

Online link to 2010 report that has 2000 information as well.
https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf

See attached spreadsheet for the item allocation rates by questions for the ACS for 2010, 2013, and 2016.

4. **What was the total survey response rate (i.e., percentage of complete questionnaires) for the 2000 long form and the 2000 short form? Of the incomplete long forms, what percentage left the citizenship question blank? Of the completed long forms, what percentage (if known) contained incorrect responses to the citizenship question?**

We do not have measures of total survey response rates from the 2000 long form and 2000 short form available at this time. The mail response rate in 2000 was 66.4 percent for short forms and 53.9 percent for long forms. No analysis that we were aware of was conducted on the incomplete long forms that left the citizenship question blank. The Census 2000 Content Reinterview Survey showed low inconsistency of the responses to the citizenship question. Only 1.8 percent of the respondents changed answers in the reinterview.

Source for 2000 mail response rates:
https://www.census.gov/pred/www/rpts/A.7.a.pdf

Source for 2000 Content Reinterview Survey. Page 32 source.
https://www.census.gov/pred/www/rpts/B.5FR_RI.PDF

5. **For the 2000 long and short forms, what was the percentage unanswered (left blank) for each question (i.e., what percentage of the responses for each question (sex, race, ethnicity, income, citizenship, etc.) were left blank)?**

For the 2000 shortform, the table in question 3a provides the percentage unanswered for each question.

For the 2000 longform, Griffin, Love and Obenski (2003) summarized the Census 2000 longform responses. Allocation rates for individual items in Census 2000 were computed, but because of the magnitude of these data, summary allocation measures were derived.

These rates summarize completeness across all data items for occupied units (households) and are the ratio of all population and housing items that had values allocated to the total number of population and housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. Fifty-four population items and 29 housing items are included in these summary measures. The analysis showed that 9.9 percent of the population question items and 12.5 percent of the housing unit question items required allocation. Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices, to impute missing values.

https://ww2.amstat.org/sections/srms/Proceedings/y2003/Files/JSM2003-000596.pdf

6. **What was the incorrect response rate for the citizenship question that was asked on the Long Form during the 2000 Decennial Census? Does the response rate on the 2000 Long Form differ from the incorrect response rate on the citizenship question for the ACS?**

   In the 2000 long form, 2.3 percent of persons have inconsistent answers, 89.4 percent have consistent answers, and 8.2 percent have missing citizenship data in the SSA Numident and/or the 2000 long form. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2000 long form, 2.6 percent have inconsistent answers and 97.4 percent have consistent answers.

   In the 2010 ACS, 3.1 percent of persons have inconsistent answers, 86.0 percent have consistent answers, and 10.8 percent have missing citizenship data in the SSA Numident and/or the 2010 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2010 ACS, 3.6 percent have inconsistent answers and 96.4 percent have consistent answers.

   In the 2016 ACS, 2.9 percent of persons have inconsistent answers, 81.2 percent have consistent answers, and 15.9 percent have missing citizenship data in the SSA Numident and/or the 2016 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2016 ACS, 3.5 percent have inconsistent answers and 96.5 percent have consistent answers.

   These ACS and 2000 Census long form rates are based on weighted data.

   This shows that inconsistent response rates are higher in the 2010 and 2016 ACS than in the 2000 long form.

7. **What is the incorrect response rate on other Decennial or ACS questions for which Census has administrative records available (for example, age, sex or income)?**

   Table 7a shows the agreement rates between the 2010 Census response and the SSA Numident for persons who could be linked and had nonmissing values, and Table 7b shows

the agreement rates between the 2010 ACS and the SSA Numident. Gender has low disagreement (0.4-0.5 percent), and white alone (0.9 percent), black alone (1.7-2 percent), and age (2.1 percent) also have low disagreement rates. Disagreement rates are greater for other races (e.g., 46.4-48.6 percent for American Indian or Alaska Native alone). Hispanic origin is not well measured in the Numident, because it contains a single race response, one of which is Hispanic.

Table 7a. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 Census Response	Percent Agreement with SSA Numident
Hispanic	54.2
Not Hispanic	99.7
White Alone	99.1
Black Alone	98.3
American Indian or Alaska Native Alone	51.4
Asian Alone	84.3
Native Hawaiian or Other Pacific Islander Alone	74.4
Some Other Race Alone	17.7
Age	97.9
Gender	99.4

Source: Rastogi, Sonya, and Amy O'Hara, 2012, "2010 Census Match Study," 2010 Census Planning Memoranda Series No. 247.

Table 7b. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 ACS Response	Percent Agreement with SSA Numident
White Alone	99.1
Black Alone	98.0
American Indian or Alaska Native Alone	53.6
Asian Alone	82.9
Native Hawaiian or Other Pacific Islander Alone	72.9
Some Other Race Alone	17.2
Age 0-2 Date of Birth	95.2
Age 3-17 Date of Birth	95.6
Age 18-24 Date of Birth	95.2
Age 25-44 Date of Birth	95.8
Age 45-64 Date of Birth	95.9
Age 65-74 Date of Birth	96.5
Age 75 and older Date of Birth	92.7
Male	99.5
Female	99.5

Source: Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon, 2014, "Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data," CARRA Working Paper #2014-14.

Abowd and Stinson (2013) find correlations of 0.75-0.89 between Survey of Income and Program Participation (SIPP) and SSA Detailed Earnings Record annual earnings between 1990-1999.[1]

8. How does the Census presently handle responses on the (A) Decennial Census and (B) the ACS when administrative records available to the Census confirm that the response on the Decennial Census or ACS is incorrect? Is the present Census approach to incorrect responses based on practice/policy or law (statute or regulation)?

We have always based the short form Decennial Census and the ACS on self-response, and while we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaire. This is a long established practice at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census. Title 13 of the U.S. Code allows the Census Bureau to use alternative data sources, like administrative records, for a variety of purposes, and we are using data in new ways in the 2020 Census. While this includes the use of administrative records data to fill in areas where a respondent does not provide an answer, we have not explored the possibility of checking or changing responses that a responding household has provided in response to the questionnaire.

9. Please explain the differences between the self-response rate analysis and the breakoff rate analysis. The range of breakoff rates between groups was far smaller than the range of self-response rates between groups.

Self-response means that a household responded to the survey by mailing back a questionnaire or by internet, and a sufficient number of core questions were answered so that an additional field interview was not required.

A breakoff occurs when an internet respondent stops answering questions prior to the end of the questionnaire. In most cases the respondent answers the core questions before breaking off, and additional fieldwork is not required. The breakoff rates are calculated separately by which question screen was the last one reached before the respondent stopped answering altogether.

The share of Hispanic respondents who broke off at some point before the end of the questionnaire (17.6 percent) is much higher than for non-Hispanic whites (9.5 percent).

[1] Abowd, John M., and Martha H. Stinson, 2013, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," Review of Economics and Statistics, Vol. 95(55), pp. 1451-1467.

Spreading the overall breakoff rates over 134 screens in the questionnaire works out to quite small rates per screen. It works out to an average breakoff rate of 0.131 percent per screen for Hispanics and 0.066 percent for non-Hispanic whites.

10. The NRFU numbers are comparatively small – approximately one additional household for NRFU per Census enumerator. Is this really a significant source of concern?

Yes, this is a significant concern. First, it gives rise to incremental NRFU cost of at least $27.5 million. This is a lower bound becaues it assumes the households that do not self-respond because we added a question on citizenship have the same follow-up costs as an average U.S. household. They won't because these households overwhelmingly contain at least one noncitzen, and that is one of our acknowledged hard-to-count subpopulations.

11. Given that the breakoff rate difference was approximately 1 percent, why did Census choose to use the 5.1 percent number for assessing the cost of Alternative B?

If a household breaks off an internet response at the citizenship, place of birth, or year of entry screens, this means it would have already responded to the core questions. This would not trigger follow-up fieldwork and thus would not involve additional fieldwork costs. In contrast, if a household does not mail back a questionnaire or give an internet response, fieldwork will be necessary and additional costs will be incurred. Thus, the 5.1 percent number for differential self-response is more appropriate for estimating the additional fieldwork cost of adding a citizenship question.

12. Alternative C states that Census would use administrative data from the Social Security Administration, Internal Revenue Service, and "other federal and state sources." What are the other sources?

In addition to continuing the acquisition of the Social Security Administration and Internal Revenue Service data, the Census Bureau is in discussion with the U.S. Citizen and Immigration Services (USCIS) staff to acquire additional citizenship data.

13. Is Census confident that administrative data will be able to be used to determine citizenship for all persons (e.g., not all citizens have social security numbers)?

We are confident that Alternative C is viable and that we have already ingested enough high-quality citizenship administrative data from SSA and IRS. The USCIS data are not required. They would, however, make the citizenship voting age tabulations better, but the administrative data we've got are very good and better than the data from the 2000 Census and current ACS. The type of activities required for Alternative C already occur daily and routinely at the Census Bureau. We have been doing this for business data products,

including the Economic Censuses, for decades. We designed the 2020 Census to use this technology too.

14. For Alternative C, the memo says, "we assume the availability of these record linkage systems and associated administrative data" – does Census already have in place access to this data or would this need to be negotiated? If negotiated, for which data sets specifically?

The Census Bureau has longstanding contractual relationships with the Social Security Administration and the Internal Revenue Service that authorize the use of data for this project. For new data acquired for this project (i.e., USCIS) we would estimate a six-month development period to put a data acquisition agreement in place. That agreement would also include terms specifying the authorized use of data for this project.

15. Are there any privacy issues / sensitive information prohibitions that might prevent other agencies from providing such data?

There are no new privacy or sensitivity issues associated with other agencies providing citizenship data. We have received such information in the past from USCIS. We are currently authorized to receive and use the data from SSA and IRS that are discussed in Alternative C.

16. How long would Census expect any negotiation for access to data take? How likely is it that negotiations would be successful? Are MOA's needed/required?

Current data available to the Census Bureau provide the quality and authority to use that are required to support this project. Additional information potentially available from USCIS would serve to supplement/validate those existing data. We are in early discussions with USCIS to develop a data acquisition agreement and at this time have no indications that this acquisition would not be successful.

17. What limitations would exist in working with other agencies like IRS, Homeland Security, etc. to share data?

The context for sharing of data for this project is for a one-way sharing of data from these agencies to the Census Bureau. Secure file transfer protocols are in-place to ingest these data into our Title 13 protected systems. For those data already in-place at the Census Bureau to support this project, provisions for sharing included in the interagency agreement restrict the Census Bureau from sharing person-level microdata outside the Census Bureau's Title 13 protections. Aggregates that have been processed through the Bureau's disclosure avoidance procedures can be released for public use.

18. If Alternative C is selected, what is Census's backup plan if the administrative data cannot be completely collected and utilized as proposed?

The backup plan is to use all of the administrative data that we currently have, which is the same set that the analyses of Alternative C used. We have verified that this use is consistent with the existing MOUs. We would then use estimation and modeling techniques similar to those used for the Small Area Income and Poverty Estimates (SAIPE) to impute missing citizenship status for those persons for whom we do not have administrative records. These models would also include estimates of naturalizations that occurred since the administrative data were ingested.

19. Does Census have any reason to believe that access to existing data sets would be curtailed if Alternative C is pursued?

No we do not believe that any access to existing data sets would be curtailed if we pursue Alternative C.

20. Has the proposed Alternative C approach ever been tried before on other data collection projects, or is this an experimental approach? If this has been done before, what was the result and what were lessons learned?

The approach in Alternative C has been routinely used in processing the economic censuses for several decades. The Bureau's Business Register was specifically redesigned for the 2002 Economic Census in order to enhance the ingestion and use of administrative records from the IRS and other sources. The data in these administrative records are used to substitute for direct responses in the economic censuses for the unsampled entities. They are also used as part of the review, edit, and imputation systems for economic censuses and surveys. On the household side, the approach in Alternative C was used extensively to build the residential characteristics for OnTheMap and OnTheMap for Emergency Management.

21. Is using sample data and administrative records sufficient for DOJ's request?

The 2020 Census data combined with Alternative C are sufficient to meet DoJ's request. We do not anticipate using any ACS data under Alternative C.

22. Under Alternative C, If Census is able to secure interagency agreements to provide needed data sets, do we know how long it would take to receive the data transmission from other agencies and the length of time to integrate all that data, or is that unknown?

With the exception of the USCIS data, the data used for this project are already integrated into the 2020 Census production schema. In mid-to late 2018, we plan to acquire the USCIS data and with those data and our existing data begin to develop models and business rules to select citizenship status from the composite of sources and attach that characteristic to

each U.S. person. We expect the development and refinement of this process to continue into 2019 and to be completed by third quarter calendar year 2019.

23. Cross referencing Census decennial responses with numerous governmental data sets stored in various databases with differing formats and storage qualities sounds like it could be complicated. Does Census have an algorithm in place to efficiently combine and cross reference such large quantities of data coming from many different sources? What cost is associated with Alternative C, and what technology/plan does Census have in place to execute?

Yes, the 2018 Census End-to-End test will be implementing processing steps to be able to match Census responses to administrative record information from numerous governmental data sets. The Census Bureau has in place the Person Identification Validation System to assign Protected Identification Keys to 2020 Census responses. The required technology for linking in the administrative records is therefore part of the 2020 Census technology. This incremental cost factored into the estimate for Alternative C is for integrating the citizenship variable specifically, since that variable is not currently part of the 2020 Census design. No changes are required to the production Person Identification Validation system to integrate the administrative citizenship data.

24. For section C-1 of the memo, when did Census do the analyses of the incorrect response rates for non-citizen answers to the long form and ACS citizenship question? Were any of the analyses published?

The comparisons of ACS, 2000 Decennial Census longform and SSA Numident citizenship were conducted in January 2018. This analysis has not been published.

25. Has Census corrected the incorrect responses it found when examining non-citizen responses? If not, why not?

In the American Community Survey (ACS), and the short form Decennial Census, we do not change self-reported answers. The Decennial Census and the ACS are based on self-response and we accept the responses provided by households as they are given. While we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaires. This is a long established process at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census.

26. Has the Department of Justice ever been made aware of inaccurate reporting of ACS data on citizenship, so that they may take this into consideration when using the data?

Not exactly. The Census Bureau is in close, regular contact with the Department of Justice (DOJ) regarding their data requirements. Our counterparts at DOJ have a solid understanding of survey methodology and the quality of survey data, and they are aware of the public documentation on sampling and accuracy surrounding the ACS. However, the specific rate of accuracy regarding responses to the ACS question on citizenship has never been discussed.

27. Why has the number of persons who cannot be linked increased from 2010 to 2016?

The linkage between the ACS and administrative data from the SSA Numident and IRS ITIN tax filings depends on two factors: (a) the quality of the personally identifiable information (PII) on the ACS response and (b) whether the ACS respondent is in the SSN/ITIN universe.

With respect to the quality of the PII on the ACS, there may be insufficient information on the ACS due to item nonresponse or proxy response for the person to allow a successful match using the production record linkage system. There may also be more than one record in the Numident or ITIN IRS tax filings that matches the person's PII. Finally, there may be a discrepancy between the PII provided to the ACS and the PII in the administrative records.

Alternatively, the person may not be in the Numident or ITIN IRS tax filing databases because they are out of the universe for those administrative systems. This happens when the person is a citizen without an SSN, or when the person is a noncitizen who has not obtained an SSN or ITIN.

Very few of the unlinked cases are due to insufficient PII in the ACS or multiple matches with administrative records. The vast majority of unlinked ACS persons have sufficient PII, but fail to match any administrative records sufficiently closely. This means that most of the nonmatches are because the ACS respondent is not in the administrative record universe.

The incidence of ACS persons with sufficient PII but no match with administrative records increased between 2010 and 2016. One contributing factor is that the number of persons linked to ITIN IRS tax filings in 2016 was only 39 percent as large as in 2010, suggesting that either fewer of the noncitizens in the 2016 ACS had ITINs, or more of them provided PII in the ACS that was inconsistent with their PII in IRS records.

28. Independent of this memo, what action does Census plan to take in response to the analyses showing that non-citizens have been incorrectly responding to the citizenship question?

The Census Bureau does not have plans to make any changes to procedures in the ACS. However, we will continue to conduct thorough evaluations and review of census and survey data. The ACS is focusing our research on the potential use of administrative records

in the survey. For instance, we are exploring whether we can use IRS data on income to reduce the burden of asking questions on income on the ACS. We are concentrating initially on questions that are high burden, e.g., questions that are difficult to answer or questions that are seen as intrusive.

29. Did Census make recommendations the last time a question was added?

Since the short form Decennial Census was established in 2010, the only requests for new questions we have received have been for the ACS. And, in fact, requests for questions prior to 2010 were usually related to the Decennial Census Long Form. We always work collaboratively with Federal agencies that request a new question or a change to a question. The first step is to review the data needs and the legal justification for the new question or requested changes. If, through this process, we determine that the request is justified, we work with the other agencies to test the question (cognitive testing and field testing). We also work collaboratively on the analysis of the results from the test which inform the final recommendation about whether or not to make changes or add the question.

30. Does not answering truthfully have a separate data standard than not participating at all?

We're not sure what you're asking here. Please clarify the question.

31. What was the process that was used in the past to get questions added to the decennial Census or do we have something similar where a precedent was established?

Because no new questions have been added to the Decennial Census (for nearly 20 years), the Census Bureau did not feed bound by past precedent when considering the Department of Justices' request. Rather, the Census Bureau is working with all relevant stakeholders to ensure that legal and regulatory requirements are filled and that questions will produce quality, useful information for the nation. As you are aware, that process is ongoing at your direction.

32. Has another agency ever requested that a question be asked of the entire population in order to get block or individual level data?

Not to our knowledge. However, it is worth pointing out that prior to 1980 the short form of the Decennial Census included more than just the 10 questions that have been on the short form since 1990.

33. Would Census linking of its internal data sets, with other data sets from places like IRS and Homeland Security, have an impact on participation as well (i.e., privacy concerns)?

The potential that concerns about the use of administrative records could have an impact on participation has always been a concern of ours, and it's a risk that we're managing on our risk register. We've worked closely with the privacy community throughout the decade, and we established a working group on our National Advisory Committee to explore this issue. We've also regularly briefed the Congress about our plans. At this stage in the decade there does not appear to be extensive concerns among the general public about our approach to using administrative records in the Nonresponse Operation or otherwise. We will continue to monitor this issue.

34. Would Alternative C require any legislation? If so, what is the estimated time frame for approval of such legislation?

No.

35. Census publications and old decennial surveys available on the Census website show that citizenship questions were frequently asked of the entire population in the past. Citizenship is also a question on the ACS. What was the justification provided for citizenship questions on the (A) short form, (B) long form, and (C) ACS?

In 1940, the Census Bureau introduced the use of a short form to collect basic characteristics from all respondents, and a long form to collect more detailed questions from only a sample of respondents. Prior to 1940, census questions were asked of everyone, though in some cases only for those with certain characteristics. For example, in 1870, a citizenship question was asked, but only for respondents who were male and over the age of 21.

Beginning in 2005, all the long-form questions – including a question on citizenship -- were moved to the ACS. 2010 was the first time we conducted a short-form only census. The citizenship question is included in the ACS to fulfill the data requirements of the Department of Justice, as well as many other agencies including the Equal Employment Opportunities Commission, the Department of Health and Human Services, and the Social Security Administration.

ACS Item Allocation Rates for United States: 2016, 2013, 2010

Title	2016	2013	2010
Overall housing allocation rate occupied and vacant housing units	4.9	5.6	5.2
Overall person allocation rate total population	9.5	8.4	5.8
Vacancy status vacant housing units	3.9	3.5	2.9
Tenure occupied housing units	1.2	1.3	1.2
Units in structure occupied and vacant housing units	1.5	1.5	1.5
Year moved in occupied housing units	3	3	3.4
Month moved in occupied housing units into which households move in the last two years	0.7	0.7	0.7
Year built occupied and vacant housing units	18.2	17.1	16.2
Lot size occupied and vacant single family and mobile homes	3.9	3.9	4.2
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4	4.2	4.4
Business on property occupied and vacant single family and mobile homes	**	2.4	3
Number of rooms occupied and vacant housing units	5	5.5	5.2
Number of bedrooms occupied and vacant housing units	5.5	4.6	4.3
Running water occupied and vacant housing units	2.4	2.1	2
Flush toilet occupied and vacant housing units	**	2.2	2
Bathtub or shower occupied and vacant housing units	2.6	2.2	2
Sink with a faucet occupied and vacant housing units	2.6	2.2	2
Stove or range occupied and vacant housing units	3.1	2.8	2.5
Refrigerator occupied and vacant housing units	3.2	2.9	2.7
Telephone occupied housing units	1.5	1.2	1.1
Number of vehicles occupied housing units	1.2	1.4	1.3
Heating fuel occupied housing units	3.4	3.4	3.3
Monthly electricity cost occupied housing units	8.1	8.2	7.3
Monthly gas cost occupied housing units	9.6	9.9	9.8

Yearly water and sewer cost occupied housing units	8.5	8.8	8.1
Yearly other fuel cost occupied housing units	7.3	8.3	10.6
Yearly food stamp recipiency - household occupied housing units	1.7	1.7	1.3
Yearly real estate taxes owner-occupied housing units	16.7	18.5	16.3
Yearly property insurance owner-occupied housing units	23.9	25.6	23.2
Mortgage status owner-occupied housing units	2.2	2.5	2.1
Monthly mortgage payment owner-occupied housing units with a mortgage	10.5	12.4	10.7
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	6.2	6.9	(X)
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	6.8	7.4	(X)
Second mortgage owner-occupied housing units	3.2	3.7	3.4
Home equity loan owner-occupied housing units	3.7	4.3	4.2
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	23.3	21.7	17.9
Property value owner-occupied housing units and vacant housing units for sale	11.6	12.9	12.3
Yearly mobile home costs occupied mobile homes and other units	21.7	21.5	19.9
Monthly condominium fee owner-occupied housing units	0.8	0.8	0.7
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	10.5	9.8	9.3
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2.1	2.1	2
Desktop/laptop/notebook computer occupied housing units	1.3	3.2	**
Handheld computer/smart mobile phone occupied housing units	**	3.3	**
Tablet or other portable wireless computer occupied housing units	1.6	**	**
Smartphone occupied housing units	1.6	**	**
Other computer occupied housing units	1.7	3.7	**
Household has internet access occupied housing units	3.3	4.4	**
Dial-up internet service occupied housing units with internet access	3.8	5.7	**
DSL internet service occupied housing units with internet access	**	5.7	**

Cable modem internet service occupied housing units with internet access	**	5.7	**
Fiber-optic internet service occupied housing units with internet access	**	5.7	**
Cellular data plan (formerly mobile broadband) occupied housing units with internet access	7.6	26.7	**
Satellite internet service occupied housing units with internet access	3.8	5.7	**
High speed internet service occupied housing units with internet accesss	3.8	**	**
Some other internet service occupied housing units with internet access	3.8	5.7	**
Race total population	1.5	1.6	1.5
Hispanic origin total population	1.8	2.1	1.8
Sex total population	0.1	0.1	0.1
Age total population	1.7	1.6	1.3
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	5.3	4.8	3
Married past 12 months total population 15 years and over, except those never married	6.9	6.6	4.7
Widowed past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Divorced past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Times married total population 15 years and over, except those never married	8.1	7.8	5.1
Year last married total population 15 years and over, except those never married	13.5	13.3	11.4
Place of birth total population	9.1	8.6	6.5
Citizenship total population	6	5.2	2.7
Year of naturalization total population naturalized citizens	22.5	22.5	16.6
Year of entry total population not born in US	14.8	13.2	10.3
Speaks another language at home total population 5 years and over	6.8	5.9	3.4
Language spoken total population 5 years and over who speak another language at home	8.3	7	5.7
English ability total population 5 years and over who speak another language at home	7.1	5.9	4
School enrollment total population 3 years and over	6.7	6	3.7
Grade level attending			

total population 3 years and over enrolled	10.2	8.9	6
Educational attainment total population 3 years and over	8.5	8	5.6
Field of degree total population 25 years and over with a bachelor's degree or higher	13.5	12.4	9.8
Mobility status total population 1 years and over	7.2	6.5	4
Migration state/foreign county total population 1 years and over movers	13.2	11.3	7.1
Migration county total population 1 years and over movers within US	14.6	12.5	8.3
Migration minor civil division total population 1 years and over movers within US	14.2	12.1	8.4
Migration place total population 1 years and over movers within US	15	12.9	8.8
Health insurance thru employer/union total population	10.7	9	6.2
Health insurance purchased directly total population	11.3	9.7	6.9
Health insurance through Medicare total population	9.5	8.1	5.2
Health insurance through Medicaid total population	12.2	10.5	7.9
Health insurance through TRICARE total population	12.5	10.8	8.1
Health insurance through VA total population	12.3	10.7	8.1
Health ins. thru Indian Health Service total population	12.8	11.1	8.5
Visual difficulty total population	7.1	6.1	3.4
Hearing difficulty total population	6.8	5.9	3.2
Physical difficulty total population 5 years and over	7.5	6.7	3.5
Difficulty remembering total population 5 years and over	7.5	6.7	3.5
Difficulty dressing total population 5 years and over	7.5	6.7	3.5
Difficulty going out total population 16 years and over	7.3	6.5	3.4
Grandchildren living in home noninstitutionalized population 30 years and over	1.1	1	0.9
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	17.7	15.7	12
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	17.2	16.1	14.9
Fertility status female total population 15-50	7.8	6.7	3.7

Veteran status total population 17 years and over	7.3	6.8	3.8
Periods of military service total population 17 years and over on active duty now or previously	9.7	9.3	6.3
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	6.8	6.6	3.9
Service-connected disability rating value total population 17 years and over with a service-connected disability	0.2	0.2	0.7
Employment status recode noninstitutionalized population 16 years and over	8.7	8.1	5.1
When last worked noninstitutionalized population 16 years and over	9.6	9.1	5.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	10.6	9.7	6.9
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	11.9	10.8	7.7
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	11.8	10.4	6.3
Place of work county noninstitutionalized population 16 years and over at work last week	12.5	11	7
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	3.6	3.3	2.1
Place of work place noninstitutionalized population 16 years and over at work last week	13.1	11.6	7.6
Transportation to work noninstitutionalized population 16 years and over at work last week	9.6	8.8	5.7
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	10.9	9.9	6.8
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	20.2	18.5	12.8
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	14.5	13.3	9.7
Class of worker total population 16 years and over who worked in the last 5 years	11.7	10.7	7.2
Industry total population 16 years and over who worked in the last 5 years	12.7	11.4	7.8
Occupation total population 16 years and over who worked in the last 5 years	13.4	11.8	8.1
Wages/salary income total population 15 years and over	19.1	19	16
Self-employment income			

total population 15 years and over	10.5	9.3	5.9
Interest, dividends, etc. income total population 15 years and over	15.2	12.6	8.8
Social security or railroad retirement total population 15 years and over	14.5	12.3	8.9
Supplemental security income total population 15 years and over	12.7	10.3	6.7
Public assistance total population 15 years and over	13.2	10.5	6.8
Retirement income total population 15 years and over	13.6	11.1	7.5
Other income total population 15 years and over	13.2	10.8	7.4
Some or all income allocated total population 15 years and over	28.4	25.3	22.4

Source: ACS 1-year data. See following links for more information:

https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/

https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html

Note:

** This item was not asked in this year.

Summary Analysis of the Key Differences Between Alternative C and Alternative D

This short note describes the Census Bureau's current assumptions about two alternatives to address the need for block level data on citizen voting age populations. The goal is to measure the citizenship status of all people enumerated in the 2020 Decennial Census. Both alternatives utilize administrative data on the citizenship status of individuals, however one option, Alternative D, proposes to also include the current American Community Survey (ACS) question on citizenship status on the 2020 Decennial Census short form.

In both alternatives described here, the methodology requires linking 2020 census response data and administrative records. However, as illustrated both alternatives would also need to assign/impute citizenship for a portion of the population. The Census Bureau will have to assign citizenship in cases of questionnaire non-response and item non-response. Additionally, it is important to note, that even when a self-response is available it is not always possible to link response data with administrative records data. Poor data quality (e.g., name and age) and nonresponse or incomplete 2020 Census responses mean that we will not have a direct measure of citizenship status for all residents enumerated in 2020. The Census Bureau will to need employ an imputation model for these cases.

One of the key differences between to the two alternatives described below is the number of cases requiring imputation. The other key difference is the impact of errors in the citizenship status reported on the 2020 Census.

In the most recent version of the 2020 Decennial Life Cycle Cost Estimate, the Census Bureau projects counting 330 million residents in 2020. Figure 1 summarizes how citizenship status will be measured under Alternative C that does not employ a citizenship question on the 2020 Census. Figure 2 summarizes how this will be done using both administrative records and a 2020 citizenship question under Alternative D.

Alternative C is a simplified process for assigning citizenship through direct linkage and modelling, without including the question on the 2020 Census. The Census Bureau will link the responses for the 330 million census records to administrative records that contain information on the citizenship status of individuals. The Census Bureau expects to successfully link and observe this status for approximately 295 million people. The Census Bureau would need to impute this status for approximately 35 million people under Alternative C whose 2020 responses cannot be linked to administrative data. Although the Census Bureau has fully developed and tested the imputation model, it has high confidence that an accurate model can be developed and deployed for this purpose. Further, we will most likely never possess a fully adequate truth deck to benchmark it to.

Measuring citizenship status is slightly more complex under Alternative D where all U.S. households will be given the opportunity to provide the citizenship status of each household member. Based on response data for the ACS citizenship and other response data research, we know that not all households that respond to the 2020 Census will answer this question, leaving the question blank or with otherwise invalid responses. Additionally, Alternative D, must also account for those households that do not respond at all or will have proxy responses. Due to these reasons, we estimate that we will get 2020 citizenship status responses for approximately 294.6 million people, a slightly higher estimate

than Alternative C. For the 35.4 million people without a 2020 citizenship response, the Census Bureau will employ the same methodology as in Alternative C, linking the 2020 Census responses to the administrative records. The Census Bureau estimates that it will be able to link these cases to administrative records where we observe citizenship status for approximately 21.5 million people. For the remaining 13.8 million will be imputed through a model as described above. Thus, there will be a need for imputing many cases across either alternative.

The Census Bureau will link the 294.6 million records from the 2020 Census with the administrative records. This will be done both for potential quality assurance purposes and to improve the quality of future modeling uses. Based on the current research from the ACS, the Census Bureau expects to successfully link approximately 272.5 million of these cases. Of these, 263 million will have citizenship statuses that agree across the 2020 response and administrative record. The Census Bureau estimates there will be 9.5 million cases where there is disagreement across the two sources. Historic Census Bureau practice is to use self-reported data in these situations. However, the Census Bureau now knows from linking ACS responses on citizenship to administrative data that nearly one third of noncitizens in the administrative data respond to the questionnaire indicating they are citizens, indicating that this practice should be revisited in the case of measuring citizenship. Finally, for those 22.2 million cases that do not link to administrative records (non-linkage occurs for the same data quality reasons discussed above), the Census Bureau will use the observed 2020 responses. Again, Census Bureau expect some quality issues with these responses. Namely, the Census Bureau estimates that just under 500 thousand noncitizens will respond as citizens.

The relative quality of Alternative C versus Alternative D will depend on the relative importance of the errors in administrative data, response data, and imputations. To be slightly more but not fully precise consider the following description of errors under both alternatives. First note that all possible measurement methods will have errors. Under Alternative C, there will be error in the administrative records, but we believe these to be relatively limited dues to the procedure following by SSA, USCIS and State. In both Alternative, the modeled cases will be subject to prediction error. Prediction error occur when the model returns the incorrect status of a case. As there are more models cases in Alternative C, prediction error will be a bigger issue there. Alternative D has an additional source or error, response error. This is where 2020 respondent give the incorrect status. Statisticians often hope these error are random and cancel out. However, we know from prior research that citizenship status responses are systematically biased for a subset of noncitizens. Response error is only an issue in alternative D. Unfortunately, the Census Bureau cannot quantify the relative magnitude of the errors across the alternatives at this time.

Figure 1

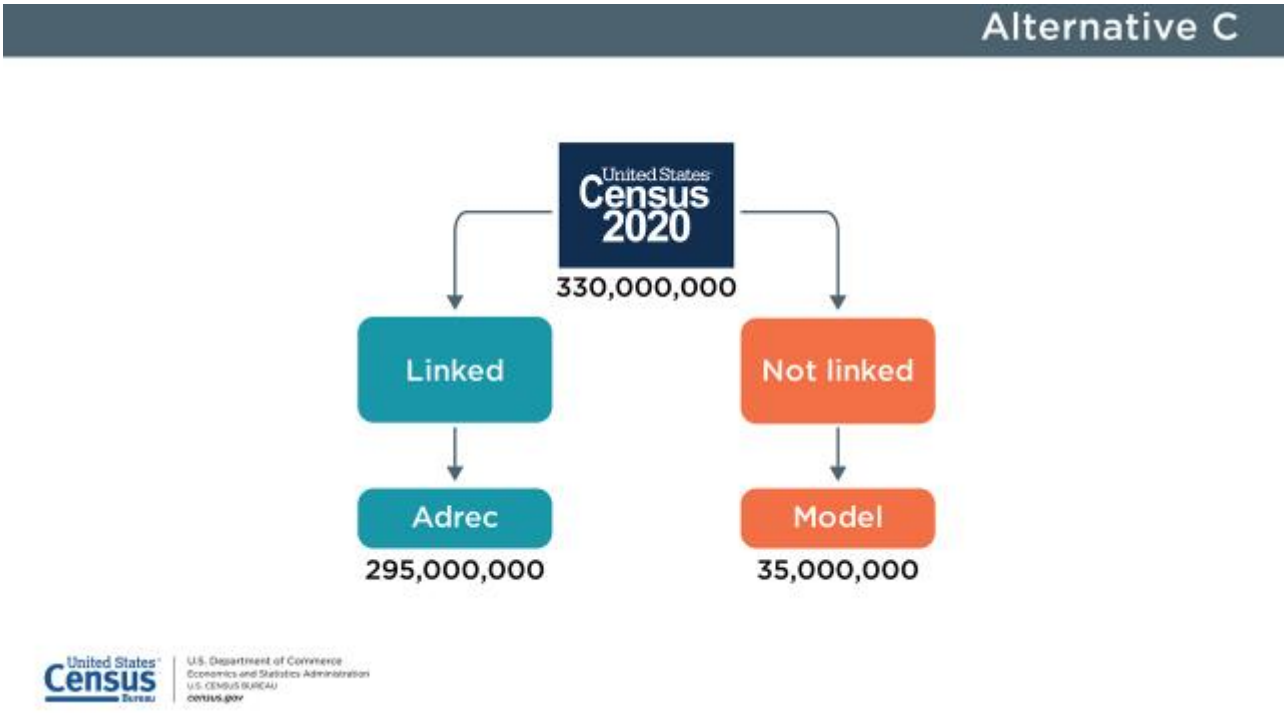
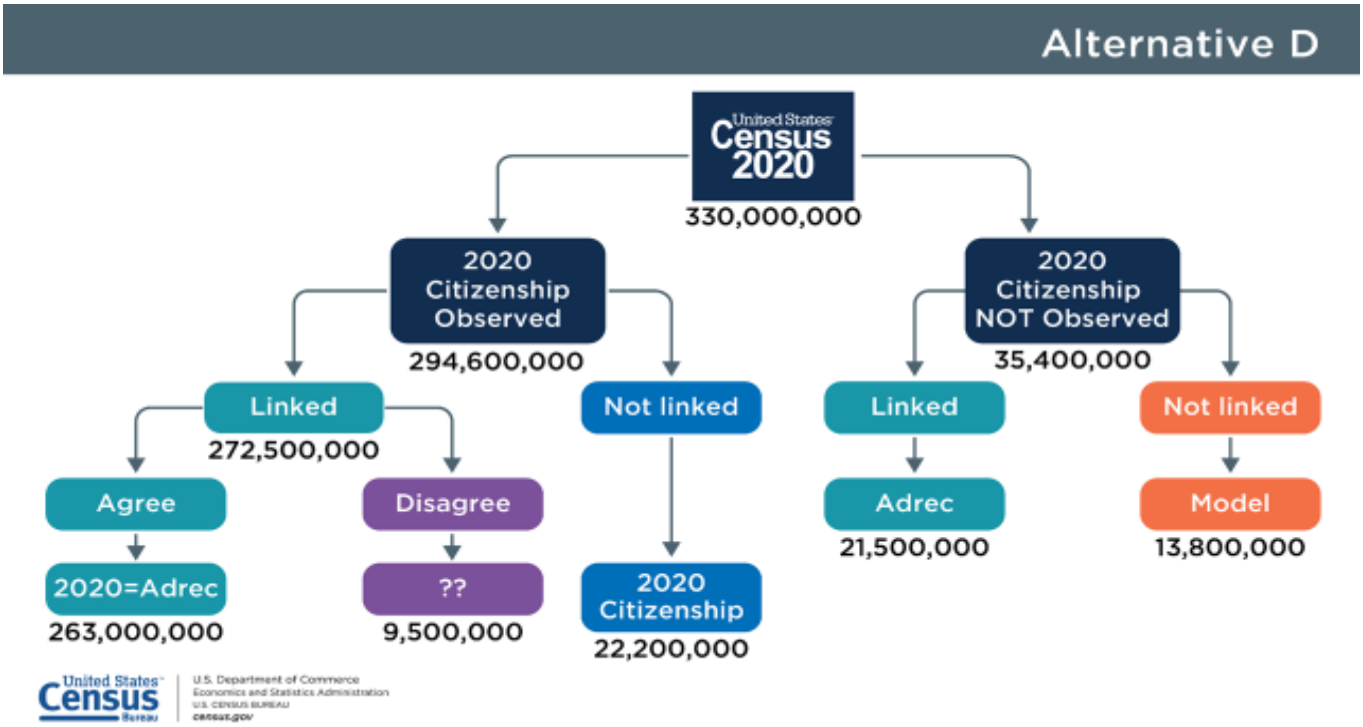

Alternative C
United States Census 2020
330,000,000
Linked
Not linked
Adrec
Model
295,000,000
35,000,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Figure 2


Alternative D
United States Census 2020
330,000,000
2020 Citizenship Observed
294,600,000
2020 Citizenship NOT Observed
35,400,000
Linked
272,500,000
Not linked
Linked
Not linked
Agree
Disagree
Adrec
21,500,000
Model
13,800,000
2020=Adrec
263,000,000
??
9,500,000
2020 Citizenship
22,200,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Reissued Jul 2013

U.S. Census Bureau
Statistical Quality Standards

The leading source of quality data about the nation's people and economy.

United States Census Bureau

U.S. CENSUS BUREAU

Thomas L. Mesenbourg, Jr.,
Deputy Director and Chief Operating Officer

Nancy A. Potok,
Associate Director for Demographic Programs

Arnold A. Jackson,
Associate Director for Decennial Census

Frank A. Vitrano,
Associate Director for 2020 Census

Ted A. Johnson,
Asssociate Director for Administration and Chief Financial Officer

Steven J. Jost,
Associate Director for Communications

Roderick Little,
Associate Director for Research and Methodology

David E. Hackbarth (Acting),
Associate Director for Field Operations

Brian E. McGrath,
Associate Director for Information Technology and Chief Information Officer

William G. Bostic, Jr.
Associate Director for Economic Programs

METHODOLOGY AND STANDARDS COUNCIL

Roderick Little,
Associate Director for Research and Methodology and Chair, Methodology and Standards Council

Ruth Ann Killion,
Chief, Demographic Statistical Methods Division

Xijian Liu,
Chief, Office of Statistical Methods and Research for Economic Programs

David Whitford,
Chief, Decennial Statistical Studies Division

Tommy Wright,
Chief, Center for Statistical Research and Methodology

TABLE OF CONTENTS

Preface[1]

1. Introduction

Purpose

This document specifies the statistical quality standards for the U.S. Census Bureau. As the largest statistical agency of the federal government, the Census Bureau strives to serve as the leading source of quality data about the nation's people and economy. The Census Bureau has developed these standards to promote quality in its information products and the processes that generate them. These standards provide a means to ensure consistency in the processes of all the Census Bureau's program areas, from planning through dissemination. By following these standards, the Census Bureau's employees and contractors will ensure the utility, objectivity, and integrity of the statistical information provided by the Census Bureau to Congress, to federal policy makers, to sponsors, and to the public.

Background

In 2002, the United States Office of Management and Budget (OMB) issued Information Quality Guidelines (OMB, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, February 22, 2002, 67 FR 8452-8460), directing all federal agencies to develop their own information quality guidelines. In October 2002, the Census Bureau issued its information quality guidelines (U.S. Census Bureau, *U.S. Census Bureau Section 515 Information Quality Guidelines*, 2002). These guidelines established a standard of quality for the Census Bureau and incorporated the information quality guidelines of the OMB and the Department of Commerce, the Census Bureau's parent agency.

Following the OMB's information quality guidelines, the Census Bureau defines information quality as an encompassing term comprising utility, objectivity, and integrity. Our definition of information quality is the foundation for these standards.

> *Utility* refers to the usefulness of the information for its intended users. We assess the usefulness of our information products from the perspective of policy makers, subject matter users, researchers, and the public. We achieve utility by continual assessment of customers' information needs, anticipation of emerging requirements, and development of new products and services.
>
> - The statistical quality standards related to utility include: Planning a Data Program (A1), Developing Data Collection Instruments and Supporting Materials (A2), Developing and Implementing a Sample Design (A3), Acquiring and Using Administrative Records (B2), Reviewing Information Products (E3), Releasing Information Products (F1), and Providing Documentation to Support Transparency in Information Products (F2).

[1] Please note that this document contains some Intranet links that are accessible only within the U.S. Census Bureau.

Objectivity focuses on whether information is accurate, reliable, and unbiased, and is presented in an accurate, clear, complete, and unbiased manner. Objectivity involves both the content of the information and the presentation of the information. It requires complete, accurate, and easily understood documentation of the sources of the information, with a description of the sources of errors that may affect the quality of the data, when appropriate.

- The statistical quality standards related to objectivity include: Developing Data Collection Instruments and Supporting Materials (A2), Developing and Implementing a Sample Design (A3), Establishing and Implementing Data Collection Methods (B1), Acquiring and Using Administrative Records (B2), Capturing Data (C1), Editing and Imputing Data (C2), Coding Data (C3), Linking Data from Multiple Sources (C4), Producing Direct Estimates from Samples (D1), Producing Estimates from Models (D2), Producing Measures and Indicators of Nonsampling Error (D3), Analyzing Data (E1), Reporting Results (E2), Reviewing Information Products (E3), Releasing Information Products (F1), Providing Documentation to Support Transparency in Information Products (F2), Addressing Information Quality Complaints (F3), and Managing Data and Documents (S2).

Integrity refers to the security of information – protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification. Several federal statutes and Census Bureau policies govern the protection of information, most notably Title 13 and Title 26.

- Protecting Confidentiality (S1) directly addresses issues concerning the integrity of the data. All the statistical quality standards contain requirements for protecting information from unauthorized access or release.

In September 2006, the OMB issued *Standards and Guidelines for Statistical Surveys*, which specify requirements for federal statistical agencies to ensure that their information products satisfy the information quality guidelines. The OMB standards are not intended to describe all the efforts that an agency may undertake to ensure the quality of its information. These Census Bureau statistical quality standards provide additional guidance that focuses on the Census Bureau's statistical programs and activities and that addresses the Census Bureau's unique methodological and operational issues.

2. Scope

The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

The Census Bureau often conducts data collections and performs associated work for sponsoring agencies on a reimbursable basis. The work performed by the Census Bureau under such contracts is in the scope of these statistical quality standards, whether performed under Title 13, Title 15, or another authorization. If a sponsor's requirements or funding constraints result in noncompliance with these standards, the Census Bureau's manager for the program must obtain a waiver, except where noted in the standards.

For the purposes of these standards, information products include printed, electronic, or digital formats (e.g., Web, CD, DVD, and tape) of: news releases; Census Bureau publications; working papers (including technical papers or reports); professional papers (including journal articles, book chapters, conference papers, poster sessions, and written discussant comments); abstracts; research reports used to guide decisions about Census Bureau programs; public presentations at external events (e.g., seminars or conferences); handouts for presentations; tabulations and custom tabulations; public-use data files; statistical graphs, figures, and maps; and the documentation disseminated with these information products.

Exclusions to the Scope

None of the following exclusions apply to Statistical Quality Standard S1, *Protecting Confidentiality*, or the requirements for protecting confidentiality in the individual standards.

These standards do not apply to:

- Information products intended for internal Census Bureau use that are not intended for public dissemination.
- Information products delivered to agencies within the Department of Commerce for their internal use.
- Internal procedural or policy manuals prepared for the management of the Census Bureau and the Department of Commerce that are not intended for public dissemination.
- Information products that result from the Census Bureau's administrative or management processes.
- Information products released in response to a Freedom of Information Act request.
- Documents intended only for communications between agencies, within agencies, or with individuals outside the Census Bureau if the documents contain no data and do not discuss analyses or methodological information.
- Informal communications between Census Bureau employees and colleagues in other organizations that do not disseminate Census Bureau data or results based on Census Bureau data.
- Information products delivered to sponsors or oversight agencies, including the Congress, relating to the management of Census Bureau programs.
- Information products authored by external researchers at the Census Bureau's Research Data Centers.
- Information products that use Census Bureau data and are authored by Special Sworn Status individuals employed by other federal agencies or organizations for their agencies (e.g., SSA, GAO, and CBO).

- Information products generated by other agencies or organizations to which the Census Bureau has given only technical assistance or training. However, Census Bureau staff providing such assistance should consider these standards as guidelines.
- Information products developed from surveys intended to measure Census Bureau customers' or users' satisfaction with Census Bureau products or to measure Census Bureau employees' job satisfaction. However, any public release of results of such surveys must explain that they do not meet the Census Bureau's statistical quality standards because the respondents are self-selected and may not be representative of all customers, all users, or all employees.
- Communications released via social media. Social media must not be used to disseminate data or statistical analyses not previously cleared for external release. Such communications must follow the Census Bureau's *Policies and Procedures Governing the Use of Social Media*.

The scope statements of the individual standards provide additional information to clarify the scope and to list exclusions specific to each standard.

3. Responsibilities

All Census Bureau employees and Special Sworn Status individuals are responsible for following the Census Bureau's statistical quality standards in their work to develop, deliver, and release information products.

Responsibilities of the Program Areas and the Supporting Directorates and Divisions

Divisions and offices within the Economic Programs, Demographic Programs, and Decennial Census plan, process, analyze, and disseminate data. The Census Bureau's Center for Statistical Research and Methodology supports all three directorates in areas of statistical, methodological, behavioral, and technological research and development. The Field Operations Directorate and Information Technology Directorate collect, transmit, and process data for demographic household surveys, the Decennial Census, the Economic Census and surveys, and the Government Census and surveys. The Census Bureau's other directorates and divisions provide various types of administrative, logistical, and strategic support to the program areas.

The responsibilities of the program areas and the supporting directorates and divisions with respect to these statistical quality standards include:

- Ensuring that the necessary resources are available to comply with the statistical quality standards.
- Implementing and verifying compliance with the statistical quality standards.

  Guidance on implementing the standards and verifying compliance can be obtained from the program area's Methodology and Standards (M&S) Council representative as shown in Table 1.

Table 1. M&S Council Representatives	
Program Directorate	**M&S Council Representative**
Decennial Census Directorate	Chief, Decennial Statistical Studies Division
Demographic Programs Directorate	Chief, Demographic Statistical Methods Division
Economic Programs Directorate	Chief, Office of Statistical Methods and Research for Economic Programs
All other directorates	Chief, Center for Statistical Research and Methodology

- Reporting situations where requirements of the standards might need revision (e.g., a program's processes or products may have changed so that some requirements of the statistical quality standards may also need to be revised).
- Following the procedure to obtain a waiver if unable to comply with one or more of the statistical quality standards.

Responsibilities of the Methodology and Standards Council

The Census Bureau's M&S Council consists of the division and office chiefs of the statistical methodology groups in the various program areas. The Council advises the Census Bureau's Program Associate Directors on policy and issues affecting research and methodology for Census Bureau programs. The Council also ensures the use of sound statistical methods and practices, and facilitates communication and coordination of statistical methodology and research throughout the Census Bureau and the broader statistical community.

The responsibilities of the M&S Council with respect to these statistical quality standards include:

- Promoting awareness of and compliance with the Census Bureau's statistical quality standards.
- Reviewing waiver requests and forwarding their recommendation for approval or denial of the waiver to the Program Associate Director.
- Conducting periodic reviews and evaluations of the standards to study how well the standards are working and to identify difficulties in implementation.
- Maintaining an archive of evaluation findings, waiver requests, and suggestions for improvement to inform future revisions of the Census Bureau's statistical quality standards.
- Updating the standards as needed.

The responsibilities of the individual M&S Council members for their directorates (See Table 1.) include:

- Provide guidance on interpreting the standards to the programs in their directorates and to directorates that participate in conducting and implementing their programs (e.g., the Field Operations Directorate).
- Provide assistance in implementing and verifying compliance with the standards to the programs in their directorates and to directorates that participate in conducting and implementing their programs (e.g., the Field Operations Directorate).

4. Interpreting and Using the Standards

The complete set of statistical quality standards includes process standards (designated with "A" through "F") and supporting standards (designated with "S"). The process standards are organized according to the different processes associated with developing and releasing information products. The organizational framework for these process standards is:

A. Planning and Development
B. Collecting and Acquiring Data
C. Capture and Processing Data
D. Producing Estimates and Measures
E. Analyzing Data and Reporting Results
F. Releasing Information

The supporting standards address issues that cut across all the process standards. The two supporting standards are S1, *Protecting Confidentiality*, and S2, *Managing Data and Documents*.

The standards are written at a broad level of detail, to apply to all the Census Bureau's programs and products. They describe *what* is required and do not delineate procedures for *how* to satisfy the requirements. Each standard has a list of key terms that are used in the standard. These terms are defined in the glossary to provide clarification on their use in relation to these standards.

To help managers interpret the requirements of the standards, examples are often provided. These examples are intended to aid the program manager in understanding the requirements and to provide guidance on the types of actions that may be useful in satisfying the requirements. It is important to note that the examples listed under a requirement are not all-inclusive; nor will every example apply to every program or product. Finally, there may be more than one acceptable way to comply with a requirement. That is, several equally acceptable actions might be performed to comply with a requirement, rather than only one unique set of actions.

Program managers must use their judgment to determine which actions must be performed for their program to comply with a requirement. The program manager is expected to carry out all the actions needed to comply with a requirement. This may include performing activities not listed in the examples. The expectation is that program managers will balance the importance of the information product and the size of the project with the constraints of budget, schedule, and resources when determining how to comply with the requirements.

If the program manager believes it is not feasible to comply with a requirement, the program manager must request a waiver. The *Waiver Procedure* provides a standard mechanism to exempt a program from compliance with a statistical quality standard when such an exemption is warranted. The Waiver Procedure also promotes proper management and control in implementing the standards. Finally, the Waiver Procedure ensures that appropriate documentation of exceptions to the standards is generated and maintained to inform future revisions of the statistical quality standards.

5. History of the Development of the Standards

The Census Bureau has a long history of delivering high quality data about the nation's people and economy. *Technical Paper 32*, *Standards for Discussion and Presentation of Errors in Data*, issued in March 1974, is an example of the Census Bureau's commitment to promote transparency in the quality of the information and data products it delivers to the public and to its sponsors.[2]

Over the years, the Census Bureau has developed additional guidance regarding the quality of its products and in 1998 began to formalize its efforts to ensure quality in its products and processes. The Census Bureau began this more formal approach by instituting a quality program based on a foundation of quality principles, standards, and guidelines. The paper, *Quality Program at the U.S. Census Bureau*, describes the beginnings of the Census Bureau's Quality Program (Proceedings of the International Conference on Quality in Official Statistics, 2001).

In 2001, the Census Bureau issued the first of eleven new statistical quality standards. Several of these standards updated the content of *Technical Paper 32*. Over the next four years, ten more standards were developed.

In 2005, after conducting a benchmarking study of the standards of other statistical organizations, the M&S Council initiated a more coordinated approach for developing a comprehensive set of statistical quality standards. While the existing standards were a good start, this approach aimed to improve consistency and cohesion among the standards, as well as to reflect all the requirements of the OMB's *Standards and Guidelines for Statistical Surveys* in the context of the Census Bureau's programs, products, and processes.

The new approach to developing statistical quality standards relied on five key components: 1) a dedicated staff to develop the standards, rather than ad hoc teams; 2) contractor assistance; 3) multiple reviews of draft standards to obtain feedback from the program areas; 4) focus groups to obtain more thoughtful and attentive input from the program areas; and 5) a documented, consistent development process.

The Census Bureau began developing these standards in May 2006. The process was completed in May 2010, when the Census Bureau issued these statistical quality standards.

---

[2] *Technical Paper 32* is available from the U.S. Government Printing Office, Washington, DC 20401. It was revised in: Gonzalez, M., Ogus, J., Shapiro, G., and Tepping, B. *Journal of the American Statistical Association*, Vol. 70, No. 351, Part 2: Standards for Discussion and Presentation of Errors in Survey and Census Data (Sep., 1975), pp. 5-23. http://www.jstor.org/stable/2286149

PLANNING AND DEVELOPMENT

Statistical Quality Standard A1
Planning a Data Program

Purpose: The purpose of this standard is to ensure that plans are developed when initiating a new or revised data program.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals that receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to planning data programs (e.g., surveys, censuses, and administrative records programs) that will release information products to the public, to sponsors, or to other customers.

Exclusions:
The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Note: Specific planning requirements for each stage of the data program are addressed in other statistical quality standards. For example, Statistical Quality Standard E1, *Analyzing Data*, includes requirements for planning data analyses.

Key Terms: Administrative records, bridge study, business identifiable information, census, data collection, data program, information products, microdata, personally identifiable information, reimbursable project, response rate, sample design, sample survey, stakeholder, target population, and users.

Requirement A1-1: The provisions of federal laws (e.g., Title 13, Title 15, and Title 26) and Census Bureau policies and procedures on privacy and confidentiality (e.g., Data Stewardship Policies) must be followed in planning and designing any programs that will collect personally identifiable information or business identifiable information. (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement A1-2: An overall program plan must be developed that includes the following:

1. A justification for the program, including:
   a. A description of the program goals.
   b. A description of stakeholder requirements and expectations.
   c. A description of the intended information products (e.g., tabulations, confidential microdata, or public-use files).
   d. A description of revisions to an ongoing program, including:
      1) Changes to key estimates, methods, or procedures.
      2) The usefulness of the revisions for conducting analyses and for informing policymakers and stakeholders.

3) Planned studies to measure the effects of the changes to key estimates and time series (e.g., overlap samples or bridge studies).

e. For sample survey and census programs (i.e., programs that do not rely solely on administrative records), a description of the steps taken to prevent unnecessary duplication with other sources of information, including a list of related (current and past) federal and non-federal studies, surveys, and reports that were reviewed.

Notes:

(1) The Office of Management and Budget's (OMB) *Guidance on Agency Survey and Statistical Information Collections* provides information on preparing OMB clearance packages for surveys used for general purpose statistics or as part of program evaluations or research studies.

(2) The OMB's *Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity* provides standards for programs collecting data on race and ethnicity.

(3) The OMB's *Standards for Defining Metropolitan and Micropolitan Statistical Areas* provides standards for collecting, tabulating, and publishing statistics for geographic areas.

2. An initial schedule that identifies key milestones for the complete program cycle from planning to data release.

Generally, the program cycle includes the following stages:

- Planning a data program (Statistical Quality Standard A1).
- Developing the data collection instrument and sample design (Statistical Quality Standards A2 and A3).
- Establishing and implementing data collection methods and acquiring administrative records (Statistical Quality Standards B1 and B2).
- Capturing and processing data (Statistical Quality Standards C1, C2, C3, and C4).
- Producing estimates and quality measures (Statistical Quality Standards D1, D2, and D3).
- Analyzing data and reporting results (Statistical Quality Standards E1 and E2).
- Reviewing information products (Statistical Quality Standard E3).
- Releasing information products (Statistical Quality Standards F1 and F2).

Note: Managers responsible for each stage of the program generally are expected to prepare milestone schedules for their stages. The overall program manager can use these individual schedules to prepare the overall milestone schedule.

3. An initial, overall cost estimate that identifies the resources needed and itemizes the costs to carry out the program.

Note: Managers responsible for each stage of the program generally are expected to prepare cost estimates for their stages. The overall program manager can use these estimates to prepare the overall cost estimate.

4. A description of deliverables to be received as the result of any contracts originated by the Census Bureau, including any documentation to be provided by contractors.

   Examples of such deliverables include:
   - Computer software or hardware.
   - Data files.
   - Advertising or outreach services and materials.
   - Specifications for software or hardware.
   - Quality control or quality assurance procedures, criteria, and results.

Sub-Requirement A1-2.1: When the sponsor of a reimbursable project requests the Census Bureau to carry out activities that do not comply with our Statistical Quality Standards or deliver products that do not conform with the standards, the program manager must:

1. Obtain a waiver to carry out the noncompliant activities or to deliver the nonconforming products before agreeing to conduct the project. (See the *Waiver Procedure* for the procedures on obtaining a waiver.)
2. Obtain from the sponsor a copy of the clearance package approved by the OMB, including any associated terms of clearance.
3. Deliver to the sponsor written documentation that describes the following for each area of noncompliance:
   a. The details regarding the noncompliance issue.
   b. The consequences of performing the noncompliant work.
   c. The actions recommended by the Census Bureau that would result in compliance.

Requirement A1-3: For sample survey and census programs**,** a preliminary survey design must be developed that describes the:

1. Target population and sampling frame.
2. Sample design.
3. Key data items and key estimates.
4. Response rate goals.
5. Data collection methods.
6. Analysis methods.

Requirement A1-4: For administrative records projects, a preliminary study design must be developed that describes the:

1. Target population.
2. Coverage of the target population by the administrative records.
3. Key data items and key estimates.
4. Methods of integrating data sources, if more than one is used.
5. Analysis methods.

Note: See the Administrative Records Handbook for complete information on planning a project that uses administrative records.

Requirement A1-5: Any contract or statement of work originated by the Census Bureau for deliverables that will be used in information products released by the Census Bureau must

include provisions that the contractor comply with the Census Bureau's statistical quality standards.

Requirement A1-6: Quality control checks must be performed to ensure the accuracy and completeness of the program plans, including all schedules, cost estimates, agreements (e.g., memoranda of understanding, statements of work, and contracts), survey designs, and study designs.

Requirement A1-7: Documentation needed to replicate and evaluate the data program must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Program plans, including cost estimates and schedules, after all revisions.
- Survey designs.
- Study designs.
- Decision memoranda.

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard A2
Developing Data Collection Instruments and Supporting Materials

Purpose: The purpose of this standard is to ensure that data collection instruments and supporting materials are designed to promote the collection of high quality data from respondents.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals that receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the development or redesign of data collection instruments and supporting materials. The types of data collection instruments and supporting materials covered by this standard include:

- Paper and electronic instruments (e.g., CATI, CAPI, Web, and touch tone data entry).
- Self-administered and interviewer-administered instruments.
- Instruments administered by telephone or in person.
- Respondent letters, aids, and instructions.
- Mapping and listing instruments used for operations, such as address canvassing, group quarters frame development, and the Local Update of Census Addresses (LUCA).

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Data collection instruments and supporting materials where the Census Bureau does not have control over the content or format, such as the paper and electronic instruments used for collecting import and export merchandise trade data.

Key Terms: Behavior coding, CAPI, CATI, cognitive interviews, data collection instrument, field test, focus group, graphical user interface (GUI), imputation, integration testing, methodological expert review, nonresponse, pretesting, questionnaire, record linkage, respondent burden, respondent debriefing, split panel test, and usability testing.

Requirement A2-1: Throughout all processes associated with data collection, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement A2-2: A plan must be produced that addresses:

1. Program requirements for the data collection instrument and the graphical user interface (GUI), if applicable (e.g., data collection mode, content, constraints, and legal requirements).
2. Supporting materials needed for the data collection (e.g., brochures, flashcards, and advance letters).
3. Pretesting of the data collection instrument and supporting materials.
4. Verification and testing to ensure the proper functioning of the data collection instrument and supporting materials.

Notes:

(1) Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including the development of schedules and costs.

(2) See the *Guidelines for Designing Questionnaires for Administration in Different Modes* and the *Economic Directorate Guidelines on Questionnaire Design* for guidance on designing data collection instruments.

(3) Data Stewardship Policy DS016, *Respondent Identification Policy*, contains policy requirements for data collection operations involving households where respondents in households provide information.

Requirement A2-3: Data collection instruments and supporting materials must be developed and tested in a manner that balances (within the constraints of budget, resources, and time) data quality and respondent burden.

Sub-Requirement A2-3.1: Specifications for data collection instruments and supporting materials, based on program requirements, must be developed and implemented.

Examples of topics that specifications might address include:

- Requirements for programming the instrument to work efficiently. For example:
  - Built-in edits or range checks for electronic data collection instruments (e.g., edits for numeric data that must be within a pre-specified range).
  - Compliance with the CATI/CAPI Screen Standards for GUI (Windows-based) Instruments and Function Key Standards for GUI Instruments. (See the Technologies Management Office's Authoring Standards *Blaise Standards for Windows Surveys*).
  - Input and output files for data collection instruments.
- Segmented boxes for paper data collection instruments to facilitate scanning.
- Paper size, color, thickness, and formatting to ensure compatibility with data capture and processing systems for paper data collection instruments.
- Frequently Asked Questions about the data collection.
- Supporting materials, such as Help materials and instructions.

Note: The Census Bureau Guideline *Presentation of Data Edits to Respondents in Electronic Self-Administered Surveys* presents recommendations for designing editing functionality, presentation, and wording in both demographic and economic self-administered electronic surveys.

Sub-Requirement A2-3.2: Data collection instruments and supporting materials must clearly state the following required notifications to respondents:

1. The reasons for collecting the information.
2. A statement on how the data will be used.
3. An indication of whether responses are mandatory (citing authority) or voluntary.
4. A statement on the nature and extent of confidentiality protection to be provided, citing authority.
5. An estimate of the average respondent burden associated with providing the information.
6. A statement requesting that the public direct comments concerning the burden estimate and suggestions for reducing this burden to the appropriate Census Bureau contact.
7. The OMB control number and expiration date for the data collection.
8. A statement that the Census Bureau may not conduct, and a person is not required to respond to, a data collection request unless it displays a currently valid OMB control number.

Sub-Requirement A2-3.3: Data collection instruments and supporting materials must be pretested with respondents to identify problems (e.g., problems related to content, order/context effects, skip instructions, formatting, navigation, and edits) and then refined, prior to implementation, based on the pretesting results.

Note: On rare occasions, cost or schedule constraints may make it infeasible to perform complete pretesting. In such cases, subject matter and cognitive experts must discuss the need for and feasibility of pretesting. The program manager must document any decisions regarding such pretesting, including the reasons for the decision. If no acceptable options for pretesting can be identified, the program manager must apply for a waiver. (See the *Waiver Procedure* for the procedures on obtaining a waiver.)

1. Pretesting must be performed when:
   a. A new data collection instrument is developed.
   b. Questions are revised because the data are shown to be of poor quality (e.g., unit or item response rates are unacceptably low, measures of reliability or validity are unacceptably low, or benchmarking reveals unacceptable differences from accepted estimates of similar characteristics).
   c. Review by cognitive experts reveals that adding pretested questions to an existing instrument may cause potential context effects.
   d. An existing data collection instrument has substantive modifications (e.g., existing questions are revised or new questions added).

   Note: Pretesting is not required for questions that performed adequately in another survey.

2. Pretesting must involve respondents or data providers who are in scope for the data collection. It must verify that the questions:
   a. Can be understood and answered by potential respondents.
   b. Can be administered properly by interviewers (if interviewer-administered).
   c. Are not unduly sensitive and do not cause undue burden.

Examples of issues to verify during pretesting:

- The sequence of questions and skip patterns is logical and easy-to-follow.
- The wording is concise, clear, and unambiguous.
- Fonts (style and size), colors, and other visual design elements promote readability and comprehension.

3. One or more of the following pretesting methods must be used:
   a. Cognitive interviews.
   b. Focus groups, but only if the focus group completes a self-administered instrument and discusses it afterwards.
   c. Usability techniques, but only if they are focused on the respondent's understanding of the questionnaire.
   d. Behavior coding of respondent/interviewer interactions.
   e. Respondent debriefings in conjunction with a field test or actual data collection.
   f. Split panel tests.

   Notes:
   (1) Methodological expert reviews generally do not satisfy this pretesting requirement. However, if a program is under extreme budget, resource, or time constraints, the program manager may request cognitive experts in the Center for Statistical Research and Methodology or on the Response Improvement Research Staff to conduct such a review. The results of this expert review must be documented in a written report. If the cognitive experts do not agree that an expert review would satisfy this requirement, the program manager must apply for a waiver.
   (2) Multiple pretesting methods should be used as budget, resources, and time permits to provide a thorough evaluation of the data collection instrument and to document that the data collection instrument "works" as expected. In addition, other techniques used in combination with the pretesting methods listed above may be useful in developing data collection instruments. (See Appendix A2, *Questionnaire Testing and Evaluation Methods for Censuses and Surveys*, for descriptions of the various pretesting methods available.)

4. When surveys or censuses are administered using multiple modes and meaningful changes to questions are made to accommodate the mode differences, all versions must be pretested.

   Meaningful changes to questions to accommodate mode differences include changes to the presentation of the question or response format to reflect mode-specific functional constraints or advantages. In these cases, the proposed wording of each version must be pretested to ensure consistent interpretation of the intent of the question across modes, despite structural format or presentation differences. As long as the proposed wording of each version is pretested, testing of the mode (e.g., paper versus electronic) is not required, although it may be advisable.

5. Data collection instruments in any languages other than English must be pretested in the languages that will be used to collect data during production. Pretesting supporting materials in these languages is not required, but is recommended.

   Note: The Census Bureau Guideline *Language Translation of Data Collection Instruments and Supporting Materials* provides guidance on translating data collection instruments and supporting materials from English to another language.

Sub-Requirement A2-3.4: Data collection instruments and supporting materials must be verified and tested to ensure that they function as intended.

Examples of verification and testing activities include:

- Verifying that the data collection instrument's specifications and supporting materials reflect the sponsor's requirements (e.g., conducting walk-throughs to verify the appropriateness of specifications).
- Verifying that the data collection instrument and supporting materials meet all specifications (e.g., verifying correctness of skip patterns, wording, instrument fills, and instrument edits).
- Conducting integration testing using mock input files with realistic scenarios to test all parts of the data collection instrument together (e.g., front, middle, and back modules).
- Conducting usability testing to discover and eliminate barriers that keep respondents from completing the data collection instrument accurately and efficiently.
- Conducting output tests to compare the output of the data collection instrument under development with that of its predecessor (if the data collection has been done with a similar instrument in the past).
- Verifying that user interfaces work according to specifications.
- Verifying that user interfaces for electronic forms adhere to IT Standard 15.0.2, *Web Development Requirements and Guidelines*, and any other guidance applicable to the program.
- Verifying that Web-based data collection instruments comply with requirements of Section 508 of the U.S. Rehabilitation Act.
- Verifying that paper data collection instruments are compatible with the program's data capture and processing systems.

Note: The Census Bureau Guideline *Computer Assisted Personal Interviewing* reflects recommended practices for ensuring the quality of CAPI.

Requirement A2-4: Documentation needed to replicate and evaluate the development of data collection instruments and supporting materials must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans for the development and testing of the data collection instrument and supporting materials.

- Specifications for the data collection instruments and supporting materials.
- Results of questionnaire development research (e.g., pretesting results, expert review reports, and site visit reports).
- Input files used to test the final production instrument and reports of testing results.
- Computer source code for the production data collection instrument along with information on the version of software used to develop the instrument.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Appendix A2[1]
Questionnaire Testing and Evaluation Methods for Censuses and Surveys

Pretesting is critical to the identification of problems for both respondents and interviewers with regard to question content, order/context effects, skip instructions, and formatting. Problems with question content, for example, include confusion over the meaning of the question as well as misinterpretation of individual terms or concepts. Problems with skip instructions may result in missing data and frustration by interviewers and/or respondents. Formatting concerns are relevant to self-administered questionnaires and may lead to respondent confusion and a loss of information.

"Pretesting" is a broad term that applies to many different methods or combinations of methods that can be used to test and evaluate questionnaires. These methods are valuable for identifying problems with draft questionnaires, but they have different strengths and weaknesses, and may be most useful at different stages of questionnaire/instrument development. Typically, using several pretesting methods is more effective in identifying problem questions and suggesting solutions than using just a single method. This appendix briefly describes the different types of pretesting methods, their strengths and weaknesses, and situations where they are most beneficial.

The enumeration and description of potential pretesting and evaluation methods in this appendix is meant to cover all the available techniques; however, some techniques do not satisfy the pretesting requirement of Statistical Quality Standard A2: *Developing Data Collection Instruments and Supporting Materials*. Other methods satisfy the requirement only under special circumstances. The pretesting requirement of Standard A2 identifies the methods that must be used to pretest census and survey questions.

Although the pretesting requirement of Standard A2 must be satisfied, the appropriateness of the methods and the resources available to implement them should be considered in determining which pretesting methods to use.

Pretesting and evaluation techniques fall into two major categories – pre-field and field techniques. Generally, pre-field techniques are used during the preliminary stages of questionnaire development. Pre-field techniques include:

- Respondent focus groups. (This method does **not** satisfy the pretesting requirement, unless the focus group completes and discusses a self-administered questionnaire.)
- Exploratory or feasibility visits to companies or establishment sites. (This method does **not** satisfy the pretesting requirement.)
- Cognitive interviews. (This method satisfies the pretesting requirement.)

[1] This appendix is based on two sources: 1) Protocol for Pretesting Demographic Surveys at the Census Bureau, prepared by Theresa DeMaio, Nancy Mathiowetz, Jennifer Rothgeb, Mary Ellen Beach, and Sharon Durant, dated June 28, 1993; and 2) Evolution and Adaptation of Questionnaire Development, Evaluation and Testing in Establishment Surveys, by Diane Willimack, Lars Lyberg, Jean Martin, Lilli Japec, and Patricia Whitridge. Monograph Paper for the International Conference on Questionnaire Development, Evaluation and Testing Methods, Charleston, SC, November, 2002.

- Usability techniques. (This method does **not** satisfy the pretesting requirement unless it is focused on respondent understanding of a self-administered or interviewer-administered questionnaire.)
- Methodological expert reviews. (This method does **not** satisfy the pretesting requirement.)

Field techniques are used to evaluate questionnaires tested under field conditions, either in conjunction with a field test or during production data collection. Using field techniques during production data collection would be appropriate only for ongoing or recurring surveys. Field techniques include:

- Behavior coding of interviewer-respondent interactions. (This method satisfies the pretesting requirement.)
- Respondent debriefings. (This method satisfies the pretesting requirement.)
- Interviewer debriefings. (This method does **not** satisfy the pretesting requirement.)
- Analysts' feedback. (This method does **not** satisfy the pretesting requirement.)
- Split panel tests. (This method satisfies the pretesting requirement.)
- Analysis of item nonresponse rates, imputation rates, edit failures, or response distributions. (This method does **not** satisfy the pretesting requirement.)

PRE-FIELD TECHNIQUES

Respondent Focus Groups are used early in the questionnaire development cycle and can be used in a variety of ways to assess the question-answering process. Generally, the focus group technique does **not** satisfy the pretesting requirement, because it does not expose respondents to a questionnaire.

The only use of focus groups that satisfies the pretesting requirement is to have the group complete a self-administered questionnaire, followed by a discussion of the experience. This provides information about the appearance and formatting of the questionnaire and reveals possible content problems.

Focus groups can be used before questionnaire construction begins to gather information about a topic, such as:

- How potential respondents structure their thoughts about a topic.
- How respondents understand general concepts or specific terminology.
- Respondents' opinions about the sensitivity or difficulty of the questions.
- How much burden is associated with gathering the information necessary to answer a question.

Focus groups can also be used to identify variations in language, terminology, or the interpretation of questions and response options. Used in this way, they may provide quicker access to a larger number of people than is possible with cognitive interviews. One of the main advantages of focus groups is the opportunity to observe an increased amount of interaction on a topic in a short time. The group interaction is of central importance – it can result in information and insights that may be less accessible in other settings. However, precisely because of this group interaction, the focus group does not permit a good test of an individual's response process

when alone. Moreover, in focus groups the researcher does not have as much control over the process as with cognitive interviews or interviewer-administered questionnaires. One or two people in the group may dominate the discussion and restrict the input from other group members.

Exploratory or Feasibility Studies are another common method for evaluating survey content relative to concepts. Economic survey practitioners typically call these studies *company or site visits* because they carry out the studies at the site of the business or institution. Because these visits are conducted before the questionnaire has been developed, they do **not** satisfy the pretesting requirement.

Because economic surveys rely heavily on business or institutional records, the primary goal of these site visits is to determine the availability of the desired data in records, their periodicity, and the definition of the concept as used in company records. Other goals include assessment of response burden and quality and the identification of the appropriate respondent.

The design of these company or site visits tends to vary a great deal. Because they are exploratory in nature, the activity may continue until the economic survey or program staff sufficiently understands the respondents' views of the concepts, resources permitting of course. Purposive or convenience samples are selected that target key data providers. Sample sizes are small, perhaps as few as five and rarely more than thirty. Typically, several members of the survey or program staff, who may or may not include questionnaire design experts, conduct meetings with multiple company employees involved in government reporting. Information gained during these visits helps determine whether the survey concepts are measurable, what the specific questions should be, how to organize or structure the questions related to the concept of interest, and to whom the form should be sent.

Exploratory or feasibility studies may be multi-purpose. In addition to exploring data availability for the concept of interest, survey or program staff may also set up reporting arrangements and review operating units to ensure correct coverage. A common by-product of these visits is to solidify relationships between the companies and the survey or program staff.

Cognitive Interviews are used in the later part of the questionnaire development cycle, after a questionnaire has been constructed based on information from focus groups, site visits, or other sources. They consist of one-on-one interviews using a draft questionnaire in which respondents describe their thoughts while answering the survey questions. Cognitive interviews provide an important means of learning about respondents' problems with the questionnaire directly from them. Because this technique tests the questionnaire with potential respondents, it satisfies the pretesting requirement.

In addition, small numbers of interviews (as few as fifteen) can yield information about major problems if respondents repeatedly identify the same questions and concepts as sources of confusion. Because sample sizes are small, iterative pretesting of an instrument is often possible. After one round of interviews is complete, researchers can diagnose problems, revise question wording to solve the problems, and conduct additional interviews to see if the new questions are successful.

Cognitive interviews may or may not be conducted in a laboratory setting. The advantage of the laboratory is that it offers a controlled environment for conducting the interview, and provides the opportunity for video as well as audio recording. However, laboratory interviews may be impractical or unsuitable. For example, economic surveys rarely conduct cognitive interviews in a laboratory setting. Rather, cognitive testing of economic surveys is usually conducted on-site at the offices or location of the business or institutional respondent. One reason for this approach is to enable business or institutional respondents' to have access to records. Another is business respondents' reluctance to meet outside their workplaces for these interviews. In many economic surveys, which tend to be relatively lengthy and require labor-intensive data retrieval from records, testing may be limited to a subset of questions or sections rather than the entire questionnaire. Thus, researchers must be careful to set the proper context for the target questions.

"Think aloud" interviews, as cognitive interviews have come to be called, can be conducted either concurrently or retrospectively – that is, the respondents' verbalizations of their thought processes can occur either during or after the completion of the questionnaire. As the Census Bureau conducts them, cognitive interviews incorporate follow-up questions by the researcher in addition to the respondent's statement of his or her thoughts.

Probing questions are used when the researcher wants to have the respondent focus on particular aspects of the question-response task. For example, the interviewer may ask how respondents chose among response choices, how they interpreted reference periods, or what a particular term meant. *Paraphrasing* (asking the respondents to repeat the question in their own words) permits the researcher to learn whether the respondent understands the question and interprets it in the manner intended, and it may reveal better wordings for questions.

In surveys of businesses or institutions, in which data retrieval often involves business records, probing and paraphrasing techniques are often augmented by questions asking respondents to describe those records and their contents or to show the records to the researcher. Since data retrieval tends to be a labor-intensive process for business respondents, frequently requiring the use of multiple sources or consultation with colleagues, it is often unrealistic for researchers to observe the process during a cognitive interview. Instead, *hypothetical probes* are often used to identify the sources of data, discover respondents' knowledge of and access to records, recreate likely steps taken to retrieve data from records or to request information from colleagues, and suggest possible estimation strategies.

Usability Techniques are used to aid development of automated questionnaires. Objectives are to discover and eliminate barriers that keep respondents from completing an automated questionnaire accurately and efficiently with minimal burden. Usability tests that are focused on respondent understanding of the questionnaire satisfy the pretesting requirement. Usability tests that are focused on the interviewers' ability to administer the instrument do not satisfy the pretesting requirement; however, they are recommended for interviewer-administered electronic questionnaires.

Aspects that deserve attention during usability testing include the language, fonts, icons, layout, organization, and interaction features, such as data entry, error recovery, and navigation. Typically, the focus is on instrument performance in addition to how respondents interpret survey questions. Problems identified during testing can then be eliminated before the instrument is finalized.

As with paper questionnaires, different usability techniques are available depending upon the stage of development. One common technique is called the *usability test*. These tests are similar to cognitive interviews – that is, one-on-one interviews that elicit information about the respondent's thought process. Respondents are given a *task*, such as "Complete the questionnaire," or smaller subtasks, such as "Send your data to the Census Bureau." The *think aloud*, *probing*, and *paraphrasing* techniques are all used as respondents complete their assigned tasks. Early in the design phase, usability testing with respondents can be done using *low fidelity questionnaire prototypes* (i.e., mocked-up paper screens). As the design progresses, versions of the automated questionnaire can be tested to choose or evaluate basic navigation features, error correction strategies, etc.

Disability accommodation testing is a form of usability testing which evaluates the ability of a disabled user to access the questionnaire through different assistive technologies, such as a screen reader. *Expert reviews* (see below) are also part of the repertoire of usability techniques.

Research has shown that as few as three participants can uncover half of the major usability problems; four to five participants can uncover 80 percent of the problems; and ten participants can uncover 90 percent of the problems (Dumas and Redish, 1999).

Finally, in a *heuristic review*, an expert compares the electronic survey instrument with usability principles that should be followed by all user interfaces (Nielsen, 1993).

Methodological Expert Reviews, conducted by survey methodologists or questionnaire-design experts, evaluate any difficulties potential interviewers and respondents may have with the questionnaire. Seasoned survey researchers who have extensive exposure to either the theoretical or practical aspects of questionnaire design use their expertise to achieve this goal. Because respondents do not provide direct input in these reviews, in general they do **not** satisfy the pretesting requirement. Usually these reviews are conducted early in the questionnaire development process and in concert with other pretest methods.

Expert reviews may be used instead of respondent-based pretesting only as a last resort, when extreme time constraints prevent the use of other pretesting methods. In such instances, survey methodology experts must conduct the reviews and document the results in a written report. The decision to use expert reviews rather than respondent-based pretesting must be made by subject-matter areas in consultation with the methodological research areas in the Center for Statistical Research and Methodology and on the Response Improvement Research Staff.

The cognitive appraisal coding system (Forsyth and Lessler, 1991) is a tool providing a systematic approach to the methodological expert review process. Like methodological expert reviews, results are used to identify questions that have potential for reporting errors. This tool is

particularly effective when used by questionnaire design experts who understand the link between the cognitive response process and measurement results. However, novice staff or subject-area staff also can use this tool as a guide in their reviews of questionnaires.

Methodological expert reviews also can be conducted as part of a usability evaluation. Typically, this review is performed with an automated version of the questionnaire, although it need not be fully functional. Experts evaluate the questionnaire for consistency and application of user-centered principles of user-control, error prevention and recovery, and ease of navigation, training, and recall.

FIELD TECHNIQUES

Field techniques may be used with pretests or pilot tests of questionnaires or instruments and survey processes. They may also be employed in ongoing periodic (or recurring) surveys. The value of testing draft questionnaires with potential survey respondents cannot be overstated, even if it simply involves observation and evaluation by questionnaire developers. However, the following pretesting methods can be used to maximize the benefits of field testing.

Behavior Coding of Respondent/Interviewer Interactions involves systematic coding of the interaction between interviewers and respondents from live or taped field or telephone interviews to collect quantitative information. Using this pretesting method satisfies the pretesting requirement.

The focus here is on specific aspects of how the interviewer asks the question and how the respondent reacts. When used for questionnaire assessment, the behaviors that are coded focus on behaviors that indicate problems with the question, the response categories, or the respondent's ability to form an adequate response. For example, if a respondent asks for clarification after hearing the question, it is likely that some aspect of the question caused confusion. Likewise, if a respondent interrupts the question before the interviewer finishes reading it, then the respondent misses information that might be important to giving a correct answer. For interviewer-administered economic surveys, the coding scheme may need to be modified from traditional household applications, because interviewers for establishment surveys tend to be allowed greater flexibility.

In contrast to the pre-field techniques described earlier, the use of behavior coding requires a sample size sufficient to address analytic requirements. For example, if the questionnaire contains many skip patterns, it is necessary to select a large enough sample to permit observation of various paths through the questionnaire. In addition, the determination of sample sizes for behavior coding should take into account the relevant population groups for which separate analysis is desired.

Because behavior coding evaluates all questions on the questionnaire, it promotes systematic detection of questions that elicit large numbers of behaviors that reflect problems. However, it is not usually designed to identify the source of the problems. It also may not be able to distinguish which of several similar versions of a question is better.

Finally, behavior coding does not always provide an accurate diagnosis of problems. It can only detect problems that are manifest in interviewer or respondent behavior. Some important problems, such as respondent misinterpretations, may remain hidden because both respondents and interviewers tend to be unaware of them. Behavior coding is not well-suited for identifying such problems.

Respondent Debriefing uses a structured questionnaire after data are collected to elicit information about respondents' interpretations of survey questions. Use of this method satisfies the pretesting requirement.

The debriefing may be conducted by incorporating structured follow-up questions at the end of a field test interview or by re-contacting respondents after they return a completed self-administered questionnaire. In economic surveys, respondent debriefings sometimes are called "response analysis surveys" ("RAS") or "content evaluations." Respondent debriefings usually are interviewer-administered, but may be self-administered. Some Census Bureau economic surveys have conducted respondent debriefings by formulating them as self-administered questionnaires and enclosing them with survey forms during pilot tests or production data collections.

Sample sizes and designs for respondent debriefings vary. Sample sizes may be as small as 20 or as large as several hundred. Designs may be either random or purposive, such as conducting debriefings with respondents who exhibited higher error rates or errors on critical items. Since the debriefing instrument is structured, empirical summaries of results may be generated.

When used for testing purposes, the primary objective of respondent debriefing is to determine whether the respondents understand the concepts and questions in the same way that the survey designers intend. Sufficient information is obtained to evaluate the extent to which reported data are consistent with survey definitions. For instance, respondents may be asked whether they included or excluded particular items in their answers, per definitions. In economic surveys, the debriefings may ask about the use of records or estimation strategies. In addition, respondent debriefings can be useful in determining the reason for respondent misunderstandings. Sometimes results of respondent debriefings show that a question is superfluous and can be eliminated from the final questionnaire. Conversely, it may be discovered that additional questions need to be included in the final questionnaire to better operationalize the concept of interest. Finally, the data may show that the intended meaning of certain concepts or questions is not clear or able to be understood.

A critical requirement to obtain a successful respondent debriefing is that question designers and researchers have a clear idea of potential problems so that good debriefing questions can be developed. Ideas about potential problems can come from pre-field techniques (e.g., cognitive interviews conducted prior to the field test), from analysis of data from a previous survey, from careful review of questionnaires, or from observation of earlier interviews.

Respondent debriefings may be able to supplement the information obtained from behavior coding. As noted above, behavior coding demonstrates the existence of problems but does not always identify the source of the problem. When designed properly, the results of respondent

debriefings can provide information about the sources of problems. Respondent debriefings also may reveal problems not evident from the response behavior.

Interviewer Debriefing has traditionally been the primary method used to evaluate field or pilot tests of interviewer-administered surveys. It also may be used following production data collection prior to redesigning an ongoing periodic or recurring survey. Interviewer debriefing consists of holding group discussions or administering structured questionnaires with the interviewers to obtain their views of questionnaire problems. The objective is to use the interviewers' direct contact with respondents to enrich the questionnaire designer's understanding of questionnaire problems. Although it is a useful evaluation component, it is not sufficient as an evaluation method and does **not** satisfy the pretesting requirement.

Interviewers may not always be accurate reporters of certain types of questionnaire problems for several reasons. When interviewers report a problem, it is not always clear if the issue caused trouble for one respondent or for many. Interviewers' reports of problem questions may reflect their own preference regarding a question, rather than respondent confusion. Finally, 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.

Interviewer debriefings can be conducted in several different ways: in a group setting, through rating forms, or through standardized questionnaires. *Group setting debriefings* are the most common method. They essentially involve conducting a focus group with the field test interviewers to learn about their experiences in administering the questionnaire. *Rating forms* obtain more quantitative information by asking interviewers to rate each question in the pretest questionnaire on selected characteristics of interest to the researchers (e.g., whether the interviewer had trouble reading the question as written, whether the respondent understood the words or ideas in the question). *Standardized interviewer debriefing questionnaires* collect information about the interviewers' perceptions of a problem, the prevalence of a problem, the reasons for a problem, and proposed solutions to a problem. Interviewer debriefings also can ask about the magnitude of specific kinds of problems, to test the interviewers' knowledge of subject-matter concepts.

Analysts' Feedback is a method of learning about problems with a questionnaire specific to the economic area. At the Census Bureau, most economic surveys are self-administered; so survey or program staff analysts in the individual subject areas, rather than interviewers, often have contact with respondents. While collecting feedback from analysts is a useful evaluation component, it does **not** satisfy the pretesting requirement.

Feedback from analysts about their interactions with respondents may serve as an informal evaluation of the questionnaire and the data collected. These interactions include "Help Desk" phone inquiries from respondents and follow-up phone calls to respondents by analysts investigating suspicious data flagged by edit failures. Analyst feedback is more useful when analysts systematically record comments from respondents in a log. The log enables qualitative evaluation of the relative severity of questionnaire problems, because strictly anecdotal feedback sometimes may be overstated.

Another way to obtain analyst feedback is for questionnaire design experts to conduct focus groups with the analysts who review data and resolve edit failures. These focus groups can identify questions that may need to be redesigned or evaluated by other methods. Regardless of how respondent feedback is captured, analysts should provide feedback early in the questionnaire development cycle of recurring surveys to identify problematic questions.

Split Panel Tests are controlled experimental tests of questionnaire variants or data collection modes to determine which one is "better" or to measure differences between them. Split panel testing satisfies the pretesting requirement.

Split panel experiments may be conducted within a field or pilot test or embedded within production data collection for an ongoing periodic or recurring survey. For pretesting draft versions of a questionnaire, the search for the "better" questionnaire requires that an a priori standard be determined by which the different versions can be judged. Split panel tests can incorporate a single question, a set of questions, or an entire questionnaire.

It is important to select adequate sample sizes when designing a split panel test so that differences of substantive interest can be measured. In addition, these tests must use randomized assignment within replicate sample designs so that differences can be attributed to the question or questionnaire and not to the effects of incomparable samples.

Another use of the split panel test is to calibrate the effect of changing questions. Although split panel tests are expensive, they are extremely valuable in the redesign and testing of surveys for which the comparability of the data collected over time is an issue. They provide an important measure of the extent to which different results following a major survey redesign are due to methodological changes, such as the survey instrument or interview mode, rather than changes over time in the subject-matter of interest. Split panel testing is recommended for data with important policy implications.

Comparing response distributions in split panel tests produces measures of differences but does not necessarily reveal whether one version of a question produces a better understanding of what is being asked than another. Other question evaluation methods, such as respondent debriefings, interviewer debriefings, and behavior coding, are useful to evaluate and interpret the differences observed in split panel tests.

Analysis of Item Nonresponse Rates, Imputation Rates, Edit Failures, or Response Distributions from the collected data can provide useful information about how well the questionnaire works. Use of this method in combination with a field test does **not** satisfy the pretesting requirement.

In household surveys, examination of item nonresponse rates can be informative in two ways. First, "don't know" rates can determine the extent to which a task is too difficult for respondents. Second, refusal rates can determine the extent to which respondents find certain questions or versions of a question to be more sensitive than others.

In economic surveys, item nonresponse may be interpreted to have various meanings, depending on the context of the survey. In some institutional surveys (e.g., hospitals, prisons, schools) where data are abstracted from individual person-level records, high item nonresponse is considered to indicate data not routinely available in those records. Item nonresponse may be more difficult to detect in other economic surveys where questions may be left blank because they are not applicable to the responding business or the response value may be zero. In these cases, the data may not be considered missing at all.

Response distributions are the frequencies with which respondents provided answers during data collection. Evaluation of the response distributions for survey items can determine whether variation exists among the responses given by respondents or if different question wordings or question sequencings produce different response patterns. This type of analysis is most useful when pretesting either more than one version of a questionnaire or a single questionnaire for which some known distribution of characteristics exists for comparative purposes.

The quality of collected data also may be evaluated by comparing, reconciling, or benchmarking to data from other sources. This is especially true for economic data, but benchmarking data are also available for some household surveys.

CONCLUSION

At least one of the following techniques must be used to satisfy the pretesting requirement:

- Cognitive interviews.
- Usability techniques focused on the respondent's understanding of the questionnaire.
- Focus groups involving the administration of questionnaires.
- Behavior coding of respondent/interviewer interactions.
- Respondent debriefings in conjunction with a field test or actual data collection.
- Split panel tests.

However, pretesting typically is more effective when multiple methods are used. Additional pretesting techniques should be carefully considered to provide a thorough evaluation and documentation of questionnaire problems and solutions. The relative effectiveness of the various techniques for evaluating survey questions depends on the pretest objectives, sample size, questionnaire design, and mode of data collection. The Census Bureau advocates that both pre-field and field techniques be undertaken, as time and funds permit.

For continuing surveys that have a pre-existing questionnaire, cognitive interviews should be used to provide detailed insights into problems with the questionnaire whenever time permits or when a redesign is undertaken. Cognitive interviews may be more useful than focus groups with a pre-existing questionnaire because they mimic the question-response process. For one-time or new surveys, focus groups are useful tools for learning what respondents think about the concepts, terminology, and sequence of topics prior to drafting the questionnaire. In economic surveys, exploratory/feasibility studies, conducted as company or site visits, also provide information about structuring and wording the questionnaire relative to data available in business/institutional records. Usability techniques are increasingly important as surveys move to automated data collection.

Pre-field methods alone may not be sufficient to test a questionnaire. Some type of testing in the field is encouraged, even if it is evaluated based only on observation by questionnaire developers. More helpful is small-to-medium-scale field or pilot testing with more systematic evaluation techniques. The various methods described in this appendix complement each other in identifying problems, the sources of problems, and potential solutions.

REFERENCES:

Dumas, J. and Redish, J., 1999. *A Practical Guide to Usability Testing*, Portland, OR: Intellect.

Forsyth, B. H., and Lessler, J. T., 1991. "Cognitive Laboratory Methods: A Taxonomy," in Measurement Errors in Surveys, Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., and Sudman, S.(eds.), New York: John Wiley and Sons, Inc., pp. 393-418.

Nielsen, Jakob, 1993. Usability Engineering. Morgan Kaufmann, New York, NY.

Statistical Quality Standard A3
Developing and Implementing a Sample Design

Purpose: The purpose of this standard is to ensure that statistically sound frames are designed and samples are selected to meet the objectives of the survey.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the design and selection of statistically sound samples used to produce estimates or make inferences. This standard covers:

- Frame development for censuses and sample surveys.
- The design and selection of samples or subsamples for surveys.
- The design and selection of samples or subsamples for secondary data analysis, evaluations, or quality assessments.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Selection of focus groups.
- Cognitive interviewing.
- Samples that will not be used to produce estimates or make inferences (e.g., samples used for operational tests, pilot studies, or quality control).
- Frames and samples provided to the Census Bureau by a sponsor.
- Activities performed to produce sample estimates (e.g., weighting, estimation, and variance estimation). Statistical Quality Standard D1, *Producing Direct Estimates from Samples*, addresses requirements related to producing estimates.

Key Terms: Cluster, coverage, cut-off samples, estimate, estimation, frame, housing unit, peer review, precision, primary sampling unit (PSU), probability of selection, probability sampling, sample design, sample size, sampling frame, sampling weights, sequential sampling, strata, stratification, systematic sampling, target population, unduplication, variance, and weights.

Requirement A3-1: Throughout all processes associated with frame development and sample design, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement A3-2: A plan must be developed that addresses:

1. Statistical requirements of the program using the sample (e.g., the target population, the key estimates, the required precision of the estimates, and the expected response rates).
2. Development of the sampling frame.
3. Sampling methodologies that improve efficiency and minimize the costs of data collection (e.g., probability sampling, oversampling, stratification, sorting, unduplication requirements, and cluster sizes).
4. Verification and testing of systems associated with the sampling operations.
5. Monitoring and evaluating the accuracy of the frame and the sample (e.g., the coverage of the target population by the frames, timeliness of the frames, efficiency of stratification, and verification of the sample).

Notes:

(1) The Census Bureau Guideline *Sample Design and Selection* identifies steps to follow and issues to consider when designing and selecting a sample.

(2) Statistical Quality Standard A1, *Planning the Data Program*, addresses overall planning requirements, including the development of schedules and costs.

Requirement A3-3: Sampling frames that meet the data collection objectives must be developed using statistically sound methods.

Examples of frame development activities include:

- Describing the target population.
- Constructing the frame using sources that promote accuracy and completeness.
- Combining multiple frames and unduplicating among them or adjusting probabilities of selection to address units appearing in multiple frames.
- Updating frames (e.g., for new construction and establishment "births" and "deaths").
- Identifying limitations of the frame, including timeliness and accuracy of the frame (e.g., misclassification, eligibility, and coverage).

Requirement A3-4: The sample design must be developed to meet the objectives of the survey, using statistically sound methods. The size and design of the sample must reflect the level of detail needed in tabulations and other information products and the precision required of key estimates. Any use of nonprobability sampling methods (e.g., cut-off) must be justified statistically.

Examples of sample design activities include:

- Setting the requirements and rules for how to define primary sampling units (PSUs), secondary units (e.g., clusters of housing units), and criteria for self-representing PSUs.
- Defining measures of size.
- Determining whether oversampling of population subgroups is needed.
- Defining sampling strata and criteria for clustering.
- Defining the sample size by stratum and the allocation methodology.
- Determining the order of selection and the probabilities of selection.
- Describing the sample selection methods (e.g., systematic sampling, sequential sampling, and probability proportional to size).

- Grouping sample units into representative panels and identifying the duration a unit will remain in sample.
- Determining sample rotation patterns.
- Addressing the issues involved with replacing a current sample design with a new one (e.g., phase-in/phase-out periods, minimizing/maximizing overlap, and accounting for any bias associated with the redesign).
- Developing and maintaining sample design information needed for weighting, estimation, and variance estimation (e.g., probabilities of selection, noninterview adjustment cells, and sample replicates).
- Assessing the potential bias from using the cut-off sampling method.

Requirement A3-5: Sampling frames must be implemented and samples selected to ensure high quality data.

Sub-Requirement A3-5.1: Specifications and procedures for creating frames and selecting samples, based on the statistical requirements, must be developed and implemented.

Examples of issues that specifications and procedures might address include:
- Stratum definitions, stratification algorithms, and clustering algorithms.
- Addition or deletion of records to update frames.
- Algorithms for creating PSUs.
- Sampling algorithms.
- Unduplication of the sample between surveys or between different waves of the same survey.
- Creation of sample replicates needed for weighting, estimation, and variance estimation.
- Assignment of sampling weights appropriate for the sample design to selected units.

Sub-Requirement A3-5.2: Systems and procedures must be verified and tested to ensure all components function as intended.

Examples of verification and testing activities include:
- Verifying that specifications conform to the technical requirements for the frame and sample design (e.g., using walk-throughs and peer reviews).
- Validating computer code against specifications.
- Performing tests of the individual modules and an integrated test of the full sample selection operation.
- Verifying the accuracy of frame information.
- Verifying the selection of the sample for accuracy (e.g., sample sizes are as expected).

Sub-Requirement A3-5.3: Systems and procedures must be developed and implemented to monitor and evaluate the accuracy of the frame development and sample selection operations and to take corrective action if problems are identified.

Examples of activities to monitor and evaluate the accuracy include:
- Comparing weighted sample counts with frame counts.

- Verifying that sample sizes are within expectations.
- Evaluating the accuracy and coverage of the frames against the target population.
- Evaluating changes in the sample design to understand how the revisions might affect the estimates.

Requirement A3-6: Documentation needed to replicate and evaluate frame development and sample design operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the systems and processes of frame development and sample selection.
- Sampling design information needed to produce estimates and variance estimates.
- Descriptions of the frame and its coverage.
- Techniques used to evaluate the coverage of the frame and the adequacy of the sample design.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

COLLECTING AND ACQUIRING DATA

Statistical Quality Standard B1
Establishing and Implementing Data Collection Methods

Purpose: The purpose of this standard is to ensure that methods are established and implemented to promote the collection of high quality data from respondents.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to establishing and implementing data collection methods for data programs that obtain information directly from respondents, including reimbursable surveys and surveys in which interviewers collect information from establishments.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Administrative records data acquired under agreements with other organizations and not collected by interviewers.

Key Terms: CAPI, CATI, coverage, data collection, dress rehearsal, fax imaging, field test, load testing, mail-out/mail-back, measurement error, nonresponse bias, nonresponse follow-up, reinterview, response error, response rate, supplemental reinterview, systems test, and touch-tone data entry (TDE).

Requirement B1-1: Throughout all processes associated with data collection, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement B1-2: A plan must be developed that addresses:

1. Data collection methods (e.g., interview mode, use of incentives, and reference periods), systems, and procedures.
2. Definitions for what constitutes an interview or response (i.e., a complete interview, a sufficient partial interview, or an insufficient partial interview).
3. Verification and testing of the data collection methods, systems, and procedures. (Statistical Quality Standard A2, *Developing Data Collection Instruments and Supporting Materials*, addresses questionnaire content pretesting and instrument testing.)
4. Training for staff involved in the data collection effort.
5. Monitoring and evaluating the quality of the data collection operations.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

Requirement B1-3: Data collection methods must be designed and implemented in a manner that balances (within the constraints of budget, resources, and time) data quality and measurement error with respondent burden.

Sub-Requirement B1-3.1: Systems and procedures must be developed to implement the data collection.

Examples of data collection activities for which systems and procedures should be developed include:

- Listing possible sampling units.
- Producing paper questionnaires and related materials (e.g., printing and assembling mail-out packages). (Statistical Quality Standard A2, *Developing Data Collection Instruments and Supporting Materials*, addresses the design of questionnaires and materials.)
- Providing OMB-required notifications to respondents.
- Providing telephone questionnaire assistance for mail-out/mail-back data collection.
- Transmitting information (by mail, electronically, the Internet, TDE, fax imaging, or other method) between respondents or interviewers and the Census Bureau.
- Formatting CAPI/CATI output files to be compatible with processing systems.
- Conducting interviews.
- Conducting nonresponse follow-up operations.

Sub-Requirement B1-3.2: Data collection systems and methods must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Verifying that the specifications and procedures reflect the requirements of the program.
- Verifying that the materials used for data collection operations meet specifications (e.g., ensure that forms are printed properly).
- Verifying the physical assembly of mailing packages (e.g., ensure that mailing pieces fit properly in the envelopes).
- Testing the electronic data management systems (e.g., the systems used to manage cases and data between headquarters and the interviewers and between headquarters and the data processing systems) for accuracy, capacity (e.g., load testing), and reliability.
- Conducting a systems test to verify the functioning of the data collection instrument in combination with the data management systems.
- Conducting a field test to test systems and methods under realistic conditions (e.g., the dress rehearsal for the decennial census).

Sub-Requirement B1-3.3: Training for field and headquarters staff involved in the data collection effort (as identified during planning) must be developed and provided.

Examples of training topics include:

- Relevant Census Bureau policies (e.g., Data Stewardship Policy DS016, *Respondent Identification Policy*).
- The goals and objectives of the data collection.
- Survey specific concepts and definitions.
- The uses of the data.
- Techniques for obtaining respondent cooperation.
- Follow-up skills.

Sub-Requirement B1-3.4: Systems and procedures must be developed and implemented to monitor and evaluate the data collection activities and to take corrective actions if problems are identified.

Examples of monitoring and evaluating activities include:

- Tracking unit response rates, progress in completing interviews, and costs of the data collection, and taking corrective action when goals are not met.
- Tracking returned cases to ensure all cases are accounted for and investigating to locate missing cases.
- Verifying that interviewers follow interviewing procedures and do not falsify data (e.g., by conducting field observations, conducting reinterviews, or monitoring telephone center interviewers) and, if necessary, taking appropriate corrective action (e.g., retraining, reassigning, or dismissing interviewers).
- Collecting, tracking, and analyzing interviewer performance statistics (e.g., refusals, completed interviews, refusal conversions, login hours, and completed interviews per login hour), and providing feedback or other corrective action when necessary.
- Verifying that analysts follow data collection review procedures, and providing feedback when necessary.
- Reviewing response data for accuracy and completeness, and taking appropriate corrective action when necessary to improve accuracy or completeness.
- Reviewing response data for unexpected results and unusual patterns (e.g., a pattern of an unusually high number of vacant households) and taking corrective action when needed (e.g., providing feedback, retraining interviewers, or conducting supplemental reinterviews).
- Conducting evaluation studies (e.g., nonresponse bias analysis, coverage evaluation study, and response error reinterview study).

Requirement B1-4: Documentation needed to replicate and evaluate the data collection methods must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the data collection.
- Test designs and results.
- Instructions to respondents and interviewers about the data collection instrument.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard B2
Acquiring and Using Administrative Records

Purpose: The purpose of this standard is to ensure the quality of information products derived from administrative records data acquired from non-Census Bureau organizations.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the acquisition and use of administrative records data (e.g., demographic, business, and geographic administrative records data), from non-Census Bureau organizations.

Exclusions:
The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Key Terms: Administrative records, data-use agreement, and record linkage.

Requirement B2-1: Throughout all processes associated with acquiring, using, and disposing of administrative records data, the provisions of federal laws (e.g., Title 13, Title 15, and Title 26), data-use agreements, and Census Bureau policies and procedures on privacy and confidentiality (e.g., Data Stewardship Policies) must be followed to protect administrative records data from unauthorized release. (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Note: For detailed procedures on acquiring, using, and disposing of administrative records data, see the Administrative Records Handbook.

Requirement B2-2: A study plan must be developed that addresses verification and evaluation of the quality of the acquired data, in addition to the requirements of the Administrative Records Handbook.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses the overall planning requirements for a data program, including estimates of schedule and costs.

Requirement B2-3: Acquired data must be reviewed to ensure that they meet the requirements specified in the data-use agreement and in the technical documentation provided by the source agency.

Examples of review activities include:
- Verifying that the data are readable and match the record layout.

- Verifying that the number of records is consistent with counts provided by the source agency.
- Comparing distributions of variables with historical averages or expected values.
- Reviewing address lists for extraneous characters and to ensure that the format of incoming information is consistent with information contained within Census Bureau databases.

Sub-Requirement B2-3.1: The quality of the acquired data must be evaluated.

Examples of evaluation activities include:
- Calculating the missing data rates within the records.
- Calculating coverage rates.
- Evaluating the accuracy of the records (e.g., values of variables are within acceptable ranges).

Sub-Requirement B2-3.2: If the data do not meet the requirements, timely feedback on the problems must be provided and corrective actions taken, following the procedures described in the Administrative Records Handbook.

Requirement B2-4: Documentation needed to replicate and evaluate administrative records projects must be produced. The documentation must be retained, to the extent allowed by applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation, in addition to the documentation specified by the Administrative Records Handbook, include:
- Descriptions of processes and procedures used to verify the data and evaluate its quality.
- Descriptions of processes and procedures used to develop estimates.
- Research reports used to guide decisions.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:
(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)
(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public ensure transparency of information products released by the Census Bureau.

CAPTURING AND PROCESSING DATA

Statistical Quality Standard C1
Capturing Data

Purpose: The purpose of this standard is to ensure that methods are established and implemented to promote the accurate capture and conversion of paper forms or image files into data files for further processing.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status (SSS) individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the development, modification, and implementation of post-collection data capture operations, such as:

- Operations to convert data on paper forms or maps into data files (e.g., key from paper (KFP) data entry, optical mark recognition (OMR), and optical character recognition (OCR)).
- Operations to convert image files (e.g., fax image files received directly from respondents and geographic image files) into data files (e.g., key from image (KFI) data entry, the Economic Programs' Paperless Fax Imaging Retrieval System (PFIRS), and operations to convert geographic image files into data files).

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Electronic data collections (e.g., CATI, CAPI, and the Web). Statistical Quality Standard A2, *Developing a Data Collection Instrument,* addresses data capture performed within an instrument during data collection.

Key Terms: Data capture, key from image (KFI), key from paper (KFP), optical character recognition (OCR), and optical mark recognition (OMR).

Requirement C1-1: Throughout all processes associated with data capture, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement C1-2: A plan must be developed that addresses:

1. Requirements for the data capture systems.
2. Required accuracy levels for data capture.
3. Verification and testing of the data capture systems.
4. Training for the staff who perform the data capture operations (including SSS contractors).

5. Monitoring and evaluation of the quality of the data capture operations.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

Requirement C1-3: Data collected on paper forms or in image files must be converted accurately into an electronic format suitable for subsequent processing.

Sub-Requirement C1-3.1: Specifications and procedures for the data capture operations must be developed and implemented.

Examples of activities that specifications and procedures might address include:

- KFP data entry.
- Scanning systems for paper forms and maps (e.g., OMR and OCR).
- Operations to convert image files (e.g., fax image files and geographic image files) into data files (e.g., KFI data entry and PFIRS).

Sub-Requirement C1-3.2: Data capture systems and procedures must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Verifying that data capture specifications reflect the system requirements.
- Verifying that data capture systems and software adhere to the specifications.
- Verifying that data capture systems and software capture data accurately.

Sub-Requirement C1-3.3: Training for the staff (including SSS contractors) who perform the data capture operations (as identified during planning) must be developed and provided.

Sub-Requirement C1-3.4: Systems and procedures must be developed and implemented to monitor and evaluate the quality of the data capture operations and to take corrective actions if problems are identified.

Examples of monitoring and evaluation activities include:

- Monitoring captured data (keyed or captured through an automated system) to ensure that it meets the specified accuracy requirements.
- Monitoring and documenting the frequency and types of errors.
- Taking corrective actions when data do not meet accuracy requirements (e.g., rejecting and repairing unacceptable batches, retraining key-entry staff, and adjusting automated systems and retesting).

Requirement C1-4: Documentation needed to replicate and evaluate the data capture operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the data capture system.
- Problems encountered and solutions implemented during the data capture operations.
- Quality measures from monitoring and evaluating the data capture operations (e.g., error rates). (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard C2
Editing and Imputing Data

Purpose: The purpose of this standard is to ensure that methods are established and implemented to promote the accurate correction of missing and erroneous values in survey, census, and administrative records data through editing and imputation.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the development and implementation of editing and imputation operations for survey, census, administrative records data, and geospatial data.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Estimation methods, such as nonresponse adjustments, that compensate for missing data. Statistical Quality Standard D1, *Providing Direct Estimates from Samples*, addresses requirements for estimation methods.

Key Terms: Editing, imputation, outliers, skip pattern, and truth deck.

Requirement C2-1: Throughout all processes associated with editing and imputation, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement C2-2: A plan must be developed that addresses:

1. Requirements for the editing and imputation systems.
2. Verification and testing of the editing and imputation systems.
3. Monitoring and evaluation of the quality of the editing and imputation operations.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

Requirement C2-3: Data must be edited and imputed using statistically sound practices, based on available information.

Sub-Requirement C2-3.1: Specifications and procedures for the editing and imputation operations must be developed and implemented to detect and correct errors or missing data in the files.

Examples of issues that specifications and procedures might address include:

- Checks of data files for missing data, duplicate records, and outliers (e.g., checks for possible erroneous extreme responses in income, price, and other such variables).
- Checks to verify the correct flow through prescribed skip patterns.
- Range checks or validity checks (e.g., to determine if numeric data fall within a prespecified range or if discrete data values fall within the set of acceptable responses).
- Consistency checks across variables within individual records to ensure non-contradictory responses (e.g., if a respondent is recorded as 5 years old and married, the record contains an error).
- Longitudinal consistency checks for data fields not measuring period to period changes.
- Editing and imputation methods and rules (e.g., internal consistency edits, longitudinal edits, hot deck edits, and analyst corrections).
- Addition of flags on the data files to clearly identify all imputed and assigned values and the imputation method(s) used.
- Retention of the unedited values in the file along with the edited or imputed values.
- Checks for topology errors in geospatial data (e.g., lack of coincidence between boundaries that should align, gaps, overshoots, and floating segments).
- Checks for address range errors in geographic data (e.g., parity inconsistencies, address range overlaps and duplicates, and address range direction irregularities).
- Checks for duplicate map features.
- Standardization of street name information in geographic data (e.g., consistency of abbreviations and directionals, and consistent formatting).
- Rules for when data not from the data collection qualify as "equivalent-quality-to-reported-data" for establishment data collections.

Sub-Requirement C2-3.2: Editing and imputation systems and procedures must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Verifying that edit and imputation specifications reflect the requirements for the edit and imputation systems.
- Validating edit and imputation instructions or programming statements against specifications.
- Verifying that the imputation process is working correctly using test files.
- Verifying that edit and imputation outcomes comply with the specifications.
- Verifying that edit and imputation rules are implemented consistently.
- Verify that the editing and imputation outcomes are consistent within records and consistent across the full file.
- Verifying that the editing and imputation outcomes that do not use randomization are repeatable.

Sub-Requirement C2-3.3: Systems and procedures must be developed and implemented to monitor and evaluate the quality of the editing and imputation operations and to take corrective actions if problems are identified.

Examples of monitoring and evaluation activities include:

- Monitoring and documenting the distributions of, and reasons for, edit and imputation changes to determine if corrections are needed in the system.
- Evaluating and documenting editing results for geospatial files (e.g., edits resulting in improvements in boundaries, feature coverage, and feature accuracy) and geographic files (e.g., address ranges, address parity, and geographic entity names and codes).
- Reviewing and verifying data when edits produce results that differ from the past.
- Using a truth deck to evaluate the accuracy of the imputed values.

Requirement C2-4: Documentation needed to replicate and evaluate the editing and imputation operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the editing and imputation systems, including edit rules.
- Distributions of changes from edits and imputations.
- Retaining original responses (before edit/imputation) on data files along with the final edited/imputed responses.
- Problems encountered and solutions implemented during the editing and imputing operations.
- Quality measures from monitoring and evaluating the editing and imputation operations (e.g., imputation rates and edit change rates). (See Statistical Quality Standard D3, *Providing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard C3
Coding Data

Purpose: The purpose of this standard is to ensure that methods are established and implemented to promote the accurate assignment of codes, including geographic entity codes, to enable analysis and tabulation of data.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the development and implementation of post-collection coding operations, including the assignment of:

- Codes to convert text and numerical data into categories.
- Geographic entity codes (geocodes) and geographic attribute codes to distinguish and describe geographic entities and their characteristics within digital databases.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Behavior coding activities associated with cognitive interviewing.

Key Terms: American National Standards Institute codes (ANSI codes), coding, geocoding, geographic entity code (geocode), Master Address File (MAF), North American Industry Classification System (NAICS), Standard Occupational Classification System (SOC), and Topologically Integrated Geographic Encoding and Referencing (TIGER).

Requirement C3-1: Throughout all processes associated with coding, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement C3-2: A plan must be developed that addresses:

1. Required accuracy levels for the coding operations, including definitions of errors.
2. Requirements for the coding systems, including requirements for input and output files.
3. Verification and testing of the coding systems.
4. Training for staff involved in the clerical coding operations.
5. Monitoring and evaluation of the quality of the coding operations.

Notes:
(1) Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

(2) The Census Bureau Guideline, *Coding Verification*, provides guidance on coding procedures.

Requirement C3-3: Processes must be developed and implemented to accurately assign codes for converting text and numerical data to categories and geocodes to identify and distinguish geographic entities and their attributes within a digital database.

Sub-Requirement C3-3.1: Specifications and procedures for the coding systems and operations must be developed and implemented.

Examples of issues that coding specifications and procedures might address include:

- A list and description of the admissible codes or values for each item on the questionnaire.
- A list of acceptable reference sources, printed and electronic, that may be used by the coding staff (e.g., Employer Name List).
- Procedures to add to the list of admissible codes or to add text responses to match existing codes.
- Consistency of codes across data collection periods.
- Procedures to assign and associate geocodes with other information within geographic files (e.g., the Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) database).

Sub-Requirement C3-3.2: Standardized codes, when appropriate, must be used to convert text data.

Examples of current coding standards include:

- American National Standards Institute (ANSI) Codes.
- North American Industry Classification System (NAICS).
- Standard Occupational Classification System (SOC).

Sub-Requirement C3-3.3: Coding systems must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Verifying that coding specifications and procedures satisfy the coding requirements.
- Validating coding instructions or programming statements against specifications.
- Verifying that coding rules are implemented consistently.
- Using a test file to ensure that the codes are assigned correctly.

Sub-Requirement C3-3.4: Training for staff involved in clerical coding operations (as identified during planning) must be developed and provided.

Sub-Requirement C3-3.5: Systems and procedures must be developed and implemented to monitor and evaluate the quality of the coding operations and to take corrective actions if problems are identified.

Examples of monitoring and evaluation activities include:

- Establishing a quality control (QC) system to check coding outcomes and providing feedback to coders or taking other corrective action.
- Monitoring QC results (such as referral rates, error rates), determining the causes of systematic errors, and taking corrective action (e.g., providing feedback or retraining to coders, updating coder reference materials, or other corrective actions).
- Incorporating a geocode verification within automated instruments and correcting geocodes when errors are detected.
- Evaluating the accuracy of geocoding and determining the cause of errors in incorrect geocodes.
- Reviewing and updating coding guidelines.
- Reviewing software and procedures to reflect any changes in the coding guidelines.

Requirement C3-4: Documentation needed to replicate and evaluate the coding operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the coding systems.
- Problems encountered and solutions implemented during the coding operations.
- Quality measures from monitoring and evaluating the coding operations (e.g., error rates and referral rates). (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard C4
Linking Data Records

Purpose: The purpose of this standard is to ensure that methods are established and implemented to promote the accurate linking of data records.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to both automated and clerical record linkage used for statistical purposes. It covers linking that uses characteristics of an entity to determine whether multiple records refer to the same entity.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Statistical attribute matching.
- Linkages performed using only a unique identifier (e.g., Protected Information Key or serial number).
- Linkages performed for quality assurance purposes.

Key Terms: Automated record linkage, blocking, clerical record linkage, field follow-up, record linkage, scoring weights, and statistical attribute matching.

Requirement C4-1: Throughout all processes associated with linking, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement C4-2: A plan must be developed that addresses:

1. Objectives for linking the files.
2. Data sets and files to be linked.
3. Verification and testing of the linking systems and processes.
4. Training for staff involved in the clerical record linkage operations.
5. Evaluation of the results of the linkage (e.g., link rates and clerical error rates).

Notes:
(1) Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

(2) The Data Stewardship Policy DS014, *Record Linkage*, states the principles that must be met for record linkage activities and a checklist that must be filled out before beginning record linkage activities.
(3) The Census Bureau Guideline *Record Linkage* provides guidance on procedures for automated and clerical record linkage.

Requirement C4-3: Record linkage processes must be developed and implemented to link data records accurately.

Sub-Requirement C4-3.1: Specifications and procedures for the record linkage systems must be developed and implemented.

Examples of issues that specifications and procedures for automated record linkage systems might address include:

- Criteria for determining a valid link.
- Linking parameters (e.g., scoring weights and the associated cut-offs).
- Blocking and linking variables.
- Standardization of the variables used in linking (e.g., state codes and geographic entity names are in the same format on the files being linked).

Examples of issues that specifications and procedures for clerical record linkage systems might address include:

- Criteria for determining that two records represent the same entity.
- Criteria for assigning records to a specific geographic entity or entities (i.e., geocoding).
- Linking variables.
- Guidelines for situations requiring referrals.
- Criteria for sending cases to field follow-up.

Sub-Requirement C4-3.2: Record linkage systems must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities for automated record linkage systems include:

- Verifying that the specifications reflect system requirements.
- Verifying that the systems and software implement the specifications accurately.
- Performing a test linkage to ensure systems work as specified.

Examples of verification and testing activities for clerical record linkage systems include:

- Verifying that the specifications reflect system requirements.
- Verifying that the instructions will accomplish what is expected.
- Testing computer systems that support clerical linking operations.

Sub-Requirement C4-3.3: Training for the staff involved in clerical record linkage (as identified during planning) must be developed and provided.

Examples of training activities include:

- Instructing clerks on how to implement the specifications.
- Providing a training database to give clerks a chance to practice their skills.
- Assessing error rates of clerks and providing feedback.

Sub-Requirement C4-3.4: Systems and procedures must be developed and implemented to monitor and evaluate the accuracy of the record linkage operations and to take corrective actions if problems are identified.

Examples of monitoring and evaluation activities for automated record linkage operations include:

- Evaluating the accuracy of automated linkages by a manual review.
- Monitoring link rates and investigating deviations from historical results, and taking corrective action if necessary.

Examples of monitoring and evaluation activities for clerical record linkage operations include:

- Establishing an acceptable error rate.
- Establishing quality control sampling rates.
- Monitoring clerks' error rates and referrals, and taking corrective action if necessary (e.g., feedback or retraining).

Requirement C4-4: Documentation needed to replicate and evaluate the linking operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the record linkage systems.
- Programs and parameters used for linking.
- Problems encountered and solutions implemented during the linking operations.
- Evaluation results (e.g., link rates and clerical error rates).

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

PRODUCING ESTIMATES AND MEASURES

Statistical Quality Standard D1
Producing Direct Estimates from Samples

Purpose: The purpose of this standard is to ensure that statistically sound practices are used for producing direct estimates from samples for information products.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the production of direct estimates from samples and estimates of their variances for Census Bureau information products. The standard applies to estimates derived from:

- Samples selected for surveys or the Economic Census.
- Samples or subsamples selected for data analyses, evaluations, or quality assessments of surveys, censuses, or programs using administrative records.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- 100 percent enumerations.
- Activities related to producing estimates from models. (See Statistical Quality Standard D2, *Producing Estimates from Models*.)

Key Terms: Calibration, coefficient of variation (CV), coverage error, cross-sectional studies, direct estimates, estimation, generalized variance function, imputation, longitudinal studies, post-stratification, raking, ratio estimation, replication methods, sanitized data, and Taylor series method for variance estimation.

Requirement D1-1: Throughout all processes associated with estimation, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement D1-2: A plan must be developed that addresses:

1. Key estimates that will be produced.
2. Estimation methodologies (e.g., population controls, post-stratification, nonresponse adjustments, ratio estimation, calibration, and raking).
3. Variance estimation methodologies (e.g., sampling formula variances, Taylor series (linearization) methods, replication methods, and generalized variance functions).
4. Verification and testing of the systems for generating estimates.
5. Verification of the estimates and evaluation of their quality.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements and development of schedules and costs.

Requirement D1-3: Estimates and their variances must be produced using statistically sound practices that account for the sample design and reduce the effects of nonresponse and coverage error.

Examples of statistically sound practices include:

- Calculating estimates and variances in ways that take into account the probabilities of selection, stratification, and clustering.
- Developing generalized variance formulas for computing variances.
- Using auxiliary data or performing post-sampling adjustments to improve the precision and the accuracy of estimates (e.g., ratio or raking weighting adjustments for unit nonresponse and post-stratification).
- Accounting for post-sampling adjustments when computing variances (e.g., imputation effects in variance estimates).
- Generating weights or adjustment factors to allow both cross-sectional and longitudinal estimates for longitudinal surveys.

Note: Statistical Quality Standard A3, *Developing and Implementing a Sample Design*, specifies requirements for the design and selection of probability samples used to produce estimates or make inferences.

Sub-Requirement D1-3.1: Specifications for the estimation systems must be developed and implemented.

Examples of issues that specifications might address include:

- Methodological requirements for generating the estimates and variances.
- Data files used or saved during the estimation process (e.g., files used for program validation, verification, and research).

Sub-Requirement D1-3.2: Estimation systems must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Verifying that specifications conform to the estimation methodologies.
- Validating computer code against specifications.
- Verifying that the estimates are computed according to the specifications.
- Using subject matter and statistical experts to review the estimation methodology.
- Conducting peer reviews (e.g., reviews of specifications, design documents, and programming code).
- Conducting verification and validation tests.
- Conducting internal user acceptance tests for estimation software.

Sub-Requirement D1-3.3: Methods and systems must be developed and implemented to verify the estimates and evaluate their quality.

Examples of verification and evaluation activities include:

- Comparing current estimates against historical results.
- Comparing the estimates derived from the survey to other independent collections of similar data.
- Comparing coefficients of variation (CVs) or variances of the estimates against historical results.
- Examining relationships among the estimates.
- Conducting studies to evaluate the performance of variance estimates.

Note: Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*, provides requirements for measuring and evaluating nonsampling error.

Requirement D1-4: Documentation needed to replicate and evaluate the estimation operations must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the estimation systems.
- Final weighting specifications, including calculations for how the final sample weights are derived.
- Final variance estimation specifications.
- Computer source code.
- Data files with weighted data and any design parameters that would be needed to replicate estimates and variances.
- Methodological documentation.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard D2
Producing Estimates from Models

Purpose: The purpose of this standard is to ensure that statistically sound practices are used to generate estimates from models for information products.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the production of estimates from models for Census Bureau information products. This standard applies to models (e.g., regression, economic, and log-linear) used to produce estimates, such as:

- Small domain estimates, including small area estimates.
- Demographic estimates and projections.
- Seasonal adjustment of estimates.
- Census coverage estimates.
- Synthetic data to protect microdata from disclosure.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Models that are not used to produce estimates for Census Bureau information products (e.g., models used for imputation or disclosure avoidance which are addressed in Statistical Quality Standard C2, *Editing and Imputing Data*, and Statistical Quality Standard S1, *Protecting Confidentiality*, respectively).

Key Terms: Autocorrelation function, autoregressive integrated moving average (ARIMA), cross-validation, goodness-of-fit, heteroscedastic, homoscedastic, model, model validation, Monte Carlo simulation, multicollinearity, projection, regression, revisions history, residual, sanitized data, seasonal adjustment, sensitivity analysis, sliding spans, small area estimation, and spectral graphs.

Requirement D2-1: Throughout all processes associated with estimation, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement D2-2: A plan must be developed that addresses:

1. Purpose and rationale for using a model (e.g., data to compute precise estimates are not available, or modeling with additional data will provide more accuracy).
2. Key estimates that will be generated and the domain of application for the model.

3. Methodologies and assumptions related to the model, such as the:
    a. Model structure (e.g., functional form, variables and parameters, error structure, and domain of interest).
    b. Model estimation procedure (e.g., least squares estimation, maximum likelihood estimation, and demographic estimation methods).
    c. Data source and how the data will be used in the model, including key modifications to the data.
4. Criteria for assessing the model fit (e.g., goodness-of-fit statistics and R-squared) and the model specification (e.g., measures of multicollinearity).
5. Verification and testing of the systems for generating estimates.
6. Verification of the modeled estimates and evaluation of their quality.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

Requirement D2-3: Models must be developed and implemented using statistically sound practices.

Examples of statistically sound model development practices include:

- Ensuring definitions of variables are accurate (e.g., definitions of the geographic areas used in the model, and eligibility criteria in administrative records).
- Specifying a model that has a basis in verified empirical relationships.
- Examining preliminary model results for internal consistency and to ensure that logical relationships among the data are maintained (e.g., population estimates are not negative, and sub-domains (e.g., counties) sum to super-domains (e.g., states)).
- Estimating measures of statistical uncertainty (e.g., prediction error variances, measures of error associated with using synthetic data, or the Bayesian equivalents of these measures).
- Modifying the functional form, the variables, or the parameters of the model to address problems revealed by the model diagnostics and error estimates.
- Having experts perform a methodological review.
- Producing estimates using weighted data, when appropriate.
- Providing justification that the sample design and selection are adequately accounted for in the estimation process.

Examples of statistically sound practices for demographic estimates and projections include:

- Basing assumptions about future relationships among variables on empirical data or on assumptions that are considered statistically sound.
- Comparing raked and unraked data to ensure logical relationships are maintained.
- Providing quantitative or qualitative assessments of uncertainty for each estimated or projected data point, whenever possible.

Examples of statistically sound practices for seasonal adjustments include:

- Before the first seasonal adjustment of a series, conducting a seasonal analysis to determine whether seasonal patterns exist and periodically repeating the analysis.
- Seasonally adjusting a time series only when data exhibit seasonal patterns.

- Seasonally adjusting only those component series that show identifiable seasonality for aggregate series derived from the combination of component series.
- Using autoregressive integrated moving average (ARIMA) extrapolations in calculating seasonal factors (e.g., the X-12-ARIMA method).
- Reviewing appropriate modeling and seasonal adjustment diagnostics (e.g., revisions history, spectral graphs, plots of the sample autocorrelation function of the model residuals, forecast performance, and sliding spans) for valuable information about model adequacy and adjustment stability.

Sub-Requirement D2-3.1: Model results must be evaluated and validated, and the results of the evaluation and validation must be documented.

Examples of evaluation and validation activities include:

- Validating the model by comparing with independent information sources.
- Generating and reviewing goodness-of-fit statistics (e.g., R-squared and F-tests).
- Generating and reviewing model diagnostics and graphical output (e.g., reviewing for outliers, multicollinearity, heteroscedasticity, homoscedasticity, and influential observations).
- Cross-validating the model using a subset of data withheld from the model fitting.
- Conducting sensitivity analyses to violations of the assumptions (e.g., Monte Carlo simulations).

Note: Evaluation and validation is required when the model is developed. Models used in a continuing production setting must be re-evaluated periodically as appropriate.

Sub-Requirement D2-3.2: Specifications for the modeling and estimation systems must be developed and implemented.

Examples of issues that specifications might address include:

- Descriptions of data files to be used in the model.
- Equations for computing estimates and variances.
- Instructions for running production software.
- Estimation algorithms.
- Convergence criteria for iterative models.

Sub-Requirement D2-3.3: Estimation systems must be verified and tested to ensure that all components function as intended.

Examples of verification and testing activities include:

- Using subject matter and statistical experts to review the estimation methodology.
- Checking that the appropriate equations were used.
- Verifying that the specifications reflect requirements.
- Validating computer code against specifications.
- Assessing computer code to ensure that the appropriate data and variables are used and the code is correctly programmed.

- Performing test runs and debugging computer code.
- Using different random starts to ensure models using maximum likelihood estimates converge consistently.

Sub-Requirement D2-3.4: Methods and systems must be developed and implemented to verify the modeled estimates and evaluate their quality.

Examples of verification and evaluation activities include:

- Performing sensitivity analyses using alternative assumptions to inform users of model stability.
- Examining measures of statistical uncertainty.
- Ensuring that variances reflect both sampling error and modeling error.
- Comparing production estimates against comparable data from other sources, including previous estimates for the program or projections from prior cycles.
- Reviewing goodness-of-fit statistics and model diagnostics and documenting unexpected results to aid the revision of the model for the next cycle.
- Reviewing (during each seasonal adjustment run) newly identified outliers and changes to previously identified extreme values that may cause large revisions in the seasonally adjusted series.

Note: Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*, provides requirements for measuring and evaluating nonsampling error.

Sub-Requirement D2-3.4.1: The seasonal adjustment process and results must be reviewed annually by the program manager (or the appropriate mathematical statistician) to identify needed changes in the X-12-ARIMA specification files. Using the required secure data transmission protocols, the program manager (or the appropriate mathematical statistician) must provide the following to the Time Series Methods Staff (TSMS) of the Office of Statistical Methods and Research for Economic Programs (OSMREP):

1. The new final X-12-ARIMA specification files and the data used.
2. The revised X-12-ARIMA specification file and the data used, whenever the seasonal adjustment options must be changed outside of the annual review period. This information must be provided immediately after release of the adjusted data.

Sub-Requirement D2-3.4.2: For indicator releases, any routine revisions to the annual review process, such as benchmarking and updating of seasonality factors, must be consolidated and released simultaneously. See Statistical Policy Directive No. 3. Deviations from this requirement must be approved as specified in the directive.

Requirement D2-4: Documentation needed to replicate and evaluate the modeling activities must be produced. The documentation must be retained, consistent with applicable policies and data use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the estimation systems.

- Data files with weighted and unweighted data.
- Computer source code.
- Results of outlier analyses, including information on cause of outliers, if available.
- Results of model diagnostics.
- Output data file with "predicted" results for every unit of analysis.
- Seasonal adjustment diagnostic measures (e.g., revisions history values and graphs, spectral graphs, forecast error values and graphs, and sliding spans results).
- Error estimates, parameter estimates, and overall performance statistics (e.g., goodness-of-fit and other such statistics).
- Methodologies used to improve the estimates.
- Quality measures and evaluation results. (See Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*.)

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

.

Statistical Quality Standard D3
Producing Measures and Indicators of Nonsampling Error

Purpose: The purpose of this standard is to ensure that measures and indicators of nonsampling error are computed and documented to allow users to interpret the results in information products, to provide transparency regarding the quality of the data, and to guide improvements to the program.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to activities associated with producing measures or indicators of nonsampling error associated with estimates for Census Bureau information products. Examples of nonsampling error sources include:

- Nonresponse (e.g., bias from household/establishment nonresponse, person nonresponse, and item nonresponse).
- Coverage (e.g., listing error, duplicates, undercoverage, overcoverage, and mismatches between the frame of administrative records and the universe of interest for the information product).
- Processing (e.g., errors due to coding, data entry, editing, weighting, linking records, disclosure avoidance methods, and inaccuracies of assumptions used to develop estimates).
- Measurement (e.g., errors due to interviewer and respondent behavior, data collection instrument design, data collection modes, definitions of reference periods, reporting unit definitions, and inconsistencies in administrative records data).

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Errors strictly associated with a modeling methodology. Statistical Quality Standard D2, *Producing Estimates from Models*, addresses these types of error.

Key Terms: Convenience sample, coverage, coverage error, coverage ratio, equivalent quality data, item allocation rate, item nonresponse, key variables, latent class analysis, longitudinal survey, measurement error, nonresponse bias, nonresponse error, nonsampling error, probability of selection, quantity response rate, reinterview, release phase, respondent debriefing, response analysis survey, total quantity response rate, and unit nonresponse.

Requirement D3-1: Throughout all processes associated with producing measures and indicators of nonsampling error, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum

of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement D3-2: A plan must be developed that addresses:

1. The general measures and indicators of nonsampling error that will be produced (e.g., coverage ratios, unit nonresponse rates, item nonresponse rates, data entry error rates, coding error rates, and interviewer quality control (QC) results).
2. Any special evaluations to be conducted (e.g., studies of interviewer variance, measurement error, and nonresponse bias). Identify the:
   a. Motivation for the study.
   b. Types of errors addressed by the study.
   c. Measures and indicators to be generated.
   d. Data needed to conduct the evaluation and their sources.
   e. Methods for collecting and analyzing the data.
3. Verification and testing of systems for producing measures and indicators of nonsampling error.
4. Evaluating the measures and indicators to guide improvements to the program.

Note: Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including estimates of schedule and costs.

Requirement D3-3: Except in the situations noted below, weighted response rates must be computed to measure unit and item nonresponse. The weights must account for selection probabilities, including probabilities associated with subsampling for nonresponse follow-up.

Response rates may be computed using unweighted data when:

1. Monitoring and managing data collection activities.
2. Making comparisons with surveys using unweighted response rates.
3. Using weighted response rates would disrupt a time series.
4. A weighted response rate would be misleading because the sampling frame population in an establishment survey is highly skewed, and a stratified sample design is employed. (See Sub-Requirement D3-3.2.)
5. The Census Bureau simply collects data for a sponsor and performs no post-collection estimation.

Note: In general, computing response rates is not appropriate for samples that are not randomly selected (e.g., convenience samples or samples with self-selected respondents).

Sub-Requirement D3-3.1: For demographic surveys and decennial censuses, when computing unit response rates, item response rates or item allocation/imputation rates (for key variables), and total item response rates:

1. Standard formulas must be used. (See Appendix D3-A.)
2. The final edited data or edited outcome codes must be used, when available. If the final edited data are not used to compute the response rates, it must be noted.
3. The definition or threshold of a sufficient partial interview must be noted if partial interviews are counted as interviews.

Sub-Requirement D3-3.2: For economic surveys and censuses, when computing unit response rates, quantity response rates (for key variables), and total quantity response rates:

1. Standard formulas must be used. (See Appendix D3-B.)
2. The type of response rate must be noted: unweighted response rate, quantity response rate, or total quantity response rate.
3. The variable used in computing the response rate must be noted (e.g., total retail sales of an establishment).
4. The definition of responding units must be noted.
5. For total quantity response rates, the sources of equivalent quality data for nonresponding tabulation units must be listed (e.g., administrative records or qualified other sources such as Security Exchange Commission (SEC) filings or company annual reports).
6. The edited data at the time of each estimate's release phase must be used, when available.
7. The final edited data for the final release must be used, when available. If the final edited data are not used to compute the response rates, it must be noted.

Sub-Requirement D3-3.3: Rates for the types of nonresponse (e.g., refusal, unable to locate, no one home, temporarily absent, language problem, insufficient data, or undeliverable as addressed) must be computed to facilitate the interpretation of the unit response rate and to better manage resources.

Sub-Requirement D3-3.4: For panel or longitudinal surveys, cumulative response rates must be computed using weighted data or cumulative total quantity response rates must be computed to reflect the total attrition of eligible units over repeated waves of data collection. If a survey uses respondents from another survey or census as its sampling frame, then the response rate of the survey (or census) serving as the frame must be included in the computation of the cumulative response rate.

Sub-Requirement D3-3.5: Cumulative response rates must be computed using weighted data over successive stages of multistage data collections (e.g., a screening interview followed by a detailed interview). If estimated probabilities of selection must be used and the accuracy of the response rate might be affected, then a description of the issues affecting the response rate must also be provided.

Note: In most situations, a simple multiplication of response rates for each stage is appropriate. In other situations, a more complex computation may be required.

Sub-Requirement D3-3.6: Nonresponse bias analyses must be conducted when unit, item, or total quantity response rates for the total sample or important subpopulations fall below the following thresholds.

1. The threshold for unit response rates is 80 percent.
2. The threshold for item response rates of key items is 70 percent.
3. The threshold for total quantity response rates is 70 percent. (Thresholds 1 and 2 do not apply for surveys that use total quantity response rates.)

Note: If response rates fall below these thresholds in a reimbursable data collection, the sponsor is responsible for conducting the nonresponse bias analysis.

Requirement D3-4: Coverage ratios must be computed to measure coverage error, as an indicator of potential bias, using statistically sound methods (e.g., computing coverage ratios as the uncontrolled estimate of population for a demographic-by-geographic group divided by the population control total for the demographic-by-geographic cell used in post-stratification adjustments or using capture-recapture methods).

Note: If computing coverage ratios is not appropriate, a description of the efforts undertaken to ensure high coverage must be made available.

Requirement D3-5: Measures or indicators of nonsampling error associated with data from administrative records must be computed to inform users of the quality of the data.

Examples of measures and indicators include:

- Coverage of the target population by the set of administrative records.
- The proportion of administrative records that have missing data items or that have been imputed to address missing data.
- The proportion of data items with edit changes because the data items were invalid.
- The proportion of records lost from the analysis or estimate due to nonmatches between linked data sets.

Requirement D3-6: Measures or indicators of nonsampling error associated with data collection and processing activities must be computed to inform users of the quality of the data.

Examples of indicators of nonsampling error include:

- Error rates for data entry/data capture operations.
- Error rates and referral rates for coding operations.
- Imputation rates and edit change rates for editing and imputation operations.

Examples of analyses or studies that generate measures or indicators of nonsampling error include:

- Geocoding evaluation studies (e.g., address matching rates and analysis of rates of allocation to higher level geographic entities based on postal place-name or ZIP Code matches).
- Analyses of geospatial accuracy (e.g., analysis of locational information in relation to geodetic control points).
- Response error evaluation studies (e.g., reinterview and latent class analysis).
- Interviewer variance studies.
- Respondent debriefing studies.
- Response analysis surveys.
- Record check or validation studies.
- Mode effect studies.

Requirement D3-7: Methods and systems for calculating measures and indicators of nonsampling error must be verified and tested to ensure all components function as intended.

Examples of verification and testing activities include:

- Verifying that calculations are correct.
- Validating computer code against specifications.
- Conducting peer reviews of specifications and coding.
- Using test data to check computer programs.

Requirement D3-8: Measures and indicators of nonsampling error must be evaluated to guide improvements to the program.

Examples of evaluation activities include:

- Analyzing the quality control results of processing systems (e.g., error rates from clerical coding and clerical record linkage) and developing improvements to the systems (e.g., improving clerical coding tools or improving training for clerks).
- Evaluating the results of nonsampling error studies (e.g., response analysis surveys, respondent debriefing studies, and response error reinterview studies) and implementing improvements (e.g., revising questionnaire wording for problematic questions, revising interviewer procedures, or revising interviewer training).
- Analyzing the results of interviewer quality control systems (e.g., Quality Control (QC) reinterviews and Computer Assisted Telephone Interviewing (CATI) monitoring, and observations) and developing improvements (e.g., improving interviewer training programs or revising questionnaires to address systemic problems).

Requirement D3-9: Documentation needed to replicate and evaluate the activities associated with producing measures and indicators of nonsampling error must be produced. The documentation must be retained, consistent with applicable policies and data use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:

- Plans, requirements, specifications, and procedures for the systems.
- Computer source code.
- Results of quality control activities.
- Results of nonsampling error studies and evaluations.
- Quality measures and indicators (e.g., final coverage ratios and response rates).

Notes:

(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)

(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Appendix D3-A
Requirements for Calculating and Reporting Response Rates: Demographic Surveys and Decennial Censuses

1. Terms and Variables

The variables needed to calculate demographic survey and decennial census response rates are based on classifications suggested by the American Association for Public Opinion Research (AAPOR), 2008. This effort helps to ensure consistency to external standards while allowing the Census Bureau to adapt the classification to our specific circumstances.

The terms and variables are partitioned into three sections. The first section describes eligibility status. Variables in this section distinguish among sample units that are known to be eligible for data collection, are known to be ineligible for data collection, or have an unknown eligibility for data collection. The data collection target population guides the distinction between eligible and ineligible units. The second section describes the response status for eligible sample units. The third section provides detail on nonrespondents by identifying the type of (or the reason for) the nonresponse.

1.1 Eligibility Status

The total number of units selected for a sample is defined as *n*. These units can be classified by their eligibility status: eligible for data collection (*E)*, ineligible for data collection (*I)*, or of unknown eligibility (*U)*. The target population determines the classification of a unit as eligible or ineligible. The target population refers to persons, households, or other units upon which inferences (estimates) are made. Specific units may be considered eligible for one census or survey but ineligible for another, depending upon the target population. For example, in a survey of housing, vacant units may be part of the target population, but these same vacant units may be outside the target population in an income survey and would therefore be classified as ineligible.

Variable p_i (Probability of selection)
Definition Probability of selecting a unit for the sample, including all subsampling, even subsampling for nonresponse follow-up.

Variable w_i (Sample weight)
Definition The inverse of the final probability of selecting a unit for the sample, including all subsampling, such as subsampling for nonresponse follow-up. $w_i = \frac{1}{p_i}$

Term *E* (Eligible)

Definition The weighted count of sample units that are eligible for data collection. A person, household, or other unit is eligible if an attempt has been made to collect data and the unit is confirmed to be a member of the target population. Both occupied and vacant units can be considered eligible.

Variable e_i – An indicator variable for whether a unit selected for the sample is eligible for data collection. If a sample unit is eligible, $e_i = 1$, else $e_i = 0$.

Computation Sum of the sample weight for all eligible units.

$$E = \sum_{i=1}^{n} (w_i * e_i)$$

Reference Equivalent to the sum of AAPOR "Interview" disposition code (1.0) and "Eligible, non-interview" disposition code (2.0).

Term *I* (Ineligible)

Definition The weighted count of sample units that are ineligible for data collection. This is the number of units for which an attempt has been made to collect data and it is confirmed that the unit is not a member of the target population.

Variable i_i – An indicator variable for whether a unit selected for the sample is confirmed as not being a member of the target population at the time of data collection. Information confirming ineligibility may come from observation, from a respondent, or from another source. Some examples of ineligible units include: demolished structure, entire household in armed forces, unit under construction, unit screened out, nonresidential unit, fax/data line or disconnected number (in random-digit dial surveys), and vacant unit. If a sample unit is ineligible, $i_i = 1$, else $i_i = 0$.

Computation Sum of the sample weight for all ineligible units.

$$I = \sum_{i=1}^{n} (w_i * i_i)$$

Reference Equivalent to AAPOR "Not Eligible" disposition code (4.0).

Term *U* (Unknown eligible)

Definition The weighted count of sample units for which eligibility is unknown.

Variable u_i – An indicator variable for whether the eligibility of a unit selected for the sample could not be determined. This occurs if data are not collected from a unit and there is no information available about whether or not the unit is a member of the target population. Some examples of units with unknown eligibility include: unable to locate unit, unable to reach/unsafe area, address never assigned/worked, or number always busy or call screening/blocking (in random digit dial surveys). If a sample unit is of unknown eligibility, $u_i = 1$, else $u_i = 0$.

Computation Sum of the sample weight for all units with an unknown eligibility.

$$U = \sum_{i=1}^{n} (w_i * u_i)$$

Note Surveys that have large number of units with unknown eligibility (e.g., random-digit-dial surveys) may estimate the proportion of cases of unknown eligibility that are eligible, *ee*. This estimated proportion may be used to adjust the estimates of *I* and *E*. The survey must have a defensible basis for estimating *ee* (e.g., assume that the ratio of eligible to not eligible cases among the known cases applies to the unknown cases). Without a defensible basis, *ee* may not be used to adjust the estimates of *I* and *E*. The number of eligible units may be adjusted by adding (*ee* * *U*) to *E*. The number of ineligible units may be adjusted by adding (*U* - (*ee* * *U*)) to *I*. The basis for estimating *ee* must be stated explicitly and the justification described clearly.

Reference Equivalent to AAPOR "Unknown Eligibility, Non-Interview" disposition code (3.0).

Term *T* (Total count)

Definition The weighted count of all units (eligible, ineligible, and of unknown eligibility) selected for the sample.

Computation Sum of the sample weights for the eligibility status outcome of all units.

$$T = \sum_{i=1}^{n} [w_i * (e_i + i_i + u_i)]$$

The relationship between *E, I, U,* and *T* is $T = E + I + U$. For the *i*th unit $e_i + i_i + u_i = 1$.

1.2 Response Status

Response status is determined only for eligible sample units. The definition of sufficient data for a unit to be classified as a response will vary across surveys and will impact the count of responding units.

Term *R* (Response)

Definition The weighted count of eligible sample units with sufficient data to be classified as a response. In a multi-mode survey or census, responses may be obtained by mail, Internet, telephone, fax, touch-tone data entry/voice recognition, or personal visit.

Variable r_i – An indicator variable for whether an eligible unit selected for the sample responded to the survey and provided sufficient data. If a unit responded, $r_i = 1$ else $r_i = 0$ (note $r_i = 0$ for units classified as U or I and units that did not respond with sufficient data).

Computation Sum of the sample weights for all response outcomes.

$$R = \sum_{i=1}^{n} (w_i * r_i)$$

Reference Equivalent to AAPOR I+P (complete interviews + partial interviews) disposition codes (1.1) and (1.2).

1.3 Reasons for Nonresponse

To improve interpretation of the response rate and better manage resources, it is recommended that whenever possible, reasons for (or types of) nonresponse be measured. Six specific terms describing nonresponse reasons are defined below. These terms (*REF*, *NOH*, *TA*, *LB*, *INSF*, and *OTH*) define specific nonresponse reasons for sample units.

Term *REF* (Refusal)
Definition The weighted count of eligible sample units that refused to respond to the survey.
Variable ref_i – An indicator variable for whether an eligible sample unit refused to respond to the survey. If a unit refused to respond, $ref_i = 1$, else $ref_i = 0$.
Computation Sum of the sample weights for all "refusal" outcomes.

$$REF = \sum_{i=1}^{n} (w_i * ref_i)$$

Reference Equivalent to AAPOR "R" (refusal and break-off) – disposition code (2.10).

Term *NOH* (No one home)
Definition The weighted count of eligible sample units that did not respond because no one was found at home during the interviewing period.
Variable noh_i – An indicator variable for whether an eligible sample unit did not respond to the survey because no one was found at home during the interviewing period. If a unit was "no one home," $noh_i = 1$, else $noh_i = 0$.
Computation Sum of the sample weights for all "no one home" outcomes.

$$NOH = \sum_{i=1}^{n} (w_i * noh_i)$$

Reference Equivalent to AAPOR "No one at residence" – disposition code (2.24).

Term *TA* (Temporarily absent)
Definition The weighted count of eligible sample units that did not respond because the occupants were temporarily absent during the interviewing period.
Variable ta_i – An indicator variable for whether an eligible sample unit did not respond to the survey because the occupants were temporarily absent during the interviewing period. If a unit was "temporarily absent," $ta_i = 1$, else $ta_i = 0$.
Computation Sum of the sample weights for all "temporarily absent" outcomes.

$$TA = \sum_{i=1}^{n} (w_i * ta_i)$$

Reference Equivalent to AAPOR "Respondent away/unavailable" – disposition code (2.25).

Term *LB* (Language barrier)

Definition The weighted count of eligible sample units that did not respond because an interviewer or interpreter was not available to conduct the interview in the required language.

Variable lb_i – An indicator variable for whether an eligible sample unit selected for the sample did not respond to the survey because an interviewer or interpreter was not available to conduct the interview in the required language. If a unit did not respond due to a language barrier, $lb_i = 1$, else $lb_i = 0$.

Computation Sum of the sample weights for all "language barrier" outcomes.

$$LB = \sum_{i=1}^{n} (w_i * lb_i)$$

Reference Equivalent to AAPOR "Language" – disposition code (2.33).

Term *INSF* (Insufficient data)

Definition The weighted count of eligible sample units selected for the sample that participated but did not provide sufficient data to qualify as a response.

Variable $insf_i$ - An indicator variable for whether an eligible sample unit that was selected for the sample returned a questionnaire, but did not provide sufficient data to qualify as a response. If a unit returned a questionnaire but fails to provide sufficient data to qualify as a response, $insf_i = 1$, else $insf_i = 0$.

Computation Sum of the sample weights for "insufficient data" outcomes.

$$INSF = \sum_{i=1}^{n} (w_i * insf_i)$$

Reference Equivalent to AAPOR "Break off" and "Break-off questionnaire too incomplete to process" – disposition code (2.12).

Term *OTH* (Other nonresponse)

Definition The weighted count of sample units that did not respond for a reason other than refusal, no one home, language barrier, temporarily absent, insufficient data, or if a unit was classified as unknown eligibility.

Variable oth_i – An indicator variable for whether a unit selected for the sample was a nonresponse for a reason other than refusal, no one home, language barrier, temporarily absent, or insufficient data or if the unit was classified as unknown eligibility. If a unit does not respond for reasons other than refusal, no one home, language barrier, temporarily absent, insufficient data, or if a unit was classified as unknown eligibility, $oth_i = 1$, else $oth_i = 0$.

Computation Sum of the sample weights for "other nonresponse" outcomes.

$$OTH = \sum_{i=1}^{n} (w_i * oth_i)$$

Reference Equivalent to AAPOR "Other," "Dead," "Physically or mentally unable," and "Miscellaneous" – disposition codes (2.30), (2.31), (2.32), and (2.35).

2. Unit Response and Nonresponse Rates

2.1 Primary Response Rates

Rate *UR* rate (Unit Response Rate)
Definition The ratio of responding units to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *UR* rate = [*R*/(*E*+*U*)] * 100
Reference Equivalent to AAPOR Response Rate 2 (RR2).

Rate *AR* rate *(*Alternative Response Rate)
Definition The ratio of responding units to estimated eligible units (expressed as a percentage).
Computation *AR* rate = [*R*/[(*E*)+*ee***U*] * 100 where: *ee* = estimated proportion of cases of unknown eligibility that are actually eligible. The survey must have a defensible basis for estimating *ee*. If such a basis does not exist, then *ee* may not be used to adjust the estimates of *I* and *E* and the survey may not estimate the *AR rate*.
Reference Equivalent to AAPOR Response Rate 3 (RR3).

Rate UR_M rate (Cumulative Unit Response Rate for multistage surveys)
Definition The product of unit response rates for all stages of the survey
Computation UR_M rate = $\prod_{j=1}^{k} UR_j$ where, UR_j is the unit response rate at stage *j* of the survey and *k* is the total number of stages. If another equation yields a more accurate estimate of the cumulative unit response rate because it uses additional information about the frame, then that equation should be used. If the cumulative response rate is misleading or inaccurate, an explanation of the problems must be documented.

2.2 Detailed Eligibility and Nonresponse Rates

Rate *REF* rate (Refusal Rate)
Definition The ratio of units classified as "refusals" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *REF* rate = [*REF*/(*E*+*U*)] * 100
Reference Equivalent to AAPOR Refusal Rate 1 (REF1).

Rate *NOH* rate (No One Home Rate)
Definition The ratio of units classified as "no one home" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *NOH* rate = [*NOH*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

Rate *TA* rate (Temporary Absent Rate)
Definition The ratio of units classified as "temporarily absent" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *TA* rate = [*TA*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

Rate *LB* rate (Language Barrier Rate)
Definition The ratio of units classified as "language barriers" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *LB* rate = [*LB*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

Rate *INSF* rate (Insufficient Data Rate)
Definition The ratio of units classified as having "insufficient data" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *INSF* rate = [*INSF*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

Rate *OTH* rate (Other Reason for Nonresponse Rate)
Definition The ratio of units classified as "other nonresponse" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *OTH* rate = [*OTH*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

Rate *U* rate (Unknown Eligibility Rate)
Definition The ratio of units classified as having an "unknown eligibility" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *U* rate = [*U*/(*E*+*U*)] * 100
Reference No AAPOR equivalent.

3. Item Response and Allocation Rates

3.1 Item Response Rates

Term $IREQ_A$ (Weighted total of responses required for data item A)
Definition The weighted count of sample units for which a response to item A is required. A response is required for item A unless it is a valid skip item.
Variable $ireq_{Ai}$ – An indicator variable for whether a response to item A is required. If a response is required, $ireq_{Ai} = 1$, else $ireq_{Ai} = 0$

Computation Sum of the sample weight for all units requiring a response to item A.

$$IREQ_A = \sum_{i=1}^{n} ireq_{Ai} * w_i$$

Term $ITEM_A$ (Total valid responses for data item A)
Definition The weighted count of sample units for which a valid response to item A is obtained.
Variable $item_{Ai}$ – An indicator variable for whether a valid response to item A is obtained. If a valid response is obtained, $item_{Ai} = 1$, else $item_{Ai} = 0$
Computation Sum of the sample weight for all units requiring a response to item A for which a valid response is obtained.

$$ITEM_A = \sum_{i=1}^{n} item_{Ai} * w_i$$

Rate IR_A rate (Item response rate for data item A)
Definition The ratio of the weighted count of units with a valid response to item A to the weighted count of units that required a response to item A.
Computation IR_A rate = $ITEM_A$ / $IREQ_A$

Rate TIR_A rate (Total item response rate for data item A)
Definition The product of the weighted item response rate for item A and either the unit response rate, reflecting the response rate to item A after accounting for both unit nonresponse and item nonresponse, or the cumulative unit response rate for multistage surveys.
Computation TIR_A rate = $IR_A * UR$ or
TIR_A rate = $IR_A * UR_M$

3.2 Item Allocation Rates

Item nonresponse is measured through the calculation of allocation rates. Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices populated by donors, to impute for missing values.

Term $ALLO_A$ (Total number of responses allocated for item A)
Definition The weighted count of sample units for which a response is allocated to item A.
Variable $allo_{Ai}$ – An indicator variable for whether a response is allocated to item A. If a response is obtained, $allo_{Ai} = 1$, else $allo_{Ai} = 0$
Computation Sum of the sample weight for all units requiring a response to item A for which a response is allocated.

$$ALLO_A = \sum_{i=1}^{n} allo_{Ai} * w_i$$

Rate	IA_A rate (Item allocation rate for data item A)
Definition	The ratio of the weighted count of units with an allocated response to item A to the weighted count of units that required a response to item A.
Computation	IA_A rate = $ALLO_A$ / $IREQ_A$ = 1 - IR_A rate

References

The American Association for Public Opinion Research. 2008. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 5th edition. Lenexa, Kansas: AAPOR.

Appendix D3-B
Requirements for Calculating and Reporting Response Rates: Economic Surveys and Censuses

1. Terms and Variables

For many economic programs, there is a need to distinguish between the survey (sampling) unit, the reporting unit, and the tabulation unit:

A **survey unit** is an entity selected from the underlying statistical population of similarly-constructed units. Examples of survey units for different economic programs include establishments, Employer Identification Numbers (EIN), firms, state and local government entities, and building permit-issuing offices. Some programs use different survey units for different segments of the total population. Examples include the Annual Retail Trade Survey (ARTS) and the Survey of Construction (SOC). The ARTS samples EINs and firms (which can be comprised of one or more establishments), and the SOC samples residential housing permits and newly constructed housing units in areas where no permit is required. For cross-sectional or longitudinal surveys, the survey unit may change in composition over time (perhaps due to mergers, acquisitions, or divestitures).

A **reporting unit** is an entity from which data are collected. Reporting units are the vehicle for obtaining data and may or may not correspond to a survey unit for several reasons. First, the composition of the originally-sampled entity can change over the sample's life cycle, as noted above. Second, for some surveys, an entity may request (or the Census Bureau may ask the entity) to report data in several separate pieces corresponding to different parts of the business or other entity type. For example, a large, diverse company in a company-based collection may request a separate form for each region or line of business in which it operates or may ask to report separately for each of its establishments to align with their record keeping practices. Similarly, many government programs have a central collection agency that provides the data for several governments, but issue additional mail-outs to obtain supplemental items that are not obtained by the central collection agency.

A **tabulation unit** houses the data used for estimation (or tabulation, in the case of a census). As with reporting units, the tabulation units may not correspond to a survey unit. Some programs consolidate establishment or plant-level data to a company level or parent government level to create tabulation units, so that the tabulation unit is often equivalent to the survey unit. Other programs create artificial units that split a reporting unit's data among the different categories in which the reporting unit operates; for example, creating separate tabulation units by industry. In this case, the tabulation unit represents a portion of a survey unit.

For each program, the "statistical period" describes the reference period for the data collection. For example, an annual program might collect data on the prior year's business; the statistical period refers to the prior year, but the data are obtained in the calendar year. During a given statistical period, all three types of units can be active, inactive, or ineligible. An active unit is in business and is in-scope for the program during the statistical period. An inactive unit is not operating or is not in-scope during the statistical period but is believed to have been active in the

past and can potentially become active and in-scope in the future; examples include seasonal businesses for monthly or quarterly programs (temporarily idle) or businesses that operate in more than one industry, with the primary activity for a given statistical period being conducted in an "out-of-scope" industry. Finally, a survey unit may become ineligible and permanently excluded from subsequent computations due to a merger or acquisition, a permanent classification category change, or a death. All units are considered active until verified evidence otherwise is provided.

Economic programs compute two different types of response rates: a unit response rate and weighted item response rates. The Unit Response Rate (URR) is defined as the ratio of the total unweighted number of "responding units" to the total number of units eligible for collection. URRs are indicators of the performance of data collection for obtaining usable responses. Consequently, the majority of business programs base URRs on their reporting units, whereas the majority of ongoing government programs base URRs on the survey units[1] that correspond to the tabulation units. Other exceptions are addressed on a case-by-case basis. The formulae for the URR provided in Section 2.1 and the detailed unit nonresponse rate breakdowns presented in the Section 2.2.1 use the term "reporting unit" for simplicity. A program can produce **at most** one URR per statistical period and per release phase[2].

Quantity and Total Quantity Response Rates (QRR and TQRR) are item-level indicators of the "quality" of each estimate. In contrast to the URR, these weighted response rates are computed for individual data items, so that a program may produce several QRRs and TQRRs per statistical period and release. Both are weighted measures that take the size of the tabulation unit into account as well as the associated sampling parameters. These rates measure the proportion of each estimate obtained directly or indirectly from the survey unit and are consequently based on the tabulation units. The QRR measures the weighted proportion of an estimate obtained directly from the respondent for the survey/census; the TQRR expands the rate to include data from equivalent quality sources.

To compute the weighted item response rates, it is necessary to determine the source of the final tabulated value of the associated data item for each tabulation unit *i*. This value could be directly obtained from respondent data, indirectly obtained from other equivalent quality data sources, or imputed. The classification process is straightforward for items that are directly obtained from the survey questionnaire (i.e., form items), less so for items that are obtained as functions of collected items (i.e., derived items). The formulae provided in Sections 2.1 and 2.2.2. can be applied to either form or derived items, but require that the item value classification process be performed immediately prior and that the classification process or rules be documented.

[1] The central collection unit may provide the responses for the majority of the program data (e.g., providing responses from all associated sample units for most of the program items). Supplemental mailings are used to obtain the rest of the items.

[2] Leading indicator surveys often have more than one official release of the same estimate. For example, a program might release a preliminary estimate for the current statistical period along with a revised estimate from the prior period. Response rates should be computed at each release phase, and it is expected that the response rates (unit or item) will generally increase for the same estimate with each release.

1.1 Eligibility Status

The total number of active reporting units in a statistical period is defined as N_{RU}. These reporting units can be classified by their eligibility status: eligible for data collection (*E*), ineligible (*IA*), unknown eligibility (*U*), or data obtained from qualified administrative sources (*A*). Reporting units that have been determined to be out-of-scope for data collection during the statistical period are excluded from all computations, as are inactive cases. Note that the *U* cases are assumed to be active and in-scope in the absence of evidence otherwise. Reporting units may be considered eligible in one survey or census but ineligible for another, depending upon the target population. For example, a reporting unit that was in business after October 2004 is eligible for the 2004 Annual Retail Trade Survey, but is ineligible for the October 2004 Monthly Retail Trade Survey.

Term	*E* (Total Eligible)
Definition	The count of reporting units that were eligible for data collection in the statistical period.
Variable	e_i – An indicator variable for whether a reporting unit is eligible for data collection in the statistical period. These include chronic refusal units (eligible reporting units that have notified the Census Bureau that they will not participate in a given program). If a reporting unit is eligible, $e_i = 1$, else $e_i = 0$.
Computation	The sum of the indicator variable for eligibility (e_i) over all the reporting units in the statistical period. $E = \sum_{i=1}^{N_{RU}} e_i$

Term	*IA* (Total Ineligible/Inactive)
Definition	The count of reporting units that were ineligible for data collection in the current statistical period.
Variable	ia_i – An indicator variable for whether a reporting unit in the statistical period has been confirmed as not a member of the target population at the time of data collection. An attempt was made to collect data, and it was confirmed that the reporting unit was not a member of the target population at that time. These reporting units are not included in the URR calculations for the periods in which they are ineligible. Information confirming ineligibility may come from observation, from a respondent, or from another source. Some examples of ineligible reporting units include firms that went out of business prior to the survey reference period, firms in an industry that is out-of-scope for the survey in question, and governments that reported data from outside of the reference period. If a reporting unit is ineligible, $ia_i = 1$, else $ia_i = 0$.
Computation	The sum of the indicator variable for ineligibility (ia_i) over all the reporting units in the statistical period. $IA = \sum_{i=1}^{N_{RU}} ia_i$

Term	*U* (Total Unknown Eligibility)

Definition The count of reporting units in the statistical period for which eligibility could not be determined.

Variable u_i – An indicator variable for whether the eligibility of a reporting unit in the statistical period could not be determined. If a reporting unit is of unknown eligibility, $u_i = 1$, else $u_i = 0$. For example, units whose returns are marked as "undeliverable as addressed" have unknown eligibility ($u_i = 1$), as do unreturned mailed forms.

Computation The sum of the indicator variable for unknown eligibility (u_i) over all the reporting units in the statistical period. $U = \sum_{i=1}^{N_{RU}} u_i$

Term *A* (Administrative data used as source)

Definition The count of reporting units in the statistical period that belong to the target population and were pre-selected to use administrative data rather than collect survey data.

Variable a_i – An indicator variable for whether administrative data of equivalent-quality-to-reported data rather than survey data was obtained for an eligible reporting unit in the statistical period. The decision not to collect survey data must have been made for survey efficiency or to reduce respondent burden and not because that reporting unit had been a refusal in the past. These reporting units are excluded from the URR calculations because they were not sent questionnaires, and thus could not respond, although their data are included in the calculation of the TQRRs. If a reporting unit is pre-selected to receive administrative data, $a_i = 1$, else $a_i = 0$.

Computation The sum of the indicator variable for units pre-selected to use administrative data (a_i) over all the reporting units in the statistical period. $A = \sum_{i=1}^{N_{RU}} a_i$

The relationship among the counts of reporting units in the statistical period in the four eligibility categories is given by $N_{RU} = E + IA + U + A$. For the i^{th} reporting unit, $e_i + ia_i + u_i + a_i = 1$. Note that the value of N_{RU} may change by statistical period.

1.2 Response Status

Response status is determined only for the eligible active reporting units in the statistical period.

Term *R* (Response)

Definition The count of reporting units in the statistical period that were eligible for data collection in the statistical period and classified as a response.

Variable r_{ui} – An indicator variable for whether an eligible reporting unit in the statistical period responded to the survey. To be classified as a response, the respondent for the reporting unit must have provided sufficient data, and the data must satisfy all

the critical edits. The definition of sufficient data will vary across surveys. Programs must designate required data items before the data collection begins. If a reporting unit responded, $r_{ui} = 1$, else $r_{ui} = 0$ (note $r_{ui} = 0$ for reporting units which were eligible but did not respond and for reporting units classified as *IA*, *U*, or *A*).

Computation The sum of the indicator variable for eligible reporting units that responded (r_{ui}) over all the reporting units in the statistical period. $R = \sum_{i=1}^{N_{RU}} r_{ui}$

1.3 Reasons for Nonresponse

To improve interpretation of the response rate and better manage resources, it is recommended that whenever possible, reasons for (or types of) nonresponse be measured on a flow basis whenever possible. These terms are used to describe "unit nonresponse" and will be presented in unweighted tabulations. Five specific terms describing nonresponse reasons are defined below. The first three terms (*REF*, *CREF*, and *INSF*) define nonresponse reasons for eligible reporting units. The final two terms (*UAA* and *OTH*) define the reasons for reporting units with unknown eligibility.

Term *REF* (Refusal)

Definition The count of eligible reporting units in the statistical period that were classified as "refusal."

Variable ref_i – An indicator variable for whether an eligible reporting unit in the statistical period refused to respond to the survey. If a reporting unit refuses to respond, $ref_i = 1$, else $ref_i = 0$.

Computation Sum of the indicator variable for "refusal" (ref_i) over all the reporting units in the statistical period. $REF = \sum_{i=1}^{N_{RU}} ref_i$

Term *CREF* (Chronic refusal)

Definition The count of eligible reporting units in the statistical period that were classified as "chronic refusals."

Variable $cref_i$ – An indicator variable for whether an eligible reporting unit in the statistical period was a "chronic refusal." A chronic refusal is a reporting unit that informed the Census Bureau that it would not participate in a given program. The Census Bureau does not send questionnaires to chronic refusals, but they are considered to be eligible reporting units. Chronic refusals comprise a subset of refusals. If a reporting unit is a chronic refusal, $cref_i = 1$, else $cref_i = 0$.

Computation The sum of the indicator variable for "chronic refusal" ($cref_i$) over all the reporting units in the statistical period. $CREF = \sum_{i=1}^{N_{RU}} cref_i$

Term *INSF* (Insufficient data)

Definition The count of eligible reporting units in the statistical period that were classified as providing insufficient data.

Variable $insf_i$ -- An indicator variable for whether an eligible reporting unit in the statistical period returned a questionnaire, but did not provide sufficient data to qualify as a response. If a reporting unit returned a questionnaire but failed to provide sufficient data to qualify as a response, $insf_i = 1$, else $insf_i = 0$.

Computation The sum of the indicator variable for "insufficient data" ($insf_i$) over all the reporting units in the statistical period. $INSF = \sum_{i=1}^{N_{RU}} insf_i$

Term *UAA* (Undeliverable as addressed)

Definition The count of reporting units in the statistical period that were classified as "undeliverable as addressed."

Variable uaa_i – An indicator variable for whether a reporting unit in the statistical period had a questionnaire returned as "undeliverable as addressed." These reporting units are of unknown eligibility. If a questionnaire is returned as "undeliverable as addressed," $uaa_i = 1$, else $uaa_i = 0$.

Computation The sum of the indicator variable for "undeliverable as addressed" (uaa_i) over all the reporting units in the statistical period. $UAA = \sum_{i=1}^{N_{RU}} uaa_i$

Term *OTH* (Other nonresponse)

Definition The count of reporting units in the statistical period that were classified as "other nonresponse."

Variable oth_i – An indicator variable for whether a reporting unit in the statistical period was a nonresponse for a reason other than refusal, insufficient data, or undeliverable as addressed. These reporting units are of unknown eligibility. If a reporting unit does not respond for reasons other than refusal, insufficient data, or undeliverable as addressed, $oth_i = 1$, else $oth_i = 0$.

Computation The sum of the indicator variable for "other nonresponse" (oth_i) over all the reporting units in the statistical period. $OTH = \sum_{i=1}^{N_{RU}} oth_i$

1.4 Quantity Response Rate Terms

The total number of active tabulation units in the statistical period is defined as N_{TU}. Recall that the number of tabulation units N_{TU} ***may*** differ from the number of reporting units N_{RU}, depending on the economic program. After a program creates tabulation units and performs any necessary data allocation procedures (from reporting unit(s) to tabulation unit(s)), the individual data items are classified according to the final source of data obtained for the units: data reported by the respondent, equivalent-quality-to-reported data obtained from the program-approved outside sources (such as company annual reports, Security Exchange Commission (SEC) sites, trade association statistics), or imputed data. Tabulation units that have been determined to be out-of-

scope for data collection during the statistical period are excluded from all computations, as are inactive cases.

Variable v_{ti} (Tabulated value of data item *t* for tabulation unit *i* in the statistical period)
Definition The quantity stored in the variable for item *t* for the i^{th} tabulation unit in the statistical period. This quantity may be reported, equivalent-quality-to-reported, or imputed.

Term R_t (Reported data tabulation units for item *t*)
Definition The count of eligible tabulation units that provided reported data during the studied statistical period for item *t* that satisfied all critical edits. This count will vary by item and by statistical period.
Variable r_{ti} – An indicator variable for whether tabulation unit *i* in the statistical period provided reported data for item *t* that satisfied all edits. If the tabulated item *t* value for unit *i* (t_i) contains reported data, then $r_{ti} = 1$, else $r_{ti} = 0$.
Computation The sum of the indicator variable for reported data (r_{ti}) over all the tabulation units (N_{TU}) in the statistical period. $R_t = \sum_{i=1}^{N_{TU}} r_{ti}$

Term Q_t (Equivalent-quality-data tabulation units for item *t*)
Definition The count of eligible tabulation units that use equivalent-quality-to-reported data for item *t*. Note that these data are **indirectly** obtained for the tabulation unit. This count will vary by item and by statistical period.
Variable q_{ti} – An indicator variable for whether tabulation unit *i* in the statistical period contains equivalent-quality-to-reported data for item *t*. Such data can come from three sources: data directly substituted from another census or survey *s* (for the same reporting unit, data item concept, and time period), administrative data *d*, or data obtained from some other equivalent source *c* validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (e.g., company annual reports, Securities and Exchange Commission (SEC) filings, trade association statistics). If the tabulated item *t* value for unit *i* (t_i) contains equivalent-quality-to-reported data then $q_{ti} = 1$, else $q_{ti} = 0$.
Computation The sum of the indicator variable for equivalent-quality-to-reported data (q_{ti}) over all tabulation units (N_{TU}) in the statistical period. $Q_t = \sum_{i=1}^{N_{TU}} q_{ti}$

Term S_t (Substituted data tabulation units for item *t*)
Definition The count of eligible tabulation units containing directly substituted data for item *t*. This count will vary by item and by statistical period.

Variable s_{ti} – An indicator variable for whether a tabulation unit in the statistical period contains directly substituted data from another census or survey for item *t*. The same reporting unit must provide the item value (in the other program), and the item concept and time period for the substituted values must agree between the

two programs. If the tabulated item *t* value for unit *i* (t_i) contains directly substituted data from another survey, $s_{ti} = 1$, else $s_{ti} = 0$.

Computation The sum of the indicator variable for directly substituted data (s_{ti}) over all tabulation units (N_{TU}) in the statistical period. $S_t = \sum_{i=1}^{N_{TU}} s_{ti}$

Term D_t (Administrative data tabulation units for item *t*)

Definition The count of eligible tabulation units containing administrative data for item *t*. This count will vary by item and by statistical period.

Variable d_{ti} – An indicator variable for whether a tabulation unit in the statistical period contains administrative data for item *t*. If the tabulated item *t* value for unit *i* (t_i) contains administrative data, $d_{ti} = 1$, else $d_{ti} = 0$.

Computation The sum of the indicator variable for administrative data (d_{ti}) over all tabulation units (N_{TU}) in the statistical period. $D_t = \sum_{i=1}^{N_{TU}} d_{ti}$

Term C_t (Equivalent source data tabulation units for item *t*)

Definition The count of eligible tabulation units containing equivalent-source data that is neither administrative data nor data substituted directly from another economic program for item *t*. This count will vary by item and by statistical period.

Variable c_{ti} – An indicator variable for whether a tabulation unit in the statistical period contains equivalent-source data validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (e.g., company annual report, SEC filings, trade association statistics) for item *t*. If the tabulated item *t* value for unit *i* (t_i) contains equivalent-source data, then $c_{ti} = 1$, else $c_{ti} = 0$.

Computation The sum of the indicator variable for equivalent-source data (c_{ti}) over all tabulation units (N_{TU}) in the statistical period. $C_t = \sum_{i=1}^{N_{TU}} c_{ti}$

Term M_t (Imputed data tabulation units for item *t*)

Definition The count of eligible tabulation units containing imputed data for item *t*. This count will vary by item and by statistical period.

Variable m_{ti} – An indicator variable for whether a tabulation unit in the statistical period contains imputed data for item *t*. If the tabulated item *t* value for unit *i* (t_i) contains imputed data, $m_{ti} = 1$, else $m_{ti} = 0$.

Computation The sum of the indicator variable for imputed data (m_{ti}) over all tabulation units (N_{TU}) in the statistical period. $M_t = \sum_{i=1}^{N_{TU}} m_{ti}$

The relationship among Q_t, S_t, D_t, and C_t for item t in a statistical period is given by $Q_t = S_t + D_t + C_t$. The relationship among the counts of tabulation units for item t in the statistical period is given by $N_{TU} = R_t + Q_t + M_t$.

Variable f_i (Nonresponse weight adjustment factor)
Definition A tabulation unit nonresponse weight adjustment factor for the i^{th} tabulation unit in the statistical period. The variable f_i is set equal to 1 for surveys that use imputation to account for unit nonresponse.

Variable w_i (Sample weight)
Definition The design weight for the i^{th} tabulation unit in the statistical period. The design weight includes subsampling factors and outlier adjustments, but excludes post-sampling adjustments for nonresponse and for coverage. This variable represents the inverse unbiased probability of selection for the tabulation unit.

Variable t_i (Design-weighted value of item t for tabulation unit i)
Definition The design-weighted tabulated quantity of the variable for item t for the i^{th} tabulation unit in the statistical period (i.e, $t_i = w_i v_{ti}$). Note that this value has not been adjusted for unit non-response.

Term T (Total value for item t)
Definition The estimated (weighted) total of data item t for the entire population represented by the tabulation units in the statistical period. T is based on the value of the data provided by the respondent, equivalent-quality-to-reported data, or imputed data. The calculation of T incorporates subsampling factors, weighting adjustment factors for unit nonresponse (adjustment-to-sample procedures only), and outlier-adjustment factors, but does not include post-stratification or other benchmarking adjustments.
Computation The product of the design weighted tabulated value of item t for the i^{th} tabulation in the statistical period (t_i) and the nonresponse weight adjustment factor (f_i), summed over all tabulation units (N_{TU}) in the statistical period. $T = \sum_{i=1}^{N_{TU}} f_i t_i$

2. Response and Nonresponse Rates

The rates defined below serve as quality indicators in the process control sense for non-negatively valued items such as total employees or total payroll. For items that can take on positive and negative values, such as income or earnings on investments, the program should plan to develop two sets of weighted item response rates (QRRs and TQRRs) – one from negatively valued data and one from non-negatively valued data – or propose alternative quality indicators that provide adequate transparency into data quality and assist in taking corrective actions.

2.1 Primary Response Rates

Rate *URR* (Unit Response Rate)
Definition The proportion of reporting units in the statistical period based on unweighted counts, that were eligible or of unknown eligibility that responded to the survey (expressed as a percentage).
Computation $URR = [R/(E+U)] * 100$

Rate *QRR* (Quantity Response Rate for data item *t*)
Definition The proportion of the estimated (weighted) total (*T*) of data item *t* reported by the active tabulation units in the statistical period (expressed as a percentage).
Computation $$QRR = \left[\frac{\sum_{i=1}^{N_{TU}} r_{ti} \times t_i}{T} \right] * 100$$

Rate *TQRR* (Total Quantity Response Rate for data item *t*)
Definition The proportion of the estimated (weighted) total (*T*) of data item *t* reported by the active tabulation units in the statistical period or from sources determined to be equivalent-quality-to-reported data (expressed as a percentage).
Computation $$TQRR = \left[\frac{\sum_{i=1}^{N_{TU}} (r_{ti} + q_{ti}) \times t_i}{T} \right] * 100$$

2.2 Detailed Response and Nonresponse Rates

2.2.1 Unit Nonresponse Rate Breakdowns

The following breakdowns provide unweighted unit nonresponse rates.

Rate *REF* rate (Refusal Rate)
Definition The ratio of reporting units in the statistical period that were classified as "refusal" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation $REF\ rate = [REF/(E+U)] * 100$

Rate *CREF* rate (Chronic Refusal Rate)
Definition The ratio of reporting units in the statistical period that were classified as "chronic refusals" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation *CREF rate* = [*CREF*/(*E*+*U*)] * 100

Rate *INSF* rate (Insufficient Data Rate)
Definition The ratio of reporting units in the statistical period that were classified as "insufficient data" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *INSF rate* = [*INSF*/(*E*+*U*)] * 100

Rate *UAA* rate (Undeliverable as Addressed Rate)
Definition The ratio of reporting units in the statistical period that were classified as "undeliverable as addressed" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *UAA rate* = [*UAA*/(*E*+*U*)] * 100

Rate *OTH* rate (Other Reason for Nonresponse Rate)
Definition The ratio of reporting units in the statistical period that were classified as "other reason for nonresponse" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *OTH rate* = [*OTH*/(*E*+*U*)] * 100

Rate *U* rate (Unknown Eligibility Rate)
Definition The ratio of reporting units in the statistical period that were classified as "unknown eligibility" to the sum of eligible units and units of unknown eligibility (expressed as a percentage).
Computation *U rate* = [*U*/(*E*+*U*)] * 100

2.2.2 Total Quantity Response Rate Breakdowns

The following breakdowns provide weighted item response rates.

Rate *Q* rate (Equivalent-Quality-to-Reported Data Rate)
Definition The proportion of the total estimate for item *t* derived from equivalent-quality-to-reported data for tabulation units in the statistical period (expressed as a percentage).
Computation $Q \text{ rate} = \left[\frac{\sum_{i=1}^{N_{TU}} (s_{ti} + d_{ti} + c_{ti}) \times t_i}{T} \right] * 100 = \left[\frac{\sum_{i=1}^{N_{TU}} q_{ti} \times t_i}{T} \right] * 100$

Rate *S* rate (Survey Substitution Rate)

Definition The proportion of the total estimate for item *t* derived from substituted other survey or census data for tabulation units in the statistical period (expressed as a percentage). To be tabulated in this rate, substituted data items must be obtained from the same reporting unit in the same time period as the target program, and the item concept between the two programs must agree.

Computation $S \text{ rate} = \left[\frac{\sum_{i=1}^{N_{TU}} s_{ti} \times t_i}{T} \right] * 100$

Rate *D* rate (Administrative Data Rate)

Definition The proportion of the total estimate of item *t* derived from administrative data for tabulation units in the statistical period (expressed as a percentage).

Computation $D \text{ rate} = \left[\frac{\sum_{i=1}^{N_{TU}} d_{ti} \times t_i}{T} \right] * 100$

Rate *C* rate (Other Source Rate)

Definition The proportion of the total estimate of item *t* derived from other source data validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (such as company annual reports, SEC filing, trade association statistics) for tabulation units in the statistical period (expressed as a percentage).

Computation $C \text{ rate} = \left[\frac{\sum_{i=1}^{N_{TU}} c_{ti} \times t_i}{T} \right] * 100$

Rate *M* rate (Imputation Rate)

Definition The proportion of the total estimate of item *t* derived from imputes for tabulation units in the statistical period (expressed as a percentage).

Computation $M \text{ rate} = \left[\frac{\sum_{i=1}^{N_{TU}} m_{ti} \times t_i}{T} \right] * 100$

ANALYZING DATA AND REPORTING RESULTS

Statistical Quality Standard E1
Analyzing Data

Purpose: The purpose of this standard is to ensure that statistical analyses, inferences, and comparisons used to develop information products are based on statistically sound practices.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards, including contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the analyses performed to generate information products. It includes analyses:

- Used to produce Census Bureau information products (e.g., reports, news releases, conference papers, journal articles, and maps), regardless of data source.
- Conducted using census data, survey data, administrative records data, or any data linked with any of these sources.
- Performed during research to develop improved methodologies for frame construction, survey design, sampling, data collection, data capture, processing, estimation, analysis, or other statistical processes.
- Performed to evaluate the quality of Census Bureau data, methodologies, and processes.
- Conducted to guide decisions about processes or information products of the Census Bureau's programs.

Exclusions:
The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Key Terms: Bonferroni correction, cluster, covariance, direct comparison, goodness-of-fit, hypothesis testing, implied comparison, multivariate analysis, outliers, parameter, peer review, regression, sample design, Scheffe's method, sensitivity analysis, significance level, statistical inference, and Tukey's method.

Requirement E1-1: Throughout all processes associated with analyzing data, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement E1-2: A plan must be developed prior to the start of the analysis that addresses, as appropriate:

1. A description of the analysis, addressing issues such as:
   - Purpose.

- Research questions or hypotheses.
- Relevant literature.

2. A description of the data, addressing issues such as:
   - The data source(s).
   - Key variables and how they relate to the concept(s) in the hypotheses.
   - Design and methods used to collect and process the data.
   - Limitations of the data.
3. A description of the methodology, addressing issues such as:
   - Analysis methods (e.g., demographic and economic analysis techniques, ANOVA, regression analysis, log-linear analysis, nonparametric approaches, box plots, and scatter plots).
   - Key assumptions used in the analysis.
   - Tests (e.g., z-tests, F-test, chi-square, and R-squared) and significance levels used to judge significance, goodness-of-fit, or degree of association.
   - Limitations of the methodology.
4. Appropriateness of the data and underlying assumptions and verification of the accuracy of the computations.

Notes:
(1) During a data analysis project, the focus of the analysis may change, as the researcher learns more about the data. The analysis plan should be updated, as appropriate, to reflect major changes in the direction of the analysis.
(2) Statistical Quality Standard A1, *Planning a Data Program*, addresses overall planning requirements, including schedule and estimates of costs.

Requirement E1-3: Statistically sound practices that are appropriate for the research questions must be used when analyzing the data.

Examples of statistically sound practices include:
- Reviewing data to identify and address nonsampling error issues (e.g., outliers, inconsistencies within records, missing data, and bias in the frame or sample from which data are obtained).
- Validating assumptions underlying the analysis, where feasible.
- Developing models appropriate for the data and the assumptions. (See Statistical Quality Standard D2, *Producing Estimates from Models*.)
- Using multiple regression and multivariate analysis techniques, when appropriate, to examine relationships among dependent variables and independent variables.
- Using a trend analysis or other suitable procedure when testing for structure in the data over time (e.g., regression, time series analysis, or nonparametric statistics).

Sub-Requirement E1-3.1: The data analysis must account for the sample design (e.g., unequal probabilities of selection, stratification, and clustering) and estimation methodology.

Notes:

(1) If it has been documented that a particular methodological feature(s) has no effect on the results of the analysis, then it is not necessary to account for that feature in the analysis (e.g., if using weighted and unweighted data produce similar results, then the analysis may use the unweighted data; if the variance properties for clustered data are similar to those for unclustered data, then the analysis need not account for clustering).

Requirement E1-4: Any conclusions derived from sample data must be supported by appropriate measures of statistical uncertainty.

Examples of measures of statistical uncertainty that support conclusions include:

- Confidence or probability intervals with specified confidence levels (e.g., 90% or 95%).
- Margins of error for specified confidence levels, provided the sample size is sufficiently large that the implied confidence interval has coverall close to the nominal level.
- P-values for hypothesis tests, such as are implied when making comparisons between groups or over time. Comparisons with p-values greater than 0.10, if reported, should come with a statement that the difference is not statistically different from zero.
- Confidence intervals, probability intervals or p-values should be statistically valid and account for the sample design (e.g., accounting for covariances when the estimates are based on clustered samples). If based on a model, then the key assumptions of the model should be checked and not contradicted by the observed data. (See Statistical Quality Standard D2, *Producing Estimates from Models*.

Note: Although the p-value does not indicate the size of an effect (or the size of the difference in a comparison), p-valuse below 0.01 constitute strong evidence against the null, p-values between 0.01 and 0.05 constitute moderate evidence, and p-values between 0.05 and 0.10 constitute weak evidence.

Sub-Requirement E1-4.1: The same significance level or confidence level must be used throughout an analysis. Table A shows the requirements for specific information products:

Table A: Significance and Confidence Levels by Information Product

Information Product	Significance Level	Confidence Level
Census Bureau publications	0.10	0.90
News releases	0.10	0.90
All other information products (e.g., working papers, professional papers, and presentations)	0.10 or less	0.90 or more

Requirement E1-5: The data and underlying assumptions must be appropriate for the analyses and the accuracy of the computations must be verified.

Examples of activities to check the appropriateness of the data and underlying assumptions and the accuracy of the computations:

- Checking that the appropriate equations were used in the analysis.
- Reviewing computer code to ensure that the appropriate data and variables are used in the analysis and the code is correctly programmed.

- Performing robustness checks (e.g., checking that unexpected results are not attributable to errors, examining plots of residuals to assess fit of models and comparing findings against historical results for reasonableness).
- Performing sensitivity analyses using alternative assumptions to assess the validity of measures, relationships, and inferences.
- Requesting peer reviews by subject matter, methodological, and statistical experts to assess analysis approach and results.

Requirement E1-6: Documentation needed to replicate and evaluate the analysis must be produced. The documentation must be retained, consistent with applicable policies and data-use agreements, and must be made available to Census Bureau employees who need it to carry out their work. (See Statistical Quality Standard S2, *Managing Data and Documents*.)

Examples of documentation include:
- Plans, requirements, specifications, and procedures relating to the analysis.
- Computer code (e.g., SAS code).
- Data files with weighted and unweighted data.
- Outlier analysis results, including information on the cause of outliers, if available.
- Error estimates, parameter estimates, and overall performance statistics (e.g., goodness-of-fit statistics).
- Results of diagnostics relating to the analysis.

Notes:
(1) The documentation must be released on request to external users, unless the information is subject to legal protections or administrative restrictions that would preclude its release. (See Data Stewardship Policy DS007, *Information Security Management Program*.)
(2) Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency of information products released by the Census Bureau.

Statistical Quality Standard E2
Reporting Results

Purpose: The purpose of this standard is to ensure that information products meet statistical reporting requirements; that they provide understandable, objective presentations of results and conclusions; and that conclusions are supported by the data.

Notes:
(1) Requirement F1-4 of Statistical Quality Standard F1, *Releasing Information Products*, contains reporting requirements regarding information products affected by serious data quality issues that may impair the suitability of the products for their intended uses.
(2) Department Administrative Order (DAO) 219-1 establishes the policy for Commerce Department employees engaging in public communications.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the reporting of results in information products such as:

- News releases.
- Census Bureau publications (i.e., information products that the program's Associate Director has reviewed and approved and the Census Bureau has affirmed their content).
- Working papers (e.g., technical papers and division reports intended for release to the public).
- Professional papers (e.g., journal articles, book chapters, conference papers, poster sessions, and written discussant comments).
- Research reports used to guide decisions about Census Bureau programs.
- Abstracts.
- Presentations at public events, such as seminars or conferences. (Statistical Quality Standard E3, *Reviewing Information Products*, defines public events.)
- Handouts for distribution at public events.
- Tabulations, including custom tabulations, estimates, and their associated documentation.
- Statistical graphs, figures, and thematic maps.

Exclusions:

In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Papers, presentations, or other public communications prepared or delivered by Census Bureau employees that are not related to programs, policies, or operations of the Department of Commerce (DOC) or the Census Bureau. (The DOC *Summary of Ethics Rules* state that you may use your Census Bureau affiliation in non-official contexts only if it is used as part of general biographic information, and it is given no

more prominence than other significant biographical details. Contact the Office of Analysis and Executive Support (OAES) for additional guidance.)

Key Terms: Census Bureau publications, coefficient of variation (CV), confidence interval, custom tabulations, derived statistics, design effect, direct comparison, estimate, implied comparison, information products, margin of error (MOE), metadata, nonsampling error, policy view, sampling error, significance level, standard error, statistical inference, statistical significance, synthetic data, transparency, and working papers.

Requirement E2-1: Throughout all processes associated with reporting results, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement E2-2: All information products must provide accurate and reliable information that promotes transparency and must present that information in an unbiased manner.

1. Information products based on data that have "serious quality issues" are not permitted except under the restrictions in Sub-Requirement F1-5.2 of Statistical Quality Standard F1, *Releasing Information Products*.

   Note: Requirement F1-5 in Statistical Quality Standard F1 describes serious data quality issues.

2. Except as noted below, information products (including each table, graph, figure, and map within an information product, and including stand-alone tables, such as custom tabulations) must include a source statement that:
   a. Indicates the program(s) that provided the data.
   b. Indicates the date of the source data.

   Note: Abstracts and presentation slides do not need source statements.

3. Except as noted below, information products (including tables, graphs, figures, and maps that stand alone) must indicate that the data are subject to error arising from a variety of sources, including (as appropriate) sampling error, nonsampling error, model error, and any other sources of error. Including one of the following in the information product will satisfy this requirement:
   a. An explicit statement indicating that the data are subject to error arising from a variety of sources.
   b. A description of the error sources.
   c. A discussion of the error sources.

   Note: Abstracts and presentation slides do not need to indicate that the data are subject to error.

4. Except as noted below, information products must include a reference (i.e., URL) to the full methodological documentation of the program(s).

   Note: Abstracts and presentation slides do not need to include a reference to the full methodological documentation.

5. All inferences and comparisons of estimates based on sample data must include appropriate measures of statistical uncertainty, such as margins of error, confidence intervals, or p-values for hypothesis tests.

   a. Results that are not statistically significant must not be discussed in a manner that implies they are significant.

   b. The same significance or confidence level must be used throughout an information product. Table A shows the requirements for specific information products:

Table A. Significance and Confidence Levels by Information Product

Information Product	Significance Level	Confidence Level
Census Bureau publications	0.10	0.90
News releases	0.10	0.90
All other information products (e.g., working papers, professional papers, and presentations)	0.10 or less	0.90 or more

   c. Direct comparison statements that are not statistically significant must include a statement conveying the lack of statistical significance, such as:

   "The 90 percent confidence interval for the change includes zero. There is insufficient evidence to conclude that the actual change is different from zero."

   Such a statement may be given in a footnote. For example, "*Sales of nondurable goods were down 0.6 percent (+/- 0.8 %)**." Footnote: "**The 90 percent confidence interval includes zero. There is insufficient evidence to conclude that the actual change is different from zero.*"

   d. The text must clearly state whether each comparison (direct or implied) is statistically significant. This must be done either by:
      1) Using a blanket statement such as, "*All comparative statements in this report have undergone statistical testing, and, unless otherwise noted, all comparisons are statistically significant at the 10 percent significance level,"* and specifically noting any implied comparison statements that are not significant.
      2) Reporting a p-value for each comparison.
      3) Stating whether or not the confidence interval includes 0.

e. Statements of equality between population quantities that are being estimated with sampling error are not allowed. For example, the following statements **are not acceptable**, since they refer to unknown underlying population quantities:

- "*The poverty rate for state A equals the rate for state B.*"
- "*The poverty rate remained statistically unchanged*" (for a comparison across time).

It is acceptable to say that the estimates are "*not statistically different*" or (for comparisons over time) "*statistically unchanged,*" if the difference in the estimates is not statistically significant. For example, the following statements **are acceptable**, since they refer to the estimates of population quantities:

- "*The estimated poverty rate for state A, 8.1 percent (± 0.2), is not statistically different from the estimated poverty rate for state B, 8.1 percent (± 0.2).*"
- "*The estimated poverty rate remained statistically unchanged for non-Hispanic whites at 8.2 percent (± 0.2).*" However, this statement must be accompanied by the abovementioned footnote.

6. Key estimates in the text must be accompanied by confidence intervals or margins of error (MOEs) or their equivalents (e.g., equivalents for Bayesian inferences or for error arising from synthetic data) for the information products indicated in the table below. Providing a URL to these measures of statistical uncertainty is not sufficient.

Table B. Confidence Intervals or MOEs for Key Estimates by Information Product

Information Product	Confidence intervals or MOEs
Census Bureau publications	Required
News releases for the economic data items listed in Appendix E2	Required
News releases for all other data (e.g., economic data items not in Appendix E2, household-level data, and person-level data)	Not required
Abstracts and presentations slides	Not required
All other information products (e.g., working papers and professional papers)	Required

Notes:
(1) In working papers and professional papers, p-values, standard errors, coefficients of variation (CV), or other appropriate measures of statistical uncertainty may be used instead of confidence intervals or MOEs.
(2) If the width of a confidence interval rounds to zero, the interval may be replaced by a statement such as "*The width of the confidence interval for this estimate rounds to zero.*"

7. Except as noted below, information products must include or make available by reference (URL) information that allows users to assess the statistical uncertainty of derived statistics as well as of the estimates themselves. For example,

- Measures of statistical uncertainty (e.g., variances, CVs, standard errors, error arising from synthetic data, or their Bayesian equivalents).
- Methods to estimate the measures of statistical uncertainty (e.g., generalized variance functions or equations and design effects).
- Methods to approximate the measures of statistical uncertainty for derived statistics, such as estimates of change or ratios of estimates.

Notes:
(1) This requirement does not apply to response rates, unless the information product analyzes the response rates or draws conclusions from them.
(2) Abstracts and presentation slides need not make available information on statistical uncertainty. Custom tabulations must provide information on statistical uncertainty as specified in Sub-Requirement E2-2.2, item 4.
(3) Maps need not portray or indicate information on statistical uncertainty, but if not, they must include a URL at which users can access measures of statistical uncertainty and other information about statistical uncertainty.
(4) When information on statistical uncertainty is made available by referencing a URL, the URL must direct users specifically to the location of the information.

8. If needed for readers to assess the results presented, the information product must include:
   a. A discussion of the assumptions made.
   b. The limitations of the data.
   c. A description of the methodology used to generate the estimates.
   d. An explanation of how the methodology and the limitations might affect the results.

9. The information presented must be technically and factually correct.

10. The information must be presented logically and any results must follow from the data and the analysis.

11. Any anomalous findings must be addressed appropriately.

12. The subject matter and methodological literature must be referenced, as appropriate.

13. Policy views must never be expressed.

14. Except as noted in Sub-Requirement E2-2.1 (item 3), personal views must not be expressed.

Sub-Requirement E2-2.1: In addition to the requirements for all information products, the requirements for working papers, professional papers, research reports, presentation slides, handouts for distribution at presentations, and similar products include the following:

1. Except as noted below, a disclaimer must be included on the title page. The author may determine the wording of the disclaimer as long as it indicates that any views expressed

are those of the author and not necessarily those of the Census Bureau. An example of a disclaimer is: "*Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.*"

Note: The disclaimer is not needed for:

- Census Bureau publications, new releases, abstracts, and handouts for advisory committee meetings.
- Information products that are distributed internally.
- Information products that have been reviewed and approved by the Associate Director as not needing a disclaimer because the documents do not contain personal views (e.g., working papers).
- Presentation slides, unless they will be distributed as handouts or published (e.g., in conference proceedings).

2. Working papers published on the Census Bureau's Web site and written entirely by non-Census Bureau individuals (e.g., external researchers at the Census Bureau's Research Data Centers) must incorporate the disclaimer described above, with an additional statement indicating that the Census Bureau has not reviewed the paper for accuracy or reliability and does not endorse its contents or conclusions.

3. Personal views may be expressed only if they are appropriate for the paper or presentation because they are on statistical, methodological, technical, or operational issues.

4. Working papers and professional papers that discuss the results of qualitative research not supported by statistical testing (e.g., based on samples that are not random, are nonrepresentative, or are too small to provide statistical support of the results) must include a caveat explaining why the qualitative methods used do not support statistical testing. The caveat also must address how the findings can (or cannot) be extended to wider populations.

5. Information products based on data with "serious data quality issues" related to nonsampling error may be written only when their purpose is not to report, analyze, or discuss characteristics of the population or economy, but to:
   - Analyze and discuss data quality issues or research on methodological improvements, or to
   - Report results of evaluations or methodological research.

   Note: Statistical Quality Standard F1, *Releasing Information Products* describes serious data quality issues and the restrictions on releasing information products with such issues.

Note: Although not a requirement of the statistical quality standards, the Census Bureau requires presentation slides to use the PowerPoint templates featuring the Census Bureau wordmark provided at the Customer Liaison and Marketing Services Office Intranet Web site.

Sub-Requirement E2-2.2: In addition to the requirements for all information products, the requirements for tabulations include the following:

1. The level of detail for tabulations must be appropriate for the level of sampling error, nonsampling error, and any other error associated with the estimates.

2. All tabulations, except as noted for custom tabulations in item 4 below, must present estimates that take into account the sample design (e.g., weighted estimates).

3. All tabulations, except as noted for custom tabulations in item 4 below, must account for missing or invalid data items (e.g., use imputed data, adjust weights, or display the weighted total of the cases where the data were not reported).

4. Custom tabulations must:
   a. Present weighted estimates unless a client requests unweighted tabulations. If unweighted tabulations are produced for a client, a discussion of the issues associated with using unweighted counts must be provided with the tabulations. Providing a reference (URL) citing the discussion is not sufficient.
   b. Account for missing or invalid data items unless a client requests custom tabulations that exclude imputed data. If tabulations are produced for a client that exclude imputed data, additional metadata must be provided with the tabulations to describe and quantify the level and the extent of the missing data. Providing a reference (URL) citing the metadata is not sufficient.
   c. Include measures of statistical uncertainty (e.g., CVs, standard errors, MOEs, confidence intervals, or their Bayesian equivalents) with weighted tabulations, or include a reference (URL) to the measures of statistical uncertainty. If a program manager thinks that computing estimates of sampling error is not feasible (e.g., for reasons of cost, schedule, or resources), the program manager must work with their research and methodology Assistant Division Chief (ADC) to provide the client with acceptable measures of statistical uncertainty or the means to compute them.

      Note: Although not a requirement of the statistical quality standards, program managers who produce custom tabulations must refer to and follow the requirements of Data Stewardship Policy DS021, *Custom Tabulations*.

5. If any differences are identified (e.g., with a footnote) as statistically significant in any table within an information product, then all statistically significant differences must be similarly identified in all the tables. However, it is not required to identify statistically significant differences in tables.

6. Tabulations must be formatted to promote clarity and comprehension of the data presented.

   Examples of formatting practices that promote clarity and comprehension include:
- Presenting at most four dimensions in a cross-tabulation.
- Labeling all variables.

- Using row or column percentages to reinforce the text description of the relationships involved.
- Labeling the type of statistics being presented (e.g., frequency, percentage, means, and standard errors).
- Presenting totals and subtotals when appropriate.
- Labeling header columns for each page in multi-page tabulations.
- Indicating when a data value is suppressed because of disclosure issues.
- Footnoting anomalous values (e.g., outliers).

7. Displaying estimates that equals zero and symbols in tables must be appropriate for the content/subject matter being presented and according to acceptable statistical practice. An estimate that equals zero should be shown as a numeric value, e.g., 0.00 for two-decimal accuracy. The exception is when the estimate is less than half of a unit of measurement from zero and there is a meaningful difference between an actual zero and a rounded zero for the particular statistics. Use the symbol without additional punctuation such as parenthesis. Use an "X" instead of "(X)".

   Examples of approved standard symbols:
   a. A 'Z' means the estimate rounds to zero.
   b. An 'S' means that the estimate is withheld because estimate did not meet publication standards.
   c. An 'X' means that the estimate is not applicable.
   d. An 'N' means that the estimate is not available or not comparable
   e. A 'D' means that the estimate is withheld to avoid disclosing data for individual companies; data are included in higher level totals

Sub--Requirement E2-2.3: In addition to the requirements for all information products, the requirements for statistical graphs, figures, and maps include the following:

1. The dimensions of graphs, figures, and maps must be consistent with the dimensions of the data (e.g., three-dimensional effects must not be used when displaying only two dimensions of data).

2. Graphs, figures, and maps must be formatted to promote clarity and comprehension of the data presented.

   Examples of formatting practices that promote clarity and comprehension include:
- Labeling axes and including the unit of measure.
- Including a legend that defines acronyms, special terms, and data values.
- Preparing graphs, figures, and maps in the same format throughout the information product.
- Using consistent scales across graphs, figures, or maps that are likely to be compared.
- Using units of measure appropriate to the scale of the graph, figure, or map.
- Starting the base of the graph or figure at zero to avoid giving an inappropriate visual impression.

- Ensuring that color hues correspond to the level of measurement (e.g., a light-to-dark color scheme corresponds with low-to-high values).
- Complying with accessibility requirements of Section 508 of the U.S. Rehabilitation Act.

  Note: The Census Bureau Guideline on the *Presentation of Statistical Graphics* and the Administrative Customer Service Division (ACSD) *Chart Publishing Guidelines* provide additional guidance on presenting graphics.

Appendix E2
Economic Indicator Variables

Direct comparison statements that are not statistically significant are not allowed, except for statements of changes in news releases for the economic indicator variables listed in the table below. In these news releases, the footnote below must be provided to indicate that the comparison is not statistically significant:

> *"The 90 percent confidence interval for the change includes zero. There is insufficient evidence to conclude that the actual change is different from zero."*

Program	Frequency	Levels	Characteristics	Current Period to Prior Period	Current Period to Same Period One Year Ago	Year-to-date to prior year-to-date
Advance Monthly Retail and Food Services Survey (MARTS)	Monthly	• Total retail plus food services • Total retail	Sales	✓	**Monthly:** ✓ **Quarterly:** ✓	✓ **(Dec. only)**
Monthly Wholesale Trade Survey (MWTS)	Monthly	• Total wholesale • Durable goods • Nondurable goods	Sales	✓	✓	
			Inventories	✓	✓	
Quarterly Services Survey (QSS)	Quarterly	• Total • 2-digit sector totals	Receipts or revenue	✓	✓	
			Expenses	✓	✓	
Manufacturing and Trade Inventories and Sales (MTIS)	Monthly	• Total manufacturing, retail and wholesale trade	Distributive trades sales plus manufacturers' shipments and total business inventories	✓	✓	
Building Permits Survey (BPS)	Monthly	• U.S. Total and by size • Region, Total, 1-unit	Authorizations	✓	✓	✓
Survey of Construction (SOC)	Monthly	• U.S. Total and by size • Region, Total, 1-unit	Starts	✓	✓	✓
			Completions	✓	✓	✓

Program	Frequency	Levels	Characteristics	Current Period to Prior Period	Current Period to Same Period One Year Ago	Year-to-date to prior year-to-date
		• U.S. and region, 1-unit	Sales	✓	✓	✓
Value Put-in-Place (VIP) Surveys	Monthly	• Total residential • Total nonresidential • Private total • Private residential • Private nonresidential • Public total • Public residential • Public highway	Construction expenditures	✓	✓	✓
Quarterly Financial Report (QFR)	Quarterly	• Total manufacturing • Durable goods manufacturing • Nondurable goods manufacturing	Seasonally-adjusted: • After-tax profits • Sales • After-tax profits per dollar of sales	✓		
		• Total manufacturing • Durable goods manufacturing • Nondurable goods manufacturing • Total mining • Wholesale trade • Retail trade	Not seasonally-adjusted: • After-tax profits • Sales • After-tax profits per dollar of sales		✓	
E-Commerce	Quarterly	• Total retail	Total sales	✓	✓	
			E-commerce sales	✓	✓	
Housing Vacancies and Homeownership (CPS/HVS)	Quarterly	• U.S. • Regions	Rental vacancy rate, Homeowner vacancy rate, Homeownership rate	✓	✓	

Statistical Quality Standard E3
Reviewing Information Products

Purpose: The purpose of this standard is to ensure that information products released by the Census Bureau receive the appropriate reviews required to ensure they are of high quality and do not disclose protected information or administratively restricted information. This standard also ensures that plans to participate at public events are reviewed and approved.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to the review and approval of:

- Information products, including internal information products that are subsequently released to the public.
- Participation at public events.

Note: Information products (e.g., professional papers, presentations, or other materials) prepared by a Census Bureau employee that pertain to the Census Bureau's programs, policies or operations and are related to the employee's job or area of expertise, are covered by this standard, even if prepared on the employee's own time, without the use of Census Bureau resources or support. See Department Administrative Order (DAO) 219-1, Section 11, Non-Official Public Communications.

Exclusions:

In addition to the global exclusions listed in the Preface, this standard does not apply to:

- Information products prepared or delivered by Census Bureau employees, but which are not related to programs, policies, or operations of the Department of Commerce or the Census Bureau. (Census Bureau employees or Special Sworn Status individuals, who want to include their Census Bureau affiliation as biographical information in the communication, should obtain guidance from the Office of Analysis and Executive Support.)

Key Terms: Census Bureau publications, custom tabulations, direct comparison, disclosure, implied comparison, information products, participation, policy view, public event, and working papers.

Requirement E3-1: All Census Bureau information products must be reviewed before release to ensure that disclosure avoidance techniques necessary to prevent unauthorized release of protected information or administratively restricted information have been implemented completely and correctly. Information protected by federal law (e.g., Title 13, Title 15, and Title 26) and by the Confidential Information Protection and Statistical Efficiency Act of 2002

(CIPSEA) is covered by this requirement. (Statistical Quality Standard S1, *Protecting Confidentiality*, addresses disclosure avoidance techniques.)

Sub-Requirement E3-1.1: The Census Bureau's Disclosure Review Board (DRB) procedures must be followed for information products that use data protected by Title 13 to prevent unauthorized release of protected information or administratively restricted information, particularly personally identifiable information or business identifiable information. (See the DRB Intranet Web site for further guidance and procedures.)

Requirement E3-2: To maintain the Census Bureau's position as unbiased and neutral with regard to policy and political issues, employees must submit an Event Participation Approval Form through their Division Chief to the Chief, International Relations Office, to receive approval to participate in public events within the United States, except for the conferences noted below. Appendix E3-A contains the Event Participation Approval Form. See the Census Bureau's Intranet Web page on participation at public events for further information.

Definitions:
(1) A "public event" means that the event is open to the general public, including events that require a registration fee.
(2) "Participation" means that the employee takes an active role in the event.

Examples of the types of activities that constitute participation and require an Event Participation Approval Form include:

- Presenting a paper or poster at a professional association conference.
- Organizing and/or chairing a session at a professional association conference.
- Acting as a discussant.
- Serving as a panelist.
- Giving a seminar or workshop at colleges, universities, the Washington Statistical Society, or other organizations.
- Making a presentation as an expert member of a working group or other group.
- Staffing a Census Bureau-sponsored booth at a professional association conference or at a trade show.
- Conducting Foreign Trade Division export compliance seminars on the Foreign Trade Regulations and electronic export reporting for U.S. exporters.

Examples of the types of activities that do not constitute participation and do not require an Event Participation Approval Form include:

- Attending a conference or seminar as a member of the audience only.
- Participating in corporate recruiting sponsored by the Human Resources Division, including conducting information sessions and other presentations at colleges or universities.

Examples of events that are not public and do not require an Event Participation Approval Form include:

- Attending a meeting with a survey sponsor.
- Attending private planning meetings for cooperation between statistical agencies.

- Attending meetings sponsored by the Census Bureau to elicit input from survey users.
- Presenting a Center for Statistical Research and Methodology (CSRM) seminar at the Census Bureau.

Notes:
(1) The Event Participation Approval Form is **not** needed for the following conferences:
- Joint Statistical Meetings (JSM) of the American Statistical Association (ASA)
- American Association of Public Opinion Research (AAPOR)
- International Conference on Establishment Surveys (ICES)
- Population Association of America (PAA)
- American Economics Association
- International Statistical Institute (ISI)
- Association of American Geographers (AAG)

(2) Multiple employees that participate in the same session of a conference need submit only one form.
(3) Contact the Chief of the International Relations Office regarding attendance at international events or for questions regarding whether an Event Participation Form must be submitted.

Requirement E3-3: All information products must undergo review and receive approval before they are released to the public, to sponsors, or to other customers. Sub-Requirements E3-3.1 through E3-3.5 describe the types and levels of review needed.

Examples of information products covered by this requirement include, but are not limited to:
- News releases.
- Census Bureau publications (i.e., information products that the program's Associate Director has reviewed and approved and the Census Bureau has affirmed their content).
- Working papers (e.g., technical papers and division reports intended for release to the public).
- Professional papers (e.g., journal articles, book chapters, conference papers, poster sessions, and written discussant comments).
- Research reports used to guide decisions about Census Bureau programs.
- Abstracts.
- Presentations at public events, such as seminars or conferences. (See Requirement E3-4 for additional requirements for presentations.)
- Handouts for distribution at public events.
- Data sets (e.g., public-use files) and their associated documentation.
- Tabulations, including custom tabulations, estimates, and their associated documentation.
- Statistical graphs, figures, and thematic maps.

Notes:
(1) Drafts of information products to be released for limited circulation (e.g., professional papers) outside the Census Bureau are subject to the requirements for a supervisory review as stated in Sub-Requirement E3-3.1. The other reviews (i.e., content/subject

matter, statistical, and policy) are not required unless the supervisor determines that the product needs any of those reviews.

(2) While not a statistical quality requirement, Census Bureau Policy requires that the Chief of the Demo-Econ Media Relations Branch in the Public Information Office (PIO) be informed of any information products or other materials being prepared for public use. (See the Census Bureau Policies and Procedures Manual, Chapter B-13 – *Clearance and Release of Public Information Materials*, for further guidance and procedures.)

Sub-Requirement E3-3.1: All information products must undergo a supervisory review and receive approval.

1. The following table specifies who must perform the supervisory review and approval.

Type of Information Product	**Supervisory Reviewers**
Census Bureau publications	• Author's immediate supervisor • Author's Division or Office Chief • Associate Director of the program releasing the information product
News releases	• Author's immediate supervisor • Author's Division or Office Chief • Associate Director of the program releasing the information product • Associate Director for Communications
All other information products	• Author's immediate supervisor • Author's Division or Office Chief

2. The supervisory reviewer must verify that the following requirements have been met.

<u>All information products</u>

a. The content of the information product is technically and factually correct.
b. All mandated disclosure avoidance procedures have been followed.
c. The provisions for reviewing and releasing information products in any data-use agreements have been followed.
d. The information product complies with the Census Bureau's statistical quality standards.
e. If the information product is a draft to be released for limited circulation outside the Census Bureau, it must include a disclaimer that states the draft is still under review and is not for distribution.

<u>All information products containing text</u>

f. No policy views are expressed in the information product.
g. No personal views are expressed in Census Bureau publications or news releases.
h. Only personal views on statistical, methodological, technical, or operational issues are expressed in information products other than Census Bureau publications and news releases.
i. A disclaimer is included on the title page in all information products except as noted below. The author may determine the wording of the disclaimer as long as it indicates that any views expressed are those of the author and not necessarily those

of the Census Bureau. An example of a disclaimer is: "*Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.*"

Note: The disclaimer is not needed for:

- Census Bureau publications, new releases, abstracts, and handouts for advisory committee meetings.
- Information products that are distributed internally.
- Information products that have been reviewed and approved by the Associate Director as not needing a disclaimer because the documents do not contain personal views (e.g., working papers).
- Presentation slides, unless they will be distributed as handouts or published (e.g., in conference proceedings).

j. The information is presented logically and any results follow from the data and the analysis.
k. Any anomalous findings are addressed appropriately.
l. Correct grammar is used.
m. Presentation slides use the required Census Bureau PowerPoint template found on the Customer Liaison Marketing and Services Office (CLMSO) Intranet Web site. (Note: This is a Census Bureau Corporate Identity Standard.)

Notes:
(1) When the author is a Division or Office Chief, the supervisory reviewer is the author's Associate Director. When the author is a higher-level manager than a Division or Office Chief, the supervisory review is waived.
(2) When the author is a Senior Technical (ST) employee, the supervisory reviewer is the Chief of the Center for Statistical Research and Methodology.

Sub-Requirement E3-3.2: All information products, except data sets and custom tabulations, must undergo a content/subject-matter review and receive approval. However, the documentation that accompanies data sets or custom tabulations must receive a content/subject matter review.

1. The following table specifies who must perform the subject matter review and approval:

Type of Information Product	**Content/Subject Matter Reviewers**
Abstracts	Author's Division or Office Chief
All other information products	Reviewers who are outside the author's organizational unit (branch), and who have expertise in the subject matter, operation, or statistical program discussed in the information product. If a qualified outside reviewer is not available, a reviewer within the author's organizational unit is permitted.

2. The content/subject matter reviewer must verify that the following requirements have been met:

a. The content of the information product is technically and factually correct.
b. The information is presented logically and any conclusions follow from the data and the analysis.
c. Any anomalous findings are addressed appropriately.
d. Subject-matter literature is referenced in the information product, as appropriate.

3. The content/subject matter reviewer must either approve the information product or provide the author with specific written instructions on issues to be revised.

4. The content/subject matter reviewer must review the information product again after the author addresses any recommended revisions. If the reviewer and the author disagree with how the comments are addressed, they must inform their supervisors so that a resolution can be reached.

Note: If an information product is generated from a program sponsored by an outside organization or uses data provided by an outside organization, the author's Division or Office Chief should determine whether to send the product to the outside organization for an additional review.

Sub-Requirement E3-3.3: All information products must undergo a statistical review and receive approval, even if the author believes the information product involves no statistical methodologies.

1. The following table specifies who must perform the statistical review and approval:

Type of Information Product	**Statistical Reviewers**
Conference papers	Reviewers who have expertise in the statistical methodology or program discussed in the information product Note: Appendix E3-B provides a list of statistical review contacts for conference papers.
Abstracts	Author's Division or Office Chief Note: If the Division or Office Chief determines that an abstract requires a more rigorous statistical review, he or she must refer the abstract to the appropriate Research and Methodology Assistant Division Chief (ADC).
All other information products	Research and Methodology ADC of the program related to the topic of the information product Note: Appendix E3-B provides a list of statistical review contacts by topic/subject matter.

2. The statistical reviewer must verify that the following requirements have been met:
a. The discussion of assumptions and limitations is accurate and appropriate.
b. The description of the reliability of the data is accurate and complete.

c. Statistical testing is performed correctly to support any comparison statements, whether expressed directly or implied.
d. Calculations and equations are accurate and statistically sound.
e. The content, conclusions, and any recommendations on technical, statistical, or operational issues are supported by the methodology used and the data presented.
f. A source statement is included in the information product. (See Requirement E2-2, item 2, in Statistical Quality Standard E2, *Reporting Results*.)
g. Statistical uncertainty is appropriately conveyed.
h. Comparison statements, such as historical comparisons, are appropriate.

3. The statistical reviewer must either approve the information product or provide the author with specific written instructions on issues to be revised.

4. The statistical reviewer must review the information product again after the author addresses any recommended revisions. If the reviewer and the author disagree on how the comments are addressed, they must inform their supervisors so that a resolution can be reached.

Notes:
(1) Media releases that do not contain estimates or discussions of statistical or survey methodology need not undergo a statistical review (e.g., media advisories such as the one titled, "Census Bureau Releases Timetable for 2008 Income, Poverty and Health Insurance Estimates and American Community Survey Data").
(2) Two types of geographic products need not undergo a statistical review:
   a) Thematic maps presenting data from the census short form if the underlying data have been reviewed and approved.
   b) Geographic reference products (e.g., reference maps, and documents showing lists and numbers of geographic entities and relationships between entities).

Sub-Requirement E3-3.4: All information products involving methodologies other than statistical must undergo a methodological review and receive approval.

1. The review must be conducted by individuals with expertise in the methodologies used in the information product (e.g., cognitive psychology, economics, demographic analysis, geographic information systems, or any other specialized methodology).
2. The methodological reviewer must either approve the information product or provide the author with specific written instructions on issues to be revised.
3. The methodological reviewer must review the information product again after the author addresses any recommended revisions. If the reviewer and the author disagree on how the comments are addressed, they must inform their supervisors so that a resolution can be reached.

Sub-Requirement E3-3.5: All information products must undergo a policy and sensitivity review by the author's Division or Office Chief. The Division Chief or Office Chief may not delegate this review.

Notes:
(1) Appendix E3-C provides a checklist developed by the Office of Analysis and Executive Support (OAES) to assist in the policy and sensitivity review. If the Division or Office Chief needs guidance on a specific issue, he or she may refer the issue to the Associate Director, the OAES, the Congressional Affairs Office (CAO), or the PIO, as appropriate.
(2) When the author is a Division or Office Chief or higher-level manager, the policy and sensitivity review is at the discretion of the author's supervisor.

Requirement E3-4: All presentations (with or without a paper) to be delivered by Census Bureau staff at meetings and conferences open to the public (including advisory and data user meetings) must undergo a dry run rehearsal.

1. A senior division manager (ADC or higher) must attend the dry run.
2. All reviewers must be invited to the dry run.
3. Authors must provide copies of their presentations and any other relevant materials to everyone invited, in advance of the dry run.

Notes:
(1) Presentations that have had a dry run and are simply being repeated at another venue do not need another dry run unless substantive changes have been made to the presentation.
(2) The dry run is optional for Division or Office Chiefs or higher and for Senior Technical (ST) employees, at the discretion of their supervisors.

Requirement E3-4.1: Authors of informal presentations (e.g., presentations without written remarks or audio-visual aids, including unwritten discussant or panelist remarks) must review their approach with their Division or Office Chief.

Note: When the author is a Division Chief, Office Chief, or other higher-level manager, this review is at the discretion of the author's supervisor.

Requirement E3-5: The results of the review and approval of information products must be documented, either electronically or on paper, and the documentation retained according to division or directorate policies and procedures.

Examples of documentation include:
- Completed approval forms.
- Approval e-mail messages from reviewers.

Appendix E3-A
Event Participation Approval Form and Instructions

Employees must submit the Event Participation Approval Form before making a firm commitment to participate in public events in order to:

- Ensure that the Census Bureau maintains neutrality on all policy and political issues and avoids any appearance of partiality on those issues.
- Keep the Deputy Director and the Executive Staff informed of staff participation in public events.

Note: This form is **not** needed for the following conferences:

- Joint Statistical Meetings (JSM) of the American Statistical Association (ASA)
- American Association of Public Opinion Research (AAPOR)
- International Conference on Establishment Surveys (ICES)
- Population Association of America (PAA)
- American Economics Association
- International Statistical Institute (ISI)
- Association of American Geographers (AAG)

In addition, this form does **not** apply to international events. Contact the Chief of the International Relations Office regarding attendance at international events, including events in U.S. territories.

Instructions

The participating employee must:

1. Provide the following information on the form:
    - The name, sponsor, date, location, and description of the event.
    - The topic of the session or panel. Complete a form for each session of a conference that has Census Bureau participants – one form for the entire conference is not sufficient if Census Bureau employees participate in more than one session.
    - The names and affiliations of all participants in the session or panel.
    - The audience anticipated.
        - Indicate whether the event is open to the public or limited to invited guests, organization members, and/or paying attendees.
        - Indicate whether media and congressional, federal, state, or local government representatives are expected to attend.
2. Obtain approval from their Division Chief.
3. Submit the completed form to the Chief, International Relations Office.

Event Participation Approval Form

Participant Name:	
Date of Invitation:	**Date Submitted:**
Event Name and Location:	
Event Date:	
Event and Sponsor Description: (e.g., event theme, organization mission statement, website address)	
Description of Panel, Session, or Discussion Topic:	
Invited Participants: (Include affiliation of each participant)	
Audience: (Indicate whether the meeting is open to the public or limited to invited guests, organization members, or paying attendees) (Indicate whether media and congressional, federal, state, or local government representatives are expected to attend)	
Attached Materials: (If any)	

Division Chief Approval _____

Submit completed forms to the Chief, International Relations Office

Appendix E3-B
Statistical Review Contacts

Statistical Review Contacts by Topic / Subject Matter

Topic / Subject Matter	Contact
Census 2000 / 2010	Assistant Division Chief (ADC) for Sampling and Estimation (DSSD)
American Community Survey (ACS)	ADC for ACS Statistical Design (DSSD)
Demographic Surveys (e.g., CPS, NHIS, and SIPP)	ADC for Sample Design and Estimation (DSMD)
Small Area Estimates	Chief, Center for Statistical Research and Methodology
Administrative records data	Chief, Center for Administrative Records Research and Applications (CARRA)
Economic Programs data	Chief, Office of Statistical Methods and Research for Economic Programs (OSMREP) or appropriate Research and Methodology ADC
International data, multiple data sources, or other data	Appropriate ADC for Research and Methodology for the author's directorate

Statistical Review Contacts for Conferences

Conference	Contact
Joint Statistical Meetings (JSM)	Chief, Center for Statistical Research and Methodology
American Association of Public Opinion Research (AAPOR)	Chief, Center for Statistical Research and Methodology
International Conference on Establishment Surveys (ICES)	Chief, Office of Statistical Methods and Research for Economic Programs (OSMREP)
Population Association of America (PAA)	ADC for Sample Design and Estimation (DSMD)
Federal Committee on Statistical Methodology (FCSM)	Chief, Center for Statistical Research and Methodology
American Economics Association (AEA)	Chief, Center for Economic Studies (CES)
Census Advisory Committee or National Academy of Sciences	Appropriate ADC for Research and Methodology for the author's directorate
International Statistical Institute (ISI)	Chief, Center for Statistical Research and Methodology
International Union for the Scientific Study of Population (IUSSP)	Chief, Demographic Statistical Methods Division (DSMD)
Other international conferences, including the UN Economic Commission for Europe and Organization for Economic Cooperation and Development	Appropriate ADC for Research and Methodology for the author's directorate
Association of American Geographers (AAG)	Appropriate ADC for Research and Methodology for the author's directorate
SAS Users	Appropriate ADC for Research and Methodology for the author's directorate
All other conferences	Appropriate ADC for Research and Methodology for the author's directorate

Appendix E3-C
Policy and Sensitivity Review Checklist for Division and Office Chiefs

This checklist should be used to determine the suitability for publication and release of Census Bureau information products.

If the answer to any of the following questions is "yes," then the information product proposed for publication/release must not be released until the issue raised by the question has been resolved appropriately.

1. Is the information product inconsistent with the Census Bureau's mission?

2. Would publication/release of the information product compromise the Census Bureau's ability to perform its mission?

3. Does the information product express views on or discuss any of the following topics in an inappropriate manner or in a way that is inconsistent with laws, Commerce Department policies, or Census Bureau policies?

   a. Laws, regulations, *Federal Register* notices, court cases, congressional testimony, or policy statements or decisions pertaining to the Commerce Department or the Census Bureau.

Examples include:

- Sections of the Commerce Department's Code of Federal Regulations.
- Chapters of the Census Bureau's Policies and Procedures Manual.
- Census Bureau's Data Stewardship Policies.
- Census Bureau's Information Technology Security Policies and Regulations.

   b. The Freedom of Information Act or the Privacy Act.

   c. Matters that are currently being investigated by Congress.

   d. Issues relating to privacy, confidentiality, data security, or access to and use of administrative records (including any issues related to personally or business identifiable information or data breaches).

   e. Budget/appropriations issues.

   f. Any issue that is politically sensitive or that has been the subject of recent news articles, correspondence, hearings, or current or potential lawsuits.

Examples of sensitive issues include:

- Current poverty estimates.

- Concerns about the American Community Survey (ACS).
- Concerns about Local Update of Census Addresses Program (LUCA).
- Concerns about the enumeration of sensitive populations (racial or ethnic populations, immigrants, the homeless, or Group Quarter's (GQ) populations such as prisoners, residents of nursing homes, or college students).
- Concerns about the enumeration of overseas Americans.
- Concerns about statistical sampling or adjustment of decennial census counts.

g. Sensitive historical issues like the internment of Japanese Americans or statistical adjustment of the decennial census.

4. Is it possible that release of the information product will affect any national policy issues related to the topics it discusses?

5. Does the information product discuss matters related to sharing confidential Title 13 and/or Title 26 information/data in a way that suggests the sharing is inconsistent with laws, Census Bureau policies, or IRS policies?

6. Does the information product suggest or imply that the Census Bureau may be cooperating in any way with an enforcement, regulatory, or administrative activity of another government agency?

An example would be a discussion of providing tabulations of public-use data to a federal law enforcement agency. It would be acceptable to discuss the Census Bureau's policy to encourage the agency to perform the tabulations and to inform the agency that any tabulations provided by the Census Bureau are subject to public disclosure.

7. Does the information product discuss specific contract/acquisitions issues or information in a manner that improperly discloses commercial proprietary information or trade secrets?

8. Does the information product single out a particular group or category of individuals to receive special treatment, consideration, or recognition (e.g., identifying key partners who contributed to the decennial census effort) in a manner that might compromise the Census Bureau's ability to perform its mission?

9. Does the information product contain any subject matter or language that might be deemed offensive, insensitive, or inappropriate?

10. Does the information product lack the disclaimer (if required) indicating that the information product represents the author's views (on statistical, methodological, technical, or operational issues) and does not necessarily represent the position of the Census Bureau? (Statistical Quality Standard E2, *Reporting Results*, specifies when the disclaimer is required.)

Note: If the disclaimer is required but missing, the author **must** add it before the information product may be published or released.

RELEASING INFORMATION

F1 Releasing Information Products

Appendix F1: Dissemination Incident Report

F2 Providing Documentation to Support Transparency in Information Products

F3 Addressing Information Quality Guideline Complaints

Appendix F3: Procedures for Correcting Information that Does Not Comply with the Census Bureau's Information Quality Guidelines

Statistical Quality Standard F1
Releasing Information Products

Purpose: The purpose of this standard is to establish quality criteria for releasing information products.

The OMB's Statistical Policy Directive No. 3 and Statistical Policy Directive No. 4 describe requirements for notifying the public of the release of information products. The Census Bureau's Product Release Notification Policy and Policies and Procedures Manual (Chapter B-13 – *Clearance and Release of Public Information Materials*) describe procedures for notifying the PIO about information products to be released to the public.

Note: Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to ensure transparency in information products released outside the Census Bureau.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

Exclusions:
In addition to the global exclusions listed in the Preface,
(1) Requirements F1-2 and F1-3 of this standard do not apply to:
- Professional papers, presentations, and similar information products.
- Information products delivered to sponsors or clients (e.g., data files and tabulations).

(2) Requirements F1-7 through F1-10 of this standard do not apply to:
- Professional papers, presentations, and similar information products.

Key Terms: Coefficient of variation (CV), coverage ratio, dissemination, estimate, information product, metadata, nonresponse bias, nonsampling error, releases of information products, response rate, sample design, and sampling error.

Requirement F1-1: Neither protected information nor administratively restricted information may be released outside the Census Bureau, except as allowed under applicable federal laws (e.g., Title 13, Title 15, and the Confidential Information Protection and Statistical Efficiency Act) and data-use agreements.

Sub-Requirement F1-1.1: Throughout all processes associated with releasing information products, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Data Stewardship Policies), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement F1-2: Information products released to the public by the Census Bureau must be released according to a dissemination plan that addresses:

1. What information product(s) are planned for release.
2. The release schedule. The release schedule for all regular or recurring information products for the upcoming year must be published on www.census.gov before January 1 of that year. (See OMB Statistical Policy Directive No. 4.)
3. The reviews and approvals needed before releasing the information products to the public.
4. The mode of release by the Census Bureau.

Requirement F1-3: Policies and procedures for disseminating information products, including those related to any planned data revisions or any corrections for data quality issues identified after an information product has been released, must be documented and published on the Census Bureau's Internet Web site.

Requirement F1-4: Information products must not be released outside the Census Bureau until they receive the appropriate reviews and approvals. (See Statistical Quality Standard E3, *Reviewing Information Products*.)

Requirement F1-5: Embargoed news releases and data files must not be released to the public by any means (including print, broadcast, Internet, podcast, blogs, or in any other form) before the specified date and time of release. (See the U.S. Census Bureau Embargo Policy.)

Requirement F1-6: Information products must comply with the Census Bureau's statistical quality standards and must be free of serious data quality issues in order to be released outside the Census Bureau without restrictions.

1. Serious data quality issues related to sampling error occur when the estimated coefficients of variation (CV) for the majority of the key estimates are larger than 30 percent.

   Notes:
   (1) This requirement does not apply to secondary estimates. For example, if the estimated month-to-month change is the key estimate, and the monthly estimates are secondary, the requirement applies only to the estimated month-to-month change.
   (2) Statistical Quality Standard A1, *Planning a Data Program*, provides requirements for identifying key estimates.

2. Serious data quality issues related to nonsampling error occur when:
   a. All products:
      1) The data suggest that the primary survey concepts are not clearly defined or that measurement of the concepts failed for some reason.
      2) The key estimates are inconsistent with our base of knowledge about the characteristic being estimated.

3) Issues that are serious enough to raise concerns about the accuracy of the data occur in sample design, sampling methods, questionnaire or forms design, data collection, data processing, estimation procedures, or the underlying assumptions of a model.

b. Products derived primarily from census or survey data:
   1) Unit response rates for surveys or censuses, or cumulative unit response rates for panel or longitudinal surveys, are below 60 percent.
   2) Sample attrition from one wave to the next wave in panel or longitudinal surveys is greater than five percent.
   3) Item response rates or total quantity response rates on key items are below 70 percent.
   4) Coverage ratios for population groups associated with key estimates are below 70 percent.
   5) Combined rates for key estimates (e.g., computed as unit response × item response × coverage) are below 50 percent.

   Notes:
   (1) These thresholds are provided because bias is often associated with low response rates or with low coverage ratios. If nonresponse bias analyses or other studies show that the bias associated with nonresponse is at an acceptable level, or that steps taken to mitigate nonresponse bias or coverage error are effective, these thresholds do not apply.
   (2) The Census Bureau conducts a few surveys that do not use probability samples. Generally, they are establishment surveys that select the largest units in the target universe and do not attempt to collect data from the small units in the universe. For these surveys, the above thresholds do not apply. These surveys have serious data quality issues if the responding units do not comprise at least 70 percent of the target universe, based on the unit response rate or the total quantity response rate, as appropriate.
   (3) Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*, specifies requirements on computing response rates.

Sub-Requirement F1-6.1: Information products with data free from the serious data quality issues described in Requirement F1-6 may be released outside the Census Bureau with no restrictions, subject to confidentiality constraints.

Sub-Requirement F1-6.2: Information products with data that have any of the serious data quality issues in Requirement F1-6 may be released outside the Census Bureau only under the restrictions described below.

1. Restrictions for information products with serious data quality issues related to sampling error:

   The information product must:
   a. Note that the CV exceeds 0.30 for a majority of the key estimates.

b. Note that data users should exercise caution when using estimates with high sampling error.
c. Indicate why the data are being released (e.g., aggregates of the estimates may be useful or the knowledge that the estimates have extremely high magnitude or extremely low magnitude may be useful).

2. Restrictions for information products with serious data quality issues related to nonsampling error:

   a. Products that are Census Bureau publications or regular or recurring products (i.e., products governed by Statistical Policy Directive No. 3 or Statistical Policy Directive No. 4):
      1) The program manager must obtain a waiver before releasing the information product.
      2) The information product must summarize any nonsampling error issues related to Requirement F1-6, item 2a (1 through 3).
      3) If response rates, coverage ratios, or the combined rates fall below the thresholds in Requirement F1-6, item 2b:
         i. The key estimates affected must be identified.
         ii. A table must be included that provides the response rates or coverage ratios for key estimates in enough detail to allow users to evaluate how the issue may affect their use of the data. Other quantitative measures of the impact of the issue should be included to the extent feasible.
      4) The information product must include details about the potential impact of the quality issues on the data.
      5) The information product must include the URL of the complete documentation on the nonsampling error issues.

   b. Products released to sponsors:
      1) The information product must summarize any nonsampling error issues related to Requirement F1-6, item 2a (1 through 3).
      2) If response rates, coverage ratios, or the combined rates fall below the thresholds in Requirement F1-6, item 2b:
         i. The key estimates affected must be identified.
         ii. A table must be included that provides the response rates or coverage ratios for key estimates in enough detail to allow users to evaluate how the issue may affect their use of the data. Other quantitative measures of the impact of the issue should be included to the extent feasible.
      3) The information product must include details about the potential impact of the quality issues on the data.
      4) The delivery of the product to the sponsor must include the complete documentation on the nonsampling error issues or a URL where the documentation is accessible.

c. Products that are not Census Bureau publications or are not regular or recurring products (e.g., custom tabulations, data files, professional papers, working papers, technical reports, and similar products):
   1) Release to the public is not allowed, except as noted in item 2) below. The information product may be released only on request. If released on request, the information product must:
      i. Include this disclaimer: "*These data are being released on request, despite concerns about their quality. The Census Bureau's policy is not to withhold data that are available, unless releasing such data would violate confidentiality requirements. The Census Bureau recommends using these data only for research or evaluation purposes, and not to make statements about characteristics of the population or economy because they do not meet the criteria outlined in the Census Bureau's Statistical Quality Standard: Releasing Information Product.*"
      ii. Summarize the nonsampling error issues.
      iii. Include summary metadata describing the issues and the impact on the data.
      iv. Provide the URL of the complete documentation on the nonsampling error issues.
   2) Release is permitted only for information products whose purpose is not to report, analyze, or discuss characteristics of the population or economy, but whose purpose is to:
      - Analyze and discuss data quality issues or research on methodological improvements, or to
      - Report results of evaluations or methodological research
   3) External researchers at the Census Research Data Centers may not have access to confidential data that are affected by serious data quality issues, except to analyze the data quality issues, including developing potential solutions. If the researcher has corrected the data quality issues and the Census Bureau has determined that the researcher's solutions are appropriate, the revised data may be used for subject-matter (e.g., poverty) analyses.

Requirement F1-7: When a data quality issue that might be serious is suspected in a previously released information product, the program manager must notify Census Bureau senior management of the issue immediately after it has been identified. At a minimum, the senior managers to be notified include:

1. The Division Chief(s) responsible for the program with the suspected data quality issue.
2. The Associate Director responsible for the program with the suspected data quality issue.

Note: These senior managers will decide whether the issue should be escalated to the Deputy Director and provide guidance on the appropriate actions to take and the specific stakeholders or organizations to notify regarding the suspected data quality issue.

Requirement F1-8: When serious data quality issues are identified in a previously released information product, a notification must be disseminated to alert the public. If the product was released to a sponsor, the notifications must be made to the sponsor.

1. The notification must be disseminated immediately after identifying a serious data quality issue, even if the issue is not yet fully understood.
    a. If appropriate the data affected by the data quality issue must be removed from the Census Bureau's Internet Web site at this time.
2. The notification must include the following components, with additional information that facilitates understanding the issue and its effects as appropriate:
    a. A description of the issue.
    b. A description of what is known about the effect on the data.
    c. A description of what is known about the cause.
    d. A statement indicating the data have been removed until the issue has been fixed (if appropriate).
    e. Plans for addressing the issue.
    f. Expected release dates of revised products.
3. If the notification is disseminated before the issue is fully understood, it must be updated when a more complete understanding is achieved.

    Note: Program managers must notify the responsible Division Chief(s) and Associate Director (Requirement F1-7) before making notifications to the public or sponsors.

Requirement F1-9: Any serious error or data quality issue identified in a previously released information product must be addressed appropriately.

Examples of appropriate actions to address serious errors and data quality issues include:

- Correct the error and re-release the product.
- Release an "errata" document for the product, describing the error and the correction.
- If it is not feasible to correct an error, release a description of the error and its likely effects on the program's estimates and results.
- If a data user or a sponsor reported the error, acknowledge the report and indicate when the issue is expected to be resolved. If the error will not be corrected, respond and explain to the user why it will not be corrected and what actions will be taken to address the error.

Sub-Requirement F1-9.1: Serious errors or data quality issues identified in a previously released information product must be documented by completing the Dissemination Incident Report found in Appendix F1 and submitting it to the Quality Program Staff.

Requirement F1-10: Information products approved for release to the public must be published on the Census Bureau's Internet Web site and must adhere to the requirements of Section 508 of the U.S. Rehabilitation Act.

Appendix F1
Dissemination Incident Report

Purpose: This report documents the nature of incidents involving the dissemination of information products with serious inaccuracies or other quality-related problems and the factors associated with those incidents as required by Sub-Requirement F1-8.1 in Statistical Quality Standard F1 (Releasing Information Products). The information gathered in this report is needed to identify common factors that contribute to disseminating inaccurate information products and to help prevent future incidents. **This report is not a substitute for your Directorate's procedures for dealing with dissemination incidents.**

The dissemination incident report has two parts:

- An initial report (see Table A) which documents the incident. Program managers complete this initial report.
- A detailed report (see Tables B, C, and D) which gathers information to promote understanding of the factors associated with the incident and its impact. The Quality Program Staff will interview program managers to gather this information.

Scope: Please prepare a dissemination incident report for:

- Releases of information products with errors or other quality-related problems (release to the public, release to sponsors, or release to other agencies within the Commerce Department).
- "Near misses" in which detection of an error, outside the normal review process, prevented the release of an erroneous product.
- Releases (or "near misses") of data without adequate disclosure avoidance measures.

Note: If a product is sent to a sponsor and an error is found as part of the normal review process, it is out of scope. However, products released with errors identified after the review would be in scope.

Instructions: Contact the Quality Program Staff (QPS) at dir.quality.program.staff.list@census.gov or on 301-763-6598 if you have any questions regarding these instructions.

1. Within one week of discovering the incident, complete the initial dissemination incident report, Table A, on page 2 and send it to the QPS at dir.quality.program.staff.list@census.gov. Do NOT report any specifics about disclosure avoidance procedures.
2. The QPS will contact you to schedule an interview to be held approximately 2 weeks after receipt of the initial report.
3. Review and answer the questions in the detailed dissemination incident report in Tables B, C, and D and be prepared to discuss these questions and answers when the QPS conducts the interview with you.
4. After the interview, review the dissemination incident report generated by the Quality Program Staff for accuracy and completeness, and provide comments to the QPS.
5. After three months, the QPS will follow-up with you to find out what preventative measures were taken as a result of the incident (see Table E, section II of the detailed dissemination incident report).
6. Determine and implement actions to prevent recurrence of the incident. Report these actions and the actual or planned dates of implementation to the QPS when they contact you in the follow-up.
7. The QPS will submit a summary report based on all the incident reports to the Program Associate Directors.

Initial Dissemination Incident Report

Please provide the following information regarding the incident in the table below and email the completed table to the Quality Program Staff at dir.quality.program.staff.list@census.gov.

TABLE A: DOCUMENTATION OF THE INCIDENT

Provide the following:	
A1. Contact Information (name and phone number)	
A2a. Directorate	
A2b. Division (use acronym)	
A3. Program or Survey Name (list all affected)	
A4a. Name of specific information products affected by the incident.	
A4b. Type(s) of specific information products affected by the incident. Select all that apply: a) Tables b) Microdata or summary files c) Reports (e.g., publications, working papers, summary brief, documentation, highlights) d) News release e) Other – please specify	
A5. Description of the incident	
A6. Incident type – Select from the following: a) Inaccurate data released b) Incomplete data released c) Wrong file released d) Improper disclosure (including releasing information prematurely) e) Geographic error f) Display error (e.g., incorrect heading or symbol) g) Other – please specify	
A7. Date the incident was detected	
A8. Description of the problem(s) that generated the incident	

TABLE A: DOCUMENTATION OF THE INCIDENT

Provide the following:	
A9. Problem type(s) – Select all that apply: a) Needed action not performed b) Wrong action performed c) Action not performed correctly d) Action performed out of sequence e) Wrong data file used f) Wrong variables used g) Communication h) Other – please specify	
A10. Date the problem(s) that generated the incident occurred	
A11. How the incident was detected	
A12. Who discovered / reported the incident – Select from the following: a) User – public b) User – sponsor c) User – Congressional or Commerce d) Bureau staff – within Branch e) Bureau staff – within Division f) Other Bureau staff g) Other - specify	
A13. Who was notified of the incident (within the Bureau and outside) – Select all that apply: a) Your Division Chief b) Your Associate Director c) Deputy Director d) Other Census Bureau divisions – specify e) BEA f) Commerce Undersecretary g) Other contact outside the Census Bureau – specify h) None of the above	
A14. Frequency of the incident (i.e., Is this the only occurrence? Has it happened before?)	

TABLE A: DOCUMENTATION OF THE INCIDENT

Provide the following:	
A14.1 Did the incident affect multiple releases of a recurring product?	
A15. Your preliminary assessment of the severity of the incident using a scale of 1 for minor to 5 for extremely severe	
a) Explain the basis of your assessment (What you know at this point – e.g., How many cases were affected? By how much were the estimates overstated or understated?)	
A16. The immediate actions taken to address the incident – Select all that apply: a) Notified sponsor b) Notified selected users c) Removed data from web d) Posted user note on web e) Posted revised data f) Other – specify	
A17. How many person-hours did it take to complete this table?	Count the time spent by anyone who participated in completing Table A.

*****After completing the initial report (Table A), please remember to review and answer the questions in the detailed report on the following pages and be prepared to discuss these questions and answers when the Quality Program Staff interviews you.*****

Detailed Dissemination Incident Report

I. Processes and Factors Associated with the Incident

Review the questions in the detailed dissemination incident report and be prepared to answer these questions when the Quality Program Staff conducts the interview with you. This information will promote understanding of the incident and its impact.

TABLE B: ORIGIN OF THE INCIDENT

Question	Y/N /NA	Obtain details from program manager
B0. Please keep track of how many person-hours it takes to review the questions in Tables B, C, and D and prepare for the interview with the Quality Program Staff. Please record the number of person-hours in Question D6.		Count the time spent by anyone who participated in reviewing Tables B, C, and D.
In what process did the incident originate?		
B1. Planning / development a) Stakeholder input / concepts to measure b) Instrument development c) Pretesting / testing d) Frame and sample development / selection e) Interview mode / timing f) Other – specify		
B2. Collecting / acquiring data a) Interviewing b) Transmitting c) Quality checks (i.e., error originated in the QC check) d) Other – specify		
B3. Capturing and processing data a) Data entry / electronic capture b) Geographic processing c) Editing d) Imputation		

TABLE B: ORIGIN OF THE INCIDENT

Question	Y/N /NA	Obtain details from program manager
e) Coding f) Geocoding g) Record linkage h) Quality checks (i.e., error originated in the QC check) i) Other – specify		
B4. Producing estimates and measures a) Weighting / post collection adjustment b) Tabulation c) Creating Microdata file d) Creating Summary file e) Variance estimation f) Modeling / seasonal adjustment g) Quality checks / analyst data review h) Other – specify		
B5. Analyzing data / reporting results a) Data analysis b) Report writing / Production of tables c) Quality checks / review of products d) Other – specify		
B6. Releasing information products a) Dissemination b) User documentation c) Other – specify		
B7. Protecting Confidentiality a) Disclosure avoidance b) Other – specify		
B8. Other – specify		

TABLE C: FACTORS ASSOCIATED WITH THE INCIDENT

Question	Y/N /NA	Explain the factors that contributed to the incident
Procedures		
C1. Did any of these factors regarding existing procedures contribute to the incident:		
a) Inadequate production procedures?		
b) Inadequate change control procedures?		
c) Inadequate version control procedures?		
d) Inadequate quality control procedures?		
C2. Did any of these factors contribute to the incident:		
a) A lack of documented procedures?		
b) Procedures that were not up-to-date?		
c) Procedures that were not followed properly?		
d) Lack of tools to ensure that procedures were followed (e.g., a checklist)?		
Requirements		
C3. Did inadequate or incomplete requirements for the processes where the problem occurred contribute to the incident?		
C4. Did the failure of requirements to reflect program needs contribute to the incident?		

TABLE C: FACTORS ASSOCIATED WITH THE INCIDENT

Question	Y/N /NA	Explain the factors that contributed to the incident
C5. Did any of these factors contribute to the incident:		
a) A lack of documented requirements?		
b) Requirements were not kept up-to-date?		
c) Requirements were not followed?		
Specifications		
C6. Did inadequate or incomplete specifications for the processes where the problem occurred contribute to the incident?		
C7. Did any of these factors contribute to the incident:		
a) A lack of documented specifications?		
b) Specifications that were not kept up-to-date?		
c) Specifications that were not followed?		
Computer Programming and Implementation		
C8. Did software errors contribute to the incident (e.g., mistakes in computer code)?		
C9. Did any of these factors contribute to the incident:		
a) Computer programs did not accurately reflect the specifications?		
b) Computer programs were not run in the correct order?		
c) Incorrect versions of the computer files were used?		
Methods		
C10. Did the application of inappropriate or incorrect methods contribute to the incident (e.g., use of inappropriate analysis methods)?		
C11. Did the use of suboptimal methods		

TABLE C: FACTORS ASSOCIATED WITH THE INCIDENT

Question	Y/N /NA	Explain the factors that contributed to the incident
contribute to the incident (e.g., methods that are labor intensive and carry risk of introducing errors)?		
C12. Did mistakes occur during the application of manual or clerical methods? (e.g., copy and paste errors, data entry errors, manual geographic edits, forgetting to perform step)		
Quality Control		
C13. Did the failure to perform any of these quality checks contribute to the incident:		
a) Test systems to ensure that they function as intended?		
b) Test and implement process or system changes?		
c) Monitor operations (e.g., monitoring data collection)?		
d) Perform the required reviews of the information products (e.g., supervisory, statistical, content, and policy)?		
e) Check the accuracy of data, results, etc.		
C13.1 Did a quality check fail to catch the error?		
Communication		
Communication Internal to the Census Bureau		
C14. Did failures in handoffs contribute to the incident?		

TABLE C: FACTORS ASSOCIATED WITH THE INCIDENT

Question	Y/N /NA	Explain the factors that contributed to the incident
C15. Did any of these communications failures contribute to the incident?		
a) Failure to communicate responsibilities?		
b) Failure to communicate changes to the people who need to know (e.g., changes to procedures, specifications, or requirements)?		
c) Failure to communicate institutional knowledge? (e.g., staff no longer work in program area and no documentation exists for staff taking over)		
d) Failure of other internal communications?		
C16. Did any of the following factors contribute to the incident?		
a) Misinterpretation of procedures?		
b) Misinterpretation of requirements?		
c) Misinterpretation of specifications?		
C17. Did inadequate training contribute to the incident?		

TABLE C: FACTORS ASSOCIATED WITH THE INCIDENT

Question	Y/N /NA	Explain the factors that contributed to the incident
Communication with Entities Outside the Census Bureau		
C18. Did insufficient communication to stakeholders (e.g., ESA, Commerce, sponsors, and users) contribute to the incident? For example, o Data correct, but apparent anomalies were not fully explained. o Data released with known problems, but discussion of these problems was inadequate.		
Resources		
C19. Did any of these factors contribute to the incident?		
a) Excessive time constraints?		
b) Budget limitations?		
C20. Did misaligned skills of staff contribute to the incident?		
C21. Did inadequate staff resources contribute to the incident?		
C22. Did the experience levels of staff or managers contribute to the incident?		
C23. Did conflicting priorities contribute to the incident?		
Other		
C24: Please select one of the following:		
a) First time the process (or procedure) was performed		
b) Change in the normal process		
c) Normal process (been done before)		
C25. What other factors contributed to the incident?		

TABLE D: FINAL ASSESSMENT OF THE INCIDENT

Question	
D1. Now that you know more about the incident, what is your final assessment of the severity of the incident using a scale of 1 for minor to 5 for extremely severe?	
D2. Explain the basis for this assessment (e.g., How many sample cases were affected? By how much were the estimates overstated or understated? What level of geography was affected?)	
D3. What actions were (will be) taken to prevent recurrence of the problems that resulted in the incident and when were (will) they implemented?	
D4. Beyond the vulnerabilities that contributed to this incident, what other operational gaps are you aware of that increase the risk for additional incidents?	
D5. (Answer if identified operational gaps in D4) What do you think you need to close those gaps (e.g., resources, software, and hardware)?	
D6. How many person-hours did it take to review the questions in Tables B, C, and D and prepare for the interview with the Quality Program Staff?	Count the time spent by anyone who participated in reviewing Tables B, C, and D.

II. Follow-up

After three months, the Quality Program Staff will follow-up with program managers to find out what preventative measures were taken as a result of the incident.

TABLE E: PREVENTATIVE MEASURES

Question	
E1. What actions were (will be) taken to prevent recurrence of the problems that resulted in the incident and when were (will) they implemented?	
E2. What actions were (will be) taken to detect incidents prior to release and when were (will) they implemented?	

Statistical Quality Standard F2
Providing Documentation to Support Transparency in Information Products

Purpose: The purpose of this standard is to specify the documentation that must be readily accessible to the public to ensure transparency and reproducibility in information products released by the Census Bureau.

The documentation required by this standard aims to provide sufficient transparency into the Census Bureau's information products so that qualified users can reproduce the estimates and results in the products. However, federal law (e.g., Title 13, Title 15, and Title 26) and Census Bureau policies require safeguarding the confidentiality of protected information or administratively restricted information. Therefore, complete transparency and reproducibility may not always be possible. At a minimum, the documentation will allow users to assess the accuracy and reliability of the estimates and results in the Census Bureau's information products.

Note: Statistical Quality Standard F1, *Releasing Information Products*, addresses the required documentation and metadata to describe any serious data quality problems and the likely effects of the problems on the data and estimates in the Census Bureau's information products.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

Exclusions:
The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Key Terms: Administratively restricted information, data program, information product, protected information, qualified user, readily accessible, reproducibility, and transparency.

Requirement F2-1: Documentation that would breach the confidentiality of protected information or administratively restricted information or that would violate data-use agreements with other agencies must not be released. (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement F2-2: Documentation must be readily accessible in sufficient detail to allow qualified users to understand and analyze the information and to reproduce (within the constraints of confidentiality requirements) and evaluate the results. The documentation must be made readily accessible by doing one or more of the following:

1. Including the documentation in the information product if it is necessary for readers to understand the results.

2. Referencing the full methodological documentation in the information product (e.g., providing a URL) and publishing the documentation on the Census Bureau's Internet Web site.
3. Delivering the full methodological documentation to the sponsors of reimbursable programs or providing them with a URL to the documentation.

Note: The Census Bureau Geospatial Product Metadata Standard (GPMS), and the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) provide additional requirements for geospatial products.

Sub-Requirement F2-2.1: Descriptions of the data program must be readily accessible.

Examples of information that describes the data program include:
- The purpose of the program (e.g., survey, census, evaluation study, or research).
- The organizational sponsor(s) of the program.
- The organization that conducted the program.
- The data source (e.g., organization or agency) and the database or systems from which the data are drawn for administrative records data.
- The universe of inference or target population for the program.

Sub-Requirement F2-2.2: Descriptions of the concepts, variables, and classifications that underlie the data must be readily accessible.

Examples of concepts, variables, and classifications that underlie the data include:
- Definitions of the primary concepts being measured.
- The wording of questions asked in surveys or censuses.
- Identification of the key variables.
- Descriptions of the concepts underlying all variables.
- Geographic levels of the data.
- The reference dates for the data and for the geographic levels.
- Descriptions of any derived measures.

Sub-Requirement F2-2.3: Descriptions of the methodology, including the methods used to collect and process the data and to produce estimates, must be readily accessible.

Examples of documentation of the methodology include:
- Discussion of methods employed to ensure data quality.
- Quality profiles. (See the Census Bureau Guideline on *Quality Profiles*.)
- Documentation of pretesting of the data collection instruments, including qualitative studies.
- Source and accuracy statement.
- Description of the sampling frame.
- Description of the sample design.
- The size of the sample.
- Information on eligibility criteria and screening procedures.

- Description of sample weights, including adjustments for nonresponse.
- The mode and methods used to collect the data.
- The dates of data collection.
- Description of any bounding methods used to control telescoping.
- Description of estimation procedures, including weighting, editing, and imputation methods.
- Reasons for not imputing the data when imputation for item nonresponse is not carried out.
- Description of how to calculate variance estimates.
- Discussion of potential nonsampling errors (e.g., nonresponse, coverage, processing, and measurement).
- Discussion of the methods to approximate the standard errors of derived statistics.
- Description of any substantial changes in procedures or methodology over time and the known impact on the data.
- References to methodological documentation maintained by the source organization supplying administrative records data.
- Model description, including assumptions and type of model.
- Equations or algorithms used to generate estimates.
- Description of seasonal adjustment methods. (See the Census Bureau Guideline on *Seasonal Adjustment Diagnostics*.)
- Description of small area estimation methods.
- Any limitations or data quality problems affecting the estimates or projections.
- Descriptions of known data anomalies and corrective actions.

Sub-Requirement F2-2.3.1: Measures and indicators of the quality of the data must be readily accessible.

Examples of measures and indicators of the quality of the data include:

- The disposition of sample cases (e.g., numbers of interviewed cases, ineligible cases, and nonresponding cases).
- Unit response rates or quantity response rates.
- Item response rates, item allocation rates, total quantity response rates, or quantity response rates for key data items.
- Rates for the types of nonresponse (e.g., refusal, unable to locate, no one home, temporarily absent, language problem, insufficient data, and undeliverable as addressed).
- Coverage ratios.
- Indicators of the statistical precision of the estimates (e.g., estimates of sampling variances, standard errors, coefficients of variation, or confidence intervals).
- Coverage of the target population by the set of administrative records.
- The proportion of administrative records that have missing data items or that contain invalid data for key variables.
- The proportion of data items with edit changes because the data items were invalid or otherwise required changes.
- The proportion of records lost from the analysis or estimate due to nonmatches when linking data sets.

- Effects on the estimates related to coverage issues, nonmatches in record linking, and missing data items in surveys, censuses, or administrative records.
- Model diagnostics (e.g., goodness of fit, coefficient of variation, and percent reduction in confidence interval of the direct estimates).

Note: Statistical Quality Standard D3, *Producing Measures and Indicators of Nonsampling Error*, contains requirements on producing measures and indicators of nonsampling error.

Sub-Requirement F2-2.3.2: The methodology and results of evaluations or studies of the quality of the data must be readily accessible.

Examples of evaluations or studies of the quality of the data include:

- Nonresponse bias analyses.
- Evaluation studies (e.g., evaluation studies of response error, interviewer variance, respondent debriefing, record check or validation, and mode effects).
- Response analysis surveys.
- Comparisons with independent sources, if available.
- Match analyses.
- Reconciliations (e.g., a comparison of import and export data).
- Periodic summaries of quality control results (e.g., interviewer quality control (QC) results and error rates measured by data entry QC and coding QC).

  Note: Results of routine reviews and verifications need not be readily accessible unless needed for data users to assess the quality of the information product.

Sub-Requirement F2-2.4: Documentation of public-use data files must be readily accessible in sufficient detail to allow a qualified user to understand and work with the files.

Examples of documentation of public-use data files include:

- File description.
- File format (e.g., SAS file or text file).
- Variable names and descriptions (e.g., data dictionary or record layout).
- Data type for each variable (e.g., numeric, alphanumeric, and length).
- Description of variables used to uniquely identify records in the data file.
- Description of flags to indicate missing and imputed items.

Statistical Quality Standard F3
Addressing Information Quality Guideline Complaints

Purpose: The purpose of this standard is to ensure that complaints alleging that information products are not in compliance with the Census Bureau's Information Quality Guidelines are addressed.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to information products released by the Census Bureau for which a party outside the Census Bureau alleges that the Census Bureau has not adhered to its information quality guidelines.

Exclusions:
In addition to the global exclusions listed in the Preface, this standard does not apply to:
- Information released by the Census Bureau before October 1, 2002.

Key Terms: Information products, information quality, and releases of information products.

Requirement F3-1: Complaints must be reviewed by the program manager responsible for the information product being challenged.

Note: The Census Bureau Information Quality Web site contains the correction procedures complainants must follow to submit complaints for information they believe does not comply with the Census Bureau's Information Quality Guidelines.

Requirement F3-2: Except as noted below, program managers must follow the procedure in Appendix F3 to investigate and resolve complaints.

Note: These programs have developed correction procedures specific to their information products and must follow their own correction procedures. The appeals process, when not separately defined in the program's procedures, will be managed as stated in Appendix F3.
- Count Question Resolution (CQR).
- Local Update of Census Addresses (LUCA).
- Governmental Unit Boundaries.
- Street and Address Range Information.
- Small Area Income and Poverty Estimates (SAIPE).
- Annual Estimates of the Total Population.
- Foreign Trade Statistics.

Requirement F3-3: Corrected information must be readily accessible on the Census Bureau's Internet Web site (www.census.gov) and subsequent issues of recurring information products, including subsequent annual reports, must reflect the corrected data.

Note: Because the Information Quality Guidelines under which these corrections will occur are for statistical information released after October 1, 2002, any correction of historical data suggested by a complaint with which the Census Bureau concurs will be performed at the discretion of the program area.

Requirement F3-4: In the case of a serious error that could potentially mislead policy makers, any published reports containing the erroneous data must be reissued.

Requirement F3-5: Complaints and the resulting actions must be documented by the program manager and submitted to the Chair of the Methodology and Standards Council.

Appendix F3
Procedures for Correcting Information that Does Not Comply with the Census Bureau's Information Quality Guidelines

The following procedures must be followed when complaints alleging that the Census Bureau has not adhered to its information quality guidelines are received.

Note: These procedures do not apply to the seven programs listed in Requirement F3-2 of Statistical Quality Standard F3, *Addressing Information Quality Guideline Complaints*. Those programs follow their own correction procedures that are specific to their data products.

1. The Census Bureau's Quality Program Staff will notify the Department of Commerce within ten business days of receiving a complaint that alleges a violation of the information quality guidelines.

2. The program manager must review:
   a. The information being challenged in consultation with the appropriate methodology staff.
   b. The processes that were used to create and disseminate the information.
   c. Whether the information conforms or does not conform to the Census Bureau's Information Quality Guidelines.

3. Based on the outcome of the above review, the Census Bureau will determine if a correction (or corrections) must be made.

4. If the Census Bureau concurs with a complaint, the responsible program manager will, with the concurrence of the area Associate Director in consultation with the Methodology and Standards Council, determine the appropriate corrective action, taking into account such factors as:
   - The nature of the information involved.
   - The significance and magnitude of the error with respect to the use of the information.
   - The cost of implementing the correction.
   - The effectiveness of the correction in terms of timeliness.

5. The Census Bureau will respond in writing to the affected person within 60 days of receiving the complaint.
   a. If the Census Bureau has completed its review, the response will explain the process that the Census Bureau followed in its review of the complaint, the findings of the review, and the resolution.
   b. If the Census Bureau has not completed its review, the response will notify the affected person that a review is underway, and provide an expected completion date. When the review is complete, the Census Bureau must again contact the affected person in writing, and explain the process that the Census Bureau followed in its review of the complaint, the findings of the review, and the resolution.
   c. If a correction is warranted, the response will include a progress report, and a subsequent written response will be sent when the correction action is complete.

d. If a correction is not warranted, the Census Bureau will explain that a correction will not be made, and why.

6. If the Census Bureau declines to correct the challenged data, and the affected party appeals, a panel appointed by the Methodology and Standards Council will manage the appeal process.
   a. The Census Bureau will respond to all requests for appeals within 60 days of receipt.
   b. If the appeal requires more than 60 days to resolve, the Census Bureau will inform the appellant that more time is required, indicate the reason why, and provide an estimated decision date.

SUPPORTING STANDARDS

Statistical Quality Standard S1
Protecting Confidentiality

Purpose: The purpose of this standard is to ensure the confidentiality of protected information and administratively restricted information.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status (SSS) individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to:

- Data collected from respondents and protected under Title 13.
- Data protected under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).
- Data collected under Title 15 and protected by legislation governing sponsoring agencies.
- Administrative records provided by source agencies, such as Federal Tax Information (FTI) protected under Title 13 and Title 26.

Exclusions:

The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Key Terms: Administratively restricted information, bottom-coding, business identifiable information, cell suppression, confidentiality, controlled rounding, controlled tabular adjustment, disclosure, noise infusion, personally identifiable information, protected information, random rounding, recoding, swapping, synthetic data, and top-coding.

Requirement S1-1: All Census Bureau employees and SSS individuals must follow the provisions of federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Information Technology (IT) Security policies and Data Stewardship policies, such as DS018 *Unauthorized Browsing Policy* and DS022 *Data Breach Policy*), and data-use agreements to prevent unauthorized release of protected information and administratively restricted information.

Sub-Requirement S1-1.1: Neither protected information nor administratively restricted information may be released outside the Census Bureau, except as allowed under applicable federal laws (e.g., Title13, Title 15, and CIPSEA) and data-use agreements.

Requirement S1-2: Disclosure avoidance techniques must be used to prevent unauthorized release of protected information and administratively restricted information, particularly personally identifiable information or business identifiable information.

Examples of disclosure avoidance techniques include:

- Random rounding.
- Controlled rounding.
- Top-coding.
- Bottom-coding.
- Recoding.
- Data swapping.
- Generating synthetic data.
- Noise infusion.
- Using rules to define sensitive cells (e.g., thresholds).
- Protecting sensitive cells (e.g., cell suppression, random rounding, controlled rounding, collapsing cells, and controlled tabular adjustment).

Notes:

(1) Contact the Census Bureau's Disclosure Review Board (DRB) for guidance on disclosure avoidance techniques.

(2) Sub-Requirement E3-1.1 of Statistical Quality Standard E3, *Reviewing Information Products*, addresses requirements for disclosure avoidance review.

(3) Statistical Policy Working Paper 22: *Report on Statistical Disclosure Limitation Methodology*, published by the Office of Management and Budget's Federal Committee on Statistical Methodology, provides information on various techniques to prevent disclosure of protected information.

Statistical Quality Standard S2
Managing Data and Documents

Purpose: The purpose of this standard is to ensure that data and documentation internal to the Census Bureau are appropriately managed (i.e., files are retained, secured, and accessible to authorized users) to promote the transparency and reproducibility of Census Bureau processes and products, and to inform future projects and improvement efforts.

> Note: Statistical Quality Standard F2, *Providing Documentation to Support Transparency in Information Products*, contains specific requirements about documentation that must be readily accessible to the public to ensure transparency in information products released by the Census Bureau.

Scope: The Census Bureau's statistical quality standards apply to all information products released by the Census Bureau and the activities that generate those products, including products released to the public, sponsors, joint partners, or other customers. All Census Bureau employees and Special Sworn Status individuals must comply with these standards; this includes contractors and other individuals who receive Census Bureau funding to develop and release Census Bureau information products.

In particular, this standard applies to activities related to managing Census Bureau data and documentation needed to replicate results (e.g., models or survey estimates) from research and evaluation studies, surveys, censuses, and administrative records.

> **Exclusions:**
> The global exclusions to the standards are listed in the Preface. No additional exclusions apply to this standard.

Key Terms: Administratively restricted information, protected information, reproducibility, and transparency, and version control.

Requirement S2-1: Throughout all processes associated with managing data and documents, unauthorized release of protected information or administratively restricted information must be prevented by following federal laws (e.g., Title 13, Title 15, and Title 26), Census Bureau policies (e.g., Information Technology (IT) Security policies and Data Stewardship policies, such as DS007 *Information Security Management Program*), and additional provisions governing the use of the data (e.g., as may be specified in a memorandum of understanding or data-use agreement). (See Statistical Quality Standard S1, *Protecting Confidentiality*.)

Requirement S2-2: A plan for data and document management must be developed that addresses:

1. Individuals and divisions responsible for managing the data and documents.
2. Data and documents to be managed.
3. Technical issues relevant to managing the data and documents (e.g., media, retention periods, storage locations, user access rules, version control, file naming conventions, and inventory of files retained).

4. Special operations needed to store and access information (e.g., scanning, encrypting, or compressing data).
5. Timetables for reviewing retained files to verify their usefulness and readability in the stored format (e.g., every five years).

Note: The Disposition of Federal Records: A Records Management Handbook provides guidance on establishing, managing, and operating a records disposition program within a Federal agency. The Census Bureau Guideline on the *Long-Term Backup of Research and Evaluation Files* and the ACSD records management Intranet page provide additional guidance on managing data files.

Requirement S2-3: Data and documentation needed to replicate and evaluate program or research results must be retained according to Census Bureau policies (e.g., Census Bureau Records Schedules, Records Management Policies in the Census Bureau's Policies and Procedures Manual, and division-level policies), data-use agreements with providers of administrative records, and appropriate Federal records disposition and archival regulations (e.g., National Archives and Records Administration's (NARA) statutes).

Examples of data and documentation to retain include:

- Data files and description of variables.
- Planning and design decisions, including the OMB (Office of Management and Budget) Information Collection Request package.
- Analysis plans.
- Field test design and results.
- Cognitive or usability testing results.
- Sampling plan and justifications, including the sampling frame used and any deviations from the plan.
- Justifications for the items on the survey instrument, including why the final items were selected.
- Instructions to respondents and interviewers.
- Description of the data collection and data processing methodologies.
- Questionnaire images.
- Description of the weighting and estimation methods, including variance estimation.
- Description of the imputation and data editing methodologies.
- Specifications and computer code (e.g., specifications and code for sampling, editing, weighting, imputation, analysis, variance estimation, and tabulation).
- Description of models used for estimates and projections.
- Documentation of disclosure avoidance techniques.
- Quality measures, including the equations and interpretations of the measures.
- Evaluation reports, including special evaluations such as nonresponse bias analyses and interviewer variance studies.
- Publicly available documentation associated with the release of data.

Sub-Requirement S2-3.1: An inventory must be developed and maintained to allow authorized users to identify and access the retained data and documents.

Note: The Census Bureau Guideline on the *Long-Term Backup of Research and Evaluation Files* provides information on producing an inventory to explain retained data and documents to potential users.

WAIVER PROCEDURE

Introduction

The Census Bureau's statistical quality standards apply to all Census Bureau's information products and the programs that develop and release those products, as described in the Scope statement in the Preface to these standards. If a program is not complying or anticipates that they may be unable to comply with any requirements of these standards, the program manager must apply for a waiver.

This waiver procedure provides a consistent mechanism to excuse a program from compliance with a statistical quality standard. Waivers will be granted when the circumstances warrant it however, no waivers to Statistical Quality Standard S1, *Protecting Confidentiality*, will be granted.

This procedure promotes proper management and control in implementing the standards and ensures that appropriate documentation of exceptions to the standards is generated and maintained. This documentation is important for providing transparency into the quality of the Census Bureau's information products and for informing future revisions of the statistical quality standards.

Procedure

1. The affected program manager, in collaboration with the program area's M&S Council representative, must prepare a waiver request using the form *Request for a Waiver of a Statistical Quality Standard*.

   The program manager must:
   - Indicate the Program(s)/Information Product(s) to be exempted by the waiver.
   - Indicate the specific requirement(s) to be waived.
   - Describe the noncompliance issue
   - Describe any anticipated effects that may result from the noncompliance.
   - Explain why the program area is not able to comply with the specific requirements of the standard.
   - Describe any actions to be taken to mitigate the effects of noncompliance.
   - Describe the corrective actions planned to achieve compliance. Include milestones dates for key accomplishments including the date the program(s) / information product(s) will be brought into compliance

2. The program manager must email the waiver request to the Quality Program Staff to review for completeness and accuracy.

3. After correcting any issues noted by the Quality Program Staff, the program manager and M&S Council representative must sign the waiver request and submit the completed waiver request to the subject matter Division Chief for concurrence.

4. The Division Chief will review the waiver request and, if concurring, sign the request and forward it to the Quality Program Staff.

5. The Quality Program Staff will schedule the waiver for review by the M&S Council and, as appropriate, by additional stakeholders.

6. The M&S Council will review the waiver request and concur or not concur with the request, noting any recommendations regarding their position.

7. The M&S Council will determine which Associate Director(s) need to approve the waiver request, based on the requirements being waived and the Program(s) / Information Product(s) involved.

8 The Quality Program Staff will forward the waiver request and the Council's recommendation to the appropriate Associate Director(s) accountable for the quality of the Program(s) / Information Product(s).

9. The Associate Director(s) will approve or deny the waiver and return the waiver request to the Quality Program Staff.

10. The Quality Program Staff will ensure that the Program Manager, Division Chief, and the M&S Council receive a copy of the approved or denied waiver request.

11. The Quality Program Staff will publish approved waiver requests on the M&S Council Intranet page.

12. The Quality Program Staff will maintain records of all waiver requests and their resolutions and use them to inform future revisions of the standard.

13. If the waiver is granted, the program manager must develop a corrective action plan and implement the corrective actions described in the waiver request, within the timeline stated on the waiver request. If the corrective actions will not be implemented on time, another waiver must be requested.

14. The Quality Program Staff must follow-up on the implementation of the corrective action plan and periodically report on the progress to the M&S Council.

15. After the corrective action has been completed, the Quality Program Staff will notify the M&S Council and update the M&S Council Intranet page to indicate when the program came into compliance.

Questions

If you have questions regarding the waiver procedure or whether a waiver is needed, contact the Quality Program Staff or the appropriate M&S Council representative.

Waiver for Quality Standard Requirement

Affected Program(s) / Information Product(s):
<Indicate the specific Program(s) / Information Product(s) to be exempted by this waiver.>

Requirement(s) to be waived:
<Requirement I.D.– Text of Requirement to be waived.
For example:
E2-2 #3a – Except as noted below, information products (including tables, graphs, figures, and maps that stand alone) must indicate that the data are subject to error arising from a variety of sources, including (as appropriate) sampling error, nonsampling error, model error, and any other sources of error. Including one of the following in the information product will satisfy this requirement:
a. An explicit statement indicating that the data are subject to error arising from a variety of sources.
b. A description of the error sources.
c. A discussion of the error sources.
Note: Abstracts and presentation slides do not need to indicate that the data are subject to error.>

Noncompliance:
Describe how the program area is or will not be in compliance.
<Requirement I.D.– Text of describing noncompliance
For Example:
E2-2 #3a – The information product for our mobile application does not indicate that the data are subject to error.>

Anticipated effects:
Describe any anticipated effects that may result from the noncompliance.
<Requirement I.D.– Text of anticipated effects.
For Example:
E2-2 #3a –Users will not be informed about errors associated with the data>

Justification:
Explain why the program area is not able to comply with the specific requirement.
<Requirement I.D.– Text justifying the noncompliance
For Example :
E2-2 #3a – There is limited screen real estate to display the notice for the mobile application.>

Mitigating Actions:
Describe any actions being taken to mitigate the effects of noncompliance.
<Requirement I.D.– Text of Requirement
For Example:
E2-2 #3a –We are working with the vendor of the application to explore options for displaying the information.>

Corrective Action Plan:
Describe the corrective actions planned to achieve compliance. Include milestones dates for key accomplishments including the date when the Program(s) / Information Product(s) will be brought into compliance.
<Requirement I.D.– Milestone date – Key accomplishment
For Example:
E2-2 #3a –
5/12/12 – Determine the appropriate placement of the notice within the mobile application and submit change requests to the vendor to correct the display.
5/30/12 – Corrective actions completed and Information Products brought into compliance.>

Waiver for Quality Standard Requirement

Program Manager: *<Signature>* *<Program Manager>* *<Program Manager Title>*	**Date:** *<mm/dd/yy>*

M&S Council Representative: *<Signature>* *<M&S Council Representative>* *<M&S Council Representative Title>*	**Date:** *<mm/dd/yy>*

Subject Matter Division Chief: *<Recommendations as appropriate.>*	***<Concurs / Does not Concur>***
<Signature> *< Subject Matter Division Chief Name >* *< Subject Matter Division Chief Title >*	**Date:** *<mm/dd/yy>*

Methodology and Standards Council: *<Recommendations as appropriate.>*	***<Concurs / Does not Concur>***
<Signature> *<Name of M&S Council Chairperson>* *<Title of M&S Council Chairperson >* Chair, Methodology and Standards Council	**Date:** *<mm/dd/yy>*

Associate Director: *<Recommendations as appropriate.>*	***<Approved / Denied>***
<Signature> *<Name of Associate Director >* *<Title of Associate Director>*	**Date:** *<mm/dd/yy>*

Statistical Quality Standards
GLOSSARY

-A-

Accuracy 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 at the same time. The difference between a sample result and the result from a complete census taken under the same conditions and at the same time is an indication of the precision of the sample result.

Administrative records and **administrative record data** refer to micro data records contained in files collected and maintained by administrative or program agencies and commercial entities. Government and commercial entities maintain these files for the purpose of administering programs and providing services. Administrative records (e.g., Title 26 data) are distinct from systems of information collected exclusively for statistical purposes, such as data from censuses and surveys that are collected under the authority of Titles 13 or 15 of the United States Code (U.S.C.). For the most part, the Census Bureau draws upon administrative records developed by federal agencies. To a lesser degree, it may use information from state, local, and tribal governments, as well as commercial entities. To obtain these data, the Census Bureau must adhere to a number of regulatory requirements.

The **Administrative Records Tracking System (ARTS)** is an electronic database on the Census Bureau's Intranet. It tracks Census Bureau administrative records agreements, agreement commitments, administrative data projects, and relevant external contacts.

Administratively restricted information (as defined in Data Stewardship Policy DS007, *Information Security Management Program*) consists of agency documentation that is not intended as a public information product and other pre-release or embargoed public information. Examples of administratively restricted information include:

- "For Official Use Only" (FOUO) information: Internal Census Bureau documentation consisting of program or operational materials (e.g., contracting, financial, budget, security, legal, policy documents) determined by management to be either protected under the Freedom of Information Act and/or of a nature that release could negatively impact the mission of the Census Bureau.
- Embargoed data or reports that have not been released, but meet Disclosure Review Board requirements for public release.
- Proprietary contractor information, such as its cost proposal and labor rates.
- All information not otherwise protected by statutory authority, but that is subject to access and/or use restrictions, as provided in a valid Agreement with the government agency or other entity supplying the information.
- All personally identifiable information (PII) not protected by an existing legal authority.
- All business identifiable information (BII) not protected by an existing legal authority.

Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices populated by donors, to impute for missing values.

American National Standards Institute codes (ANSI codes) are a standardized set of numeric or alphabetic codes issued by the American National Standards Institute (ANSI) to ensure uniform identification of geographic entities through all federal government agencies.

The **autocorrelation function** of a random process describes the correlation between the processes at different points in time.

Automated record linkage is the pairing of data, primarily via computer software.

An **autoregressive integrated moving average (ARIMA)** model is a generalization of an autoregressive moving average or (ARMA) model for nonstationary time series. A nonstationary time series is a time series not in equilibrium about a constant mean level. In a nonstationary time series, the mean or variance of the series may not be the same at all time periods. The model is generally referred to as an ARIMA(p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated (differencing), and moving average parts of the model respectively.

An **autoregressive moving average (ARMA)** model is a stationary model of time series data where the current data point and current stochastic error are each modeled as finite linear regressions of previous data points or stochastic errors respectively. The regression for the data points is referred to as an autoregression. The regression for the stochastic errors is referred to as a moving average. Symbolically, the model is denoted as an ARMA (p,q) model where p and q are integers greater than or equal to zero and refer to the order of the autoregressive and moving average parts of the model respectively. A stationary time series is a time series in equilibrium about a constant mean level. These models are fitted to time series data either to better understand the data or to predict future points in the series.

-B-

Behavior coding of respondent/interviewer interactions involves systematic coding of the interaction between interviewers and respondents from live or taped field or telephone interviews to collect quantitative information. When used for questionnaire assessment, the behaviors that are coded focus on behaviors indicative of a problem with the question, the response categories, or the respondent's ability to form an adequate response.

Bias is the difference between the expected value of an estimator and the actual population value.

Blocking is grouping the records of a set into mutually exclusive, exhaustive pieces by using a set of fields (e.g., state, last name, first initial). Usually used in the context of record linkage.

Bonferroni correction is a method used to address the problem of multiple comparisons. It is based on the idea that if an experimenter is testing n dependent or independent hypotheses on a set of data, then one way of maintaining the family-wise error rate is to test each individual hypothesis at a statistical significance level of 1/n times what it would be if only one hypothesis were tested.

Bottom-coding is a disclosure limitation technique that involves limiting the minimum value of a variable allowed on the file to prevent disclosure of individuals or other units with extreme values in a distribution.

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

Business identifiable information is information defined in the Freedom of Information Act (FOIA) as trade secrets or commercial or financial information, that is obtained from a person representing a business entity, and which is privileged and confidential (e.g., Title 13) and exempt from automatic release under FOIA. Also included is commercial or other information that, although it may not be exempt from release under the FOIA, is exempt from disclosure by law (e.g., Title 13). Also see **Personally identifiable information**.

-C-

The **calibration** approach to estimation for finite populations consists of: (a) a computation of weights that incorporate specified auxiliary information and are restrained by calibration equation(s); (b) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units; (c) an objective to obtain nearly design unbiased estimates as long as nonresponse and other nonsampling errors are absent.

Cell suppression is a disclosure limitation technique where sensitive cells are generally deleted from a table and flags are inserted to indicate this condition.

A **census** is a data collection that seeks to obtain data directly from all eligible units in the entire target population. It can be considered a sample with a 100 percent sampling rate. The Economic Census may use administrative records data rather than interviews for some units.

Census Bureau publications are information products that are backed and released by the Census Bureau to the public. "Backed and released by the Census Bureau" means that the Census Bureau's senior management officials (at least through the Associate Director responsible for the product) have reviewed and approved the product and the Census Bureau affirms its content. Because publications do not contain personal views, these information products do not include a disclaimer.

Clerical record linkage is record matching that is primarily performed manually.

A **cluster** is a set of units grouped together on the basis of some well-defined criteria. For example, the cluster may be an existing grouping of the population such as a city block, a hospital, or a household; or may be conceptual such as the area covered by a grid imposed on a map.

Coding is the process of categorizing response data using alphanumeric values so that the responses can be more easily analyzed.

Coefficient of variation (CV) is a measure of dispersion calculated by dividing the standard deviation of an estimate by its mean. It is also referred to as the relative standard error.

Cognitive interviews are used as a pretesting technique consisting of one-on-one interviews using a draft questionnaire to find out directly from respondents about their problems with the questionnaire. In a typical cognitive interview, respondents report aloud everything they are thinking as they attempt to answer a survey question.

Computer-assisted personal interviewing (CAPI) is an interviewing technique similar to computer-assisted telephone interviewing, except that the interview takes place in person instead of over the telephone. The interviewer sits in front of a computer terminal and enters the answers into the computer.

Computer-assisted telephone interviewing (CATI) is an interviewing technique, conducted using a telephone, in which the interviewer follows a script provided by a software application. The software is able to customize the flow of the questionnaire based on the answers provided, as well as information already known about the participant.

A **confidence interval** is a range of values determined in the process of estimating a population parameter. The likelihood that the true value of the parameter falls in that range is chosen in advance and determines the length of the interval. That likelihood is called the confidence level. Confidence intervals are displayed as (lower bound, upper bound) or as *estimate* ± *MOE*, where *MOE = z-value * standard error of the associated estimate* (when the confidence level = 90%, the *z-value* = 1.645).

Confidence level is the probability that an assertion about the value of a population parameter is correct.

Confidence limits are the upper and lower boundaries of the confidence interval.

Confidentiality involves the protection of personally identifiable information and business identifiable information from unauthorized release.

Controlled rounding is a form of random rounding, but it is constrained to have the sum of the published entries in each row and column equal the appropriate published marginal totals.

Controlled tabular adjustment is a perturbative method for statistical disclosure limitation in tabular data. This method perturbs sensitive cell values until they are considered safe and then rebalances the nonsensitive cell values to restore additivity.

A **convenience sample** is a nonprobability sample, from which inferences cannot be made. Convenience sampling involves selecting the sample from the part of the population that is convenient to reach. Convenience sampling is not allowed for Census Bureau information products.

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

Coverage refers to the extent to which elements of the target population are listed on the sampling frame. **Overcoverage** refers to the extent that elements in the population are on the frame more than once and **undercoverage** refers to the extent that elements in the population are missing from the frame.

Coverage error which includes both undercoverage and overcoverage, is the error in an estimate that results from (1) failure to include all units belonging to the target population or failure to include specified units in the conduct of the survey (undercoverage), and (2) inclusion of some units erroneously either because of a defective frame or because of inclusion of unspecified units or inclusion of specified units more than once in the actual survey (overcoverage).

A **coverage ratio** is the ratio of the population estimate of an area or group to the independent estimate for that area or group. The coverage ratio is sometimes referred to as a coverage rate and may be presented as percentage.

Cross-sectional studies (also known as cross-sectional analysis) form a class of research methods that involve observation of some subset of a population of items all at the same time. The fundamental difference between cross-sectional and longitudinal studies is that cross-sectional studies take place at a single point in time and that a longitudinal study involves a series of measurements taken on the same units over a period of time. See **Longitudinal survey**.

Cross-validation is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis.

Custom tabulations are tables prepared by the Census Bureau at the request of a data user or program sponsor. This terminology does not apply to tables produced by Census Bureau software (e.g., FERRET or American Fact Finder).

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 capture is the conversion of information provided by a respondent into electronic format suitable for use by subsequent processes.

Data collection involves activities and processes that obtain data about the elements of a population, either directly by contacting respondents to provide the data or indirectly by using administrative records or other data sources. Respondents may be individuals or organizations.

Data collection instrument refers to the device used to collect data, such as a paper questionnaire or computer assisted interviewing system.

A **data program** is a program that generates information products, often on a regular schedule. These programs include efforts such as the censuses and surveys that collect data from respondents. Data programs also include operations that generate information products from administrative records and operations that combine data from multiple sources, such as various surveys, censuses, and administrative records. Specific examples of multiple source data programs include the Small Area Income and Poverty Estimates (SAIPE) program, the Population Division's "Estimates and Projections" program, the National Longitudinal Mortality Study, and the Annual Survey of Manufactures (ASM). One-time surveys also are considered data programs.

Data-use agreements for administrative records are signed documents between the Census Bureau and other agencies to acquire restricted state or federal data or data from vendors. These are often called Memoranda of Understanding (MOU).

Derived statistics are calculated from other statistical measures. For example, population figures are statistical measures, but population-per-square-mile is a derived quantity.

The **design effect** is the ratio of the variance of a statistic, obtained from taking the complex sample design into account, to the variance of the statistic from 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.

A **direct comparison** is a statement that explicitly points out a difference between estimates.

Direct estimates are estimates of the true values of the target populations, based on the sample design and resulting survey data collected on the variable of interest, only from the time period of interest and only from sample units in the domain of interest. Direct estimates may be adjusted using explicit or implicit models (e.g., ratio adjustment, hot or cold deck imputation, and non-response adjustment) to correct for nonresponse and coverage errors.

Disclosure is the release of personally identifiable information or business identifiable information outside the Census Bureau.

Dissemination means Census Bureau-initiated or sponsored distribution of information to the public (e.g., publishing information products on the Census Bureau Internet Web site). Dissemination does not include distribution limited to government employees or agency contractors or grantees; intra-agency or inter-agency use or sharing of government information; and response to requests for agency records under the Freedom on Information Act, the Privacy

Act, or other similar law. This definition also does not include distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas, or adjudicative processes.

A **dress rehearsal** is a complete test of the data collection components on a small sample under conditions that mirror the full-implementation. See **Field test**.

-E-

Editing is the process of identifying and examining missing, invalid, and inconsistent entries and changing these entries according to predetermined rules, other data sources, and recontacts with respondents with the intent to produce more accurate, cohesive, and comprehensive data. Some of the editing checks involve logical relationships that follow directly from the concepts and definitions. Others are more empirical in nature or are obtained through the application of statistical tests or procedures.

Equivalent quality data is data obtained from another source than the respondent, which have quality equivalent to data reported by the respondent. Equivalent quality data have three possible sources: 1) data directly substituted from another census or survey (for the same reporting unit, question wording, and time period); 2) data from administrative records; or 3) data obtained from some other equivalent source that has been validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (e.g., company annual reports, Securities and Exchange Commission (SEC) filings, and trade association statistics).

An **estimate** is a numerical quantity for some characteristic or attribute calculated from sample data as an approximation of the true value of the characteristic in the entire population. An estimate can also be developed from models or algorithms that combine data from various sources, including administrative records.

Estimation is the process of using data from a survey 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.

Exploratory studies (also called **Feasibility studies**) are common methods for specifying and evaluating survey content relative to concepts. In economic surveys, these studies often take the form of company or site visits.

External users – see **Users**.

-F-

Fax imaging is properly called Paperless Fax Imaging Retrieval System (PFIRS). This collection method mails or faxes a paper instrument to respondents. The respondents fax it back to the Census Bureau, where it is automatically turned into an image file.

Feasibility studies (also called **Exploratory studies**) are common methods for specifying and evaluating survey content relative to concepts. In economic surveys, these studies often take the form of company or site visits.

Field follow-up is a data collection procedure involving personal visits by enumerators to housing units to perform the operations such as, resolving inconsistent and/or missing data items on returned questionnaires, conducting a vacant/delete check, obtaining information for blank or missing questionnaires, and visiting housing units for which no questionnaire was checked in.

A **field test** is a test of some of the procedures on a small scale that mirrors the planned full-scale implementation. See **Dress rehearsal**.

A **focus group** is a pretesting technique whereby respondents are interviewed in a group setting to guide the design of a questionnaire based on the respondent's reaction to the subject matter and the issues raised during the discussion.

A **frame** consists of one or more lists of the units comprising the universe from which respondents can be selected (e.g., Census Bureau employee telephone directory). The frame may include elements not in the universe (e.g., retired employees). It may also miss elements that are in the universe (e.g., new employees).

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

-G-

Geocoding is the conversion of spatial information into computer-readable form. As such, geocoding, both the process and the concepts involved, determines the type, scale, accuracy, and precision of digital maps.

A **geographic entity** is a spatial unit of any type, legal or statistical, such as a state, county, place, county subdivision, census tract, or census block.

A **geographic entity code (geocode)** is a code used to identify a specific geographic entity. For example, the geocodes needed to identify a census block for Census 2000 data are the state code, county code, census tract number, and block number. Every geographic entity recognized by the Census Bureau is assigned one or more geographic codes. "To geocode" means to assign an address, living quarters, establishment, etc., to one or more geographic codes that identify the geographic entity or entities in which it is located.

A **generalized variance function** is a mathematical model that describes the relationship between a statistic (such as a population total) and its corresponding variance. Generalized variance function models are used to approximate standard errors of a wide variety of characteristics of the target population.

Goodness-of-fit means how well a statistical model fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values

expected under a model. Such measures can be used in statistical hypothesis testing (e.g., to test for normality of residuals, to test whether two samples are drawn from identical distributions, or to test whether outcome frequencies follow a specified distribution).

A **graphical user interface (GUI)** emphasizes the use of pictures for output and a pointing device such as a mouse for input and control whereas a command line interface requires the user to type textual commands and input at a keyboard and produces a single stream of text as output.

-H-

Random variables are **heteroscedastic** if they have different variances. The complementary concept is called homoscedasticity.

Random variables are **homoscedastic** if they have the same variance. This is also known as homogeneity of variance. The complement is called heteroscedasticity.

A **housing unit** is a house, an apartment, a mobile home or trailer, a group of rooms or a single room occupied as separate living quarters or, if vacant, intended for occupancy as separate living quarters. The Census Bureau's estimates program prepares estimates of housing units for places, counties, states, and the nation.

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-

An **implied comparison** between two (or more) estimates is one that readers might infer, either because of proximity of the two estimates in the text of the report or because the discussion presents the estimates in a manner that makes it likely readers will compare them. For an implied comparison to exist between two estimates:

- The estimates must be for similar subgroups that it makes sense to compare (e.g., two age subgroups, two race subgroups).
- The estimates must be of the same type (e.g., percentages, rates, levels).
- The subgroups must differ by only one characteristic (e.g., teenage males versus teenage females; adult males versus adult females; teenage males versus adult males). If they differ by more than one characteristic an implied comparison does not exist (e.g., teenage males versus adult females).
- The estimates appear close enough to each other in the report that the reader would make a connection between them. Two estimates in the same paragraph that satisfy the first three criteria will always constitute an implied comparison. However, if the two estimates were in different sections of a report they would not constitute an implied comparison.

Estimates presented in tables do not constitute implied comparisons. However, if a table displays the difference between two estimates, it is a direct comparison.

Imputation is a procedure for entering a value for a specific data item where the response is missing or unusable.

Information products may be in print or electronic format and include news releases; Census Bureau publications; working papers (including technical papers or reports); professional papers (including journal articles, book chapters, conference papers, poster sessions, and written discussant comments); abstracts; research reports used to guide decisions about Census Bureau programs; presentations at public events (e.g., seminars or conferences); handouts for presentations; tabulations and custom tabulations; public-use data files; statistical graphs, figures, and maps; and the documentation disseminated with these information products.

Information quality is an encompassing term comprising utility, objectivity, and integrity.

Integration testing is the phase of software testing in which individual software modules are combined and tested as a group. The purpose of integration testing is to verify functional, performance and reliability requirements placed on major design items. Integration testing can expose problems with the interfaces among program components before trouble occurs in real-world program execution.

Integrity refers to the security of information – protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.

Internal users – see **Users**.

Interviewer debriefing has traditionally been the primary method used to evaluate field or pilot tests of interviewer-administered surveys. Interviewer debriefing consists of group discussions or structured questionnaires with the interviewers who conducted the test to obtain their views of questionnaire problems.

An **item allocation rate** is the proportion of the estimated (weighted) total (T) of item t that was imputed using statistical procedures, such as within-household or nearest neighbor matrices populated by donors, for that item.

Item nonresponse occurs when a respondent provides some, but not all, of the requested information, or if the reported information is not useable.

-J-

Joint partners refers to projects where both the Census Bureau and another agency are collecting the data together, but for their own use. It is a collaborative effort to reduce overall costs to the government and increase efficiency.

-K-

Key from image (KFI) is an operation in which keyers enter questionnaire responses by referring to a scanned image of a questionnaire for which entries could not be recognized by optical character or optical mark recognition with sufficient confidence.

Key from paper (KFP) is an operation in which keyers enter information directly from a hard-copy questionnaire that could not be read by optical character or optical mark recognition with sufficient confidence.

Key variables are main classification variables (e.g., geography, demographic attributes, economic attributes, industry etc.) of units to be studied.

-L-

Latent class analysis is a method for estimating one or more components of the mean squared error or an estimator.

Linear regression is a method that models a parametric relationship between a dependent variable Y, explanatory variables Xi, i = 1, ..., p, and a random term ε. This method is called "linear" because the relation of the response (the dependent variable Y) to the independent variables is assumed to be a linear function of the parameters.

Linking – see **Record linkage**.

Load testing is the process of putting demand on a system or device and measuring its response. Load testing generally refers to the practice of modeling the expected usage of a software program by simulating multiple users accessing the program concurrently.

Logistic regression is a model used for prediction of the probability of occurrence of an event. It models the logit of the probability as a linear function of the parameters using explanatory variables X_i, i = 1, ..., p.

A **longitudinal survey** is a correlational research study that involves repeated observations of the same items over long periods of time, often many decades.

Longitudinal studies are often used in psychology to study developmental trends across the life span. The reason for this is that unlike cross-sectional studies, longitudinal studies track the same unit of observation, and therefore the differences observed in those people are less likely to be the result of cultural differences across generations.

-M-

Mail-out/mail-back is a method of data collection in which the U.S. Postal Service delivers addressed questionnaires to housing units. Residents are asked to complete and mail the questionnaires to a specified data capture center.

The **margin of error (MOE)** is a measure of the precision of an estimate at a given level of confidence (e.g., 90%). The larger the margin of error, the less confidence one should have that the reported results are close to the "true" figures; that is, the figures for the whole population.

Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER) is a topologically integrated geographic database in which the topological structures define the location, connection, and relative relationship of streets, rivers, railroads, and other features to each other, and to the numerous geographic entities for which the Census Bureau tabulates data for its censuses and sample surveys.

Matching – see **Record linkage**.

Measurement error is the difference between the true value of the measurement and the value obtained during the measurement process.

Metadata are data about data. Metadata are used to facilitate the understanding, use and management of data. An item of metadata may describe an individual datum or content item, or a collection of data including multiple content items.

Methodological expert reviews are independent evaluations of an information product conducted by one or more technical experts. These experts may be within the Census Bureau or outside the Census Bureau, such as advisory committees. See also **Peer reviews**.

A **microdata** file includes the detailed information about people or establishments. Microdata come from interviews and administrative records.

A **model** is a formal (e.g., mathematical) description of a natural system. The formal system is governed by rules of inference; the natural system consists of some collection of observable and latent variables. It is presumed that the rules of inference governing the formal system mimic in some important respect the causal relations that govern the natural system (e.g., the formal laws of arithmetic apply to counting persons).

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

Monte Carlo simulation is a technique that converts uncertainties in input variables of a model into probability distributions. By combining the distributions and randomly selecting values from them, it recalculates the simulated model many times and brings out the probability of the output.

In **multi-stage sampling**, a sample of clusters is selected and then a subsample of units is selected within each sample cluster. If the subsample of units is the last stage of sample selection, it is called a two-stage design. If the subsample is also a cluster from which units are again selected, it is called a three-stage design, etc.

Multicollinearity is a statistical term for the existence of a high degree of linear correlation amongst two or more explanatory variables in a multiple regression model. In the presence of multicollinearity, it is difficult to assess the effect of the independent variables on the dependent variable.

Multivariate analysis is a generic term for many methods of analysis that are used to investigate relationships among two or more variables.

-N-

Noise infusion is a method of disclosure avoidance in which values for each establishment are perturbed prior to table creation by applying a random noise multiplier to the magnitude data (e.g., characteristics such as first-quarter payroll, annual payroll, and number of employees) for each company.

Nonresponse means the failure to obtain information from a sample unit for any reason (e.g., no one home or refusal). There are two types of nonresponse – see **Unit nonresponse** and **Item nonresponse**.

Nonresponse bias is the deviation of the expected value of an estimate from the population parameter due to differences between respondents and nonrespondents. The impact of nonresponse on a given estimate is affected by both the degree of nonresponse and the degree that the respondents' reported values differ from what the nonrespondents would have reported.

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.

Nonresponse follow-up is an operation whose objective is to obtain completed questionnaires from housing units for which the Census Bureau did not have a completed questionnaire in mail areas (mailout/mailback, update/leave, and urban update/leave).

Nonresponse subsampling is a method for reducing nonresponse bias in which new attempts are made to obtain responses from a subsample of sampling units that did not provide responses to the first attempt.

Nonsampling errors are survey errors caused by factors other than sampling (e.g., nonsampling errors include errors in coverage, response errors, non-response errors, faulty questionnaires, interviewer recording errors, and processing errors).

The **North American Industry Classification System (NAICS)** is the standard used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. Canada, Mexico, and the U.S. jointly developed the NAICS to provide new comparability in statistics about business activity across North America. NAICS coding has replaced the U.S. Standard Industrial Classification (SIC) system (for more information, see www.census.gov/epcd/www/naics.html).

-O-

Objectivity focuses on whether information is accurate, reliable, and unbiased, and is presented in an accurate, clear, complete, and unbiased manner.

Optical character recognition (OCR) is a technology that uses an optical scanner and computer software to "read" human handwriting and convert it into electronic form.

Optical mark recognition (OMR) is a technology that uses an optical scanner and computer software to recognize the presence of marks in predesignated areas and assign a value to the mark depending on its specific location and intensity on a page.

Outliers in a set of data are values that are so far removed from other values in the distribution that their presence cannot be attributed to the random combination of chance causes.

-P-

The ***p*-value** is the probability that the observed value of the test statistic or a value that is more extreme in the direction of the alternative hypothesis, calculated when H_0 is true, is obtained.

Parameters are unknown, quantitative measures (e.g., total revenue, mean revenue, total yield or number of unemployed people) for the entire population or for specified domains that are of interest. A parameter is a constant in the equation of a curve that can be varied to yield a family of similar curves or a quantity (such as the mean, regression coefficient, or variance) that characterizes a statistical population and that can be estimated by calculations from sample data.

Participation means that the employee takes an active role in the event.

A peer review is an independent evaluation of an information product conducted by one or more technical experts.

Personally identifiable information refers to any information about an individual maintained by the Census Bureau which can be used to distinguish or trace an individual's identity, such as their name, social security number, date and place of birth, biometric records, etc., including any other personal information which is linked or linkable to an individual. Also see **Business identifiable information**.

Census Bureau information products must not contain **policy views**. The Census Bureau's status as a statistical agency requires us to absolutely refrain from taking partisan political positions. Furthermore, there is an important distinction between producing data and using that data to advocate for program and policy changes. The Census Bureau's duty is to produce high quality, relevant data that the nation's policy makers can use to formulate public policy and programs. The Census Bureau should not, however, insert itself into a debate about the program or policy implications of the statistics it produces. We produce poverty statistics; we do not advocate for programs to alleviate poverty.

Population estimates (post-censal or intercensal estimates) are prepared for demographic groups and geographic areas. These estimates usually are developed from separate measures of the

components of population change (births, deaths, domestic net migration, and net international migration) in each year but may be supplemented with other methodologies in the absence of current measures of components.

Post-stratification is applied to survey data by stratifying sample units after data collection using information collected in the survey and auxiliary information to adjust weights to population control totals or for nonresponse adjustment.

Precision of survey results refers to how closely the results from a sample can be obtained across repeated samples conducted using the same techniques from the same population at the same time. A precise estimate is stable over replications.

Pretesting is a broad term that incorporates many different techniques for identifying problems for both respondents and interviewers with regard to question content, order/context effects, skip instructions, and formatting**.**

Primary sampling units (PSU) are clusters of reporting units selected in the first stage of a multi-stage sample.

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 frame. The advantage of probabilistic sampling methods is that sampling error can be calculated without reference to a model assumption. Such methods include random sampling, systematic sampling, and stratified sampling.

The **probability of selection** is the probability that a population (frame) unit will be drawn in a sample. In a simple random selection, this probability is the number of elements drawn in the sample divided by the number of elements on the sampling frame.

Probability sampling is an approach to sample selection that satisfies certain conditions:
1. We can define the set of samples that are possible to obtain with the sampling procedure.
2. A known probability of selection is associated with each possible sample.
3. The procedure gives every element in the population a nonzero probability of selection.
4. We select one sample by a random mechanism under which each possible sample receives exactly its probability of selection.

A **project** is a temporary endeavor undertaken to create a unique product, service, or result.

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

Protected information (as defined in Data Stewardship Policy DS007, *Information Security Management Program*) includes information about individuals, businesses, and sensitive statistical methods that are protected by law or regulation. The Census Bureau classifies the following as protected information:
- Individual census or survey responses.

- Microdata or paradata, containing original census or survey respondent data and/or administrative records data that do not meet the disclosure avoidance requirements.
- Address lists and frames, including the Master Address File (MAF).
- Pre-release Principal Economic Indicators and Demographic Time-Sensitive Data.
- Aggregate statistical information produced for internal use or research that do not meet the Disclosure Review Board disclosure avoidance requirements, or that have not been reviewed and approved for release.
- Internal use methodological documentation in support of statistical products such as the primary selection algorithm, swapping rates, or Disclosure Review Board checklists.
- All personally identifiable information (PII) protected by an existing legal authority (such as Title 13, Title 15, Title 5, and Title 26).
- All business identifiable information (BII) protected by an existing legal authority.

A **public event** means that the event is open to the general public, including events that require a registration fee.

-Q-

A **qualified user** is a user with the experience and technical skills to meaningfully understand and analyze the data and results. For example, a qualified user of direct estimates produced from samples understands sampling, estimation, variance estimation, and hypothesis testing.

A **quantity response rate** is the proportion of the estimated (weighted) total (T) of data item t reported by tabulation units in the sample (expressed as a percentage). [Note: Because the value of economic data items can be negative (e.g., income), the absolute value must be used in the numerators and denominators in all calculations.]

A **questionnaire** is a set of questions designed to collect information from a respondent. A questionnaire may be interviewer-administered or respondent-completed, using paper-and-pencil methods for data collection or computer-assisted modes of completion.

-R-

Raking is a method of adjusting sample estimates to known marginal totals from an independent source. For a two-dimensional case, the procedure uses the sample weights to proportionally adjust the weights so that the sample estimates agree with one set of marginal totals. Next, these adjusted weights are proportionally adjusted so that the sample estimates agree with the second set of marginal totals. This two-step adjustment process is repeated enough times until the sample estimates converge simultaneously to both sets of marginal totals.

In **random rounding,** cell values are rounded, but instead of using standard rounding conventions a random decision is made as to whether they will be rounded up or down.

Ratio estimation is a method of estimating from sample data. In ratio estimation**,** an auxiliary variate x_i, correlated with y_i is obtained for each unit in the sample. The population total X of the x_i must be known. The goal is to obtain increased precision by taking advantage of the

correlation between y_i and x_i. The ratio estimate of Y, the population total of y_i, is $\hat{Y}_R = X\left(\frac{y}{x}\right)$, where y and x are the sample totals of y_i and x_i respectively.

Readily accessible means that users can access the documentation when they need it, not that it is only available on request.

Recoding is a disclosure limitation technique that involves collapsing/regrouping detail categories of a variable so that the resulting categories are safe.

Record linkage is the process of linking or matching two or more records that are determined to refer to the same person or establishment.

Regression is a statistical method which tries to predict the value of a characteristic by studying its relationship with one or more other characteristics.

A **regression model** is a statistical model used to depict the relationship of a dependent variable to one or more independent variables**.**

Reimbursable projects are those for which the Census Bureau receives payment (in part or in total) from a customer for products or services rendered.

Reinterview is repeated measurement of the same unit intended to estimate measurement error (response error reinterview) or designed to detect and deter falsification (quality control reinterview).

A **release phase** refers to the point in the statistical process where you release the data. It may be to the public, the sponsor, or any other user for whom the data was created.

Releases of information products are the delivery or the dissemination of information products to government agencies, organizations, sponsors, or individuals outside the Census Bureau, including releases to the public.

Replication methods are variance estimation methods that take repeated subsamples, or replicates, from the data, re-compute the weighted estimate for each replicate, and then compute the variance based on the deviations of these replicate estimates from the full-sample estimate. The subsamples are generated to properly reflect the variability due to the sample design.

Reproducibility means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased). If the Census Bureau applies the reproducibility test to specific types of original or supporting data, the associated guidelines shall provide relevant definitions of reproducibility (e.g., standards for replication of laboratory data). With respect to analytic results, "capable of being substantially reproduced" means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error.

A **residual** is the observed value minus the predicted value.

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

Respondent debriefing is a pretesting technique that involves using a structured questionnaire following data collection to elicit information about respondents' interpretations of survey questions.

A **response analysis survey** is a technique for evaluating questionnaires from the perspective of the respondent. It is typically a respondent debriefing conducted after a respondent has completed the main survey.

Response error is the difference between the true answer to a question and the respondent's answer. It may be caused by the respondent, the interviewer, the questionnaire, the survey procedure or the interaction between the respondent and the interviewer.

A **response rate** measures the proportion of the selected sample that is represented by the responding units.

Revisions history is a stability diagnostic to compare regARIMA modeling and seasonal adjustment results over lengthening time spans. History analysis begins with a shortened series. Series values are added, one at a time, and the regARIMA model and seasonal adjustment are reestimated. Comparing different sets of adjustment options for the same series may indicate that one set of options is more stable. Among adjustment options whose other diagnostics indicate acceptable quality, options that result in fewer large revisions, that is, fewer large changes as data are added, usually are preferred.

-S-

The **sample design** describes the target population, frame, sample size, and the sample selection methods.

The **sample size** is the number of population units or elements selected for the sample, determined in relation to the required precision and available budget for observing the selected units.

A **sample survey** is a data collection that obtains data from a sample of the population.

The **sampled population** is the collection of all possible observation units (objects on which measurements are taken) that might have been chosen in the sample. For example, in a presidential poll taken to determine who people will vote for, the target population might be all persons who are registered to vote. The sampled population might be all registered voters who can be reached by telephone.

Sampling is the process of selecting a segment of a population to observe and facilitate the estimation and analysis of something of interest about the population. The set of sampling units selected is referred to as the sample. If all the units are selected, the sample is referred to as a census.

Sampling error is the uncertainty associated with an estimate that is based on data gathered from a sample of the population rather than the full population.

A **sampling frame** is any list or device that, for purposes of sampling, de-limits, identifies, and allows access to the sampling units, which contain elements of the frame population. The frame may be a listing of persons, housing units, businesses, records, land segments, etc. One sampling frame or a combination of frames may be used to cover the entire frame population.

Sampling units are the basic components of a sampling frame. The sampling unit may contain, for example, defined areas, houses, people, or businesses.

Sampling weight is a weight assigned to a given sampling unit that equals the inverse of the unit's probability of being included in the sample and is determined by the sample design. This weight may include a factor due to subsampling.

Sanitized data, used for testing, may be totally fictitious or based on real data that have been altered to eliminate the ability to identify the information of any entity represented by the data.

Scheffé's method is a method for adjusting significance levels in a linear regression analysis to account for multiple comparisons. It is particularly useful in analysis of variance, and in constructing simultaneous confidence bands for regressions involving basis functions. Scheffé's method is a single-step multiple comparison procedure which applies to the set of estimates of all possible contrasts among the factor level means, not just the pairwise differences considered by the Tukey method.

A **scoring weight** is the amount of value assigned when a pair of records agree or disagree on the same matching variable. Each matching variable is assigned two scoring weights --- a positive weight for agreement and a negative weight for disagreement. After comparing all matching variables on a matching variable by matching variable basis, the resulting set of assigned weights are added to get a total score for the total record. Pairs of records with scores above a predetermined cut-off are classified as a match; pairs of records with scores below a second predetermined cut-off are classified as a non-match.

Seasonal adjustment is a statistical technique that consists of estimating seasonal factors and applying them to a time series to remove the seasonal variations in the estimates.

Sensitivity analysis is designed to determine how the variation in the output of a model (numerical or otherwise) can be apportioned, qualitatively or quantitatively, to changes in input parameter values and assumptions. This type of analysis is useful in ascertaining the capability of a given model, as well its robustness and reliability.

Sequential sampling is a sampling method in which samples are taken one at a time or in successive predetermined groups, until the cumulative result of their measurements (as assessed against predetermined limits) permits a decision to accept or reject the population or to continue sampling. The number of observations required is not determined in advance, but the decision to terminate the operation depends, at each stage, on the results of the previous observations. The plan may have a practical, automatic termination after a certain number of units have been examined.

Significance level refers to the probability of rejecting a true null hypothesis.

Simple random sampling (SRS) is a basic probability selection scheme that uses equal probability sampling with no strata.

A **skip pattern** in a data collection instrument is the process of skipping over non-applicable questions depending upon the answer to a prior question.

Sliding spans diagnostics are seasonal adjustment stability diagnostics for detecting adjustments that are too unstable. X-12-ARIMA creates up to four overlapping subspans of the time series, seasonally adjusts each span, then compares the adjustments of months (quarters with quarterly data) common to two or more spans. Months are flagged whose adjustments differ by more than a certain cutoff. (The default cutoff is 3% for most comparisons.) If too many months are flagged, the seasonal adjustment is rejected for being too unstable. The series should not be adjusted unless other software options are found that lead to an adjustment with an acceptable number of flagged months. Sliding spans diagnostics can include comparisons of seasonally adjusted values, seasonal factors, trading day factors, month-to-month changes and year-to-year changes. (Year-to-year change results are not used to accept or reject an adjustment.)

Small area estimation is a statistical technique involving the estimation of parameters for small sub-populations where a sample has insufficient or no sample for the sub-populations to be able to make accurate estimates for them. The term "small area" may refer strictly to a small geographical area such as a county, but may also refer to a "small domain," i.e., a particular demographic within an area. Small area estimation methods use models and additional data sources (such as census data) that exist for these small areas in order to improve estimates for them.

Special sworn status (SSS) is conferred upon individuals for whom the Census Bureau approves access to confidential Census Bureau data in furtherance of a Title 13 purpose. SSS individuals are subject to same legal penalties for violation of confidentiality as employees.

Spectral graphs are diagnostic graphs that indicate the presence of seasonal or trading day effects. Visually significant peaks at the marked seasonal and/or trading day frequencies usually indicate the presence of these effects, in some cases as residual effects after an adjustment that is not fully successful for the span of data from which the spectrum is calculated. Spectral graphs are available for the prior-adjusted series (or original series if specified), regARIMA model residuals, seasonally adjusted series, and modified irregular.

Split panel tests refer to controlled experimental testing of questionnaire variants or data collection modes to determine which one is "better" or to measure differences between them.

Stakeholders include Congress, federal agencies, sponsors, state and local government officials, advisory committees, trade associations, or organizations that fund data programs, use the data, or are affected by the results of the data programs.

The **standard deviation** is the square root of the variance and measures the spread or dispersion around the mean of a data set.

The **standard error** is a measure of the variability of an estimate due to sampling.

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

Statistical attribute matching consists of comparing two records, determining if they refer to "similar" entities (but not necessarily the same entity), and augmenting data from one record to the other.

Statistical inference is inference about a population from a random or representative sample drawn from it. It includes point estimation, interval estimation, and statistical significance testing.

A **statistical model** consists of a series of assumptions about a data generating process that explicitly involve probability distributions and functions on those distributions, in order to construct an estimate or a projection of one or more phenomena.

Statistical purposes refer to the description, estimation, or analysis of the characteristics of groups without identifying the individuals or organizations that compose such groups.

Statistical significance is attained when a statistical procedure applied to a set of observations yields a *p*-value that exceeds the level of probability at which it is agreed that the null hypothesis will be rejected.

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

Stratification involves dividing the sampling frames into subsets (called strata) prior to the selection of a sample for statistical efficiency, for production of estimates by stratum, or for operational convenience. Stratification is done such that each stratum contains units that are relatively homogeneous with respect to variables that are believed to be highly correlated with the information requested in the survey.

Stratified sampling is a sampling procedure in which the population is divided into homogeneous subgroups or strata and the selection of samples is done independently in each stratum.

Sufficient data is determined for a survey by whether the respondent completes enough items for the case to be considered a completed response.

Supplemental reinterview allows the regional offices to select any field representative (FR) with an original interview assignment for reinterview. All assigned cases that are not selected for reinterview are available as inactive supplemental reinterview cases. The regional office may place a field representative in supplemental reinterview for various reasons: the FR was not selected for reinterview; the FR was hired during the assignment period; or the regional office needs to reinterview additional cases to investigate the FR for suspected falsification.

Swapping is a disclosure limitation technique that involves selecting a sample of records, finding a match in the database on a set of predetermined variables, and swapping all other variables.

Synthetic data are microdata records created to improve data utility while preventing disclosure of confidential respondent information. Synthetic data is created by statistically modeling original data and then using those models to generate new data values that reproduce the original data's statistical properties. Users are unable to identify the information of the entities that provided the original data.

Systematic sampling is a method of sample selection in which the sampling frame is listed in some order and every k^{th} element is selected for the sample, beginning from a random start between 1 and k.

A **systems test** is used to test the data collection instrument along with the data management systems.

-T-

The **target population** is the complete collection of observations under study. For example, in a presidential poll taken to determine who people will vote for, the target population might be all persons who are registered to vote. The sampled population might be all registered voters who can be reached by telephone.

A **Taylor series** is a representation of a function as an infinite sum of polynomial terms calculated from the values of its derivatives at a single point.

The **Taylor series method for variance estimation** is used to estimate variances for non-linear estimators such as ratio estimators. If the sample size is large enough so that estimator can be closely approximated by the first order (linear) terms in the Taylor series, then the variances can be approximated by using variance methods appropriate for linear statistics. The Taylor series approximation to the ratio estimator is: $\hat{Y}_R \approx Y + (y - Y) - (Y/X)(x - X)$. This approximation is linear in the survey sample totals x and y.

Testing is a process used to ensure that methods, systems or other components function as intended.

A **time series** is a sequence of data values obtained over a period of time, usually at uniform intervals.

Timeliness of information reflects the length of time between the information's availability and the event or phenomenon it describes.

Top-coding is a disclosure limitation technique that involves limiting the maximum value of a variable allowed on the file to prevent disclosure of individuals or other units with extreme values in a distribution.

Topologically Integrated Geographic Encoding and Referencing (TIGER) – see definition for Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER).

A **total quantity response rate** is the proportion of the estimated (weighted) total (T) of data item t reported by tabulation units in the sample or from sources determined to be equivalent-quality-to-reported data (expressed as a percentage).

Touch-tone data entry (TDE) is a data collection method that uses an electronic instrument to collect and capture data by telephone.

Transparency refers to providing documentation about the assumptions, methods, and limitations of an information product to allow qualified third parties to reproduce the information, unless prevented by confidentiality or other legal constraints.

Truth decks are used to test imputation methods by comparing the imputed values to the original values for the items flagged as missing. The truth deck originates as a file of true responses. Certain responses are then blanked in a manner that reflects the probable nonresponse in the sample. The truth deck is then run through the imputation process in order to evaluate the accuracy of the imputed values.

Tukey's method is a single-step multiple comparison procedure and statistical test generally used in conjunction with an ANOVA to find which means are significantly different from one another. Named after John Tukey, it compares all possible pairs of means, and is based on a studentized range distribution q (this distribution is similar to the distribution of t from the t-test).

-U-

Unduplication involves the process of deleting units that are erroneously in the frame more than once to correct for overcoverage.

Unit nonresponse occurs when a sampled unit fails to respond or a sampled unit response does not meet a minimum threshold and is classified as not having responded at all.

Usability testing in surveys is the process whereby a group of representative users are asked to interact and perform tasks with survey materials (e.g., computer-assisted forms) to determine if the intended users can carry out planned tasks efficiently, effectively, and satisfactorily.

A **user interface** is the aspects of a computer system or program that can be seen (or heard or otherwise perceived) by the human user, and the commands and mechanisms the user uses to control its operation and input data.

Users are organizations, agencies, the public, or any others expected to use the information products. Census Bureau employees, contractors, and other Special Sworn Status individuals affiliated with the Census Bureau are **internal users**. Users outside of the Census Bureau, including Congress, federal agencies, sponsors, other Special Sworn Status individuals, and the public, are **external users**.

Utility refers to the usefulness of the information for its intended users.

-V-

Variance is a measurement of the error associated with nonobservation, that is, the error that occurs because all members of the frame population are not measured. The measurement is the average of the squared differences between data points and the mean.

Version Control is the establishment and maintenance of baselines and the identification of changes to baselines that make it possible to return to the previous baseline. A baseline, in the context of documentation, is a document that has been formally reviewed and agreed on.

-W-

Weights are values associated with each sample unit that are intended to account for probabilities of selection for each unit and other errors such as nonresponse and frame undercoverage so that estimates using the weights represent the entire population. A weight can be viewed as an estimate of the number of units in the population that the sampled unit represents.

Working papers are information products that are prepared by Census Bureau employees (or contractors), but the Census Bureau does not necessarily affirm their content. They include technical papers or reports, division reports, research reports, and similar documents that discuss analyses of subject matter topics or methodological, statistical, technical or operational issues. The Census Bureau releases working papers to the public, generally on the Census Bureau's Web site. Working papers must include a disclaimer, unless the Associate Director responsible for the program determines that a disclaimer is not appropriate.

Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census

by

J. David Brown
U.S. Census Bureau

Misty L. Heggeness
U.S. Census Bureau

Suzanne M. Dorinski
U.S. Census Bureau

Lawrence Warren
U.S. Census Bureau

Moises Yi
U.S. Census Bureau

CES 18-38 **August, 2018**

Abstract

This paper examines the quality of citizenship data in self-reported survey responses compared to administrative records and evaluates options for constructing an accurate count of resident U.S. citizens. Person-level discrepancies between survey-collected citizenship data and administrative records are more pervasive than previously reported in studies comparing survey and administrative data aggregates. Our results imply that survey-sourced citizenship data produce significantly lower estimates of the noncitizen share of the population than would be produced from currently available administrative records; both the survey-sourced and administrative data have shortcomings that could contribute to this difference. Our evidence is consistent with noncitizen respondents misreporting their own citizenship status and failing to report that of other household members. At the same time, currently available administrative records may miss some naturalizations and capture others with a delay. The evidence in this paper also suggests that adding a citizenship question to the 2020 Census would lead to lower self-response rates in households potentially containing noncitizens, resulting in higher fieldwork costs and a lower-quality population count.

[*] We thank career staff and statistical experts within the Bureau who graciously gave their time and effort to review, comment, edit, and make improvements to this document. The analysis, thoughts, opinions, and any errors presented here are solely those of the authors and do not necessarily reflect any official position of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The Disclosure Review Board release numbers are DRB-B0093-CDAR-20180621, DRB-B0103-CDAR-20180712, and DRB-B0113-CDAR-20180806. Republication in whole or part must be cleared with the authors.

J. David Brown is a Senior Economist in the Center for Economic Studies at the U.S. Census Bureau and the corresponding author on this paper, 4600 Silver Hill Road, Washington, DC 20233, j.david.brown@census.gov. Misty L. Heggeness is Senior Advisor for Evaluations and Experiments in the Research and Methodology Directorate at the U.S. Census Bureau. Suzanne M. Dorinski is a Mathematical Statistician currently on detail with the Social, Economic, and Housing Statistics Division at the U.S. Census Bureau. Lawrence Warren is an Economist in the Center for Economic Studies at the U.S. Census Bureau. Moises Yi is an Economist in the Center for Economic Studies at the U.S. Census Bureau.

1. Introduction

National statistical agencies are charged with collecting and reporting accurate information about society, including individuals, households, and businesses. This information is used to produce official statistics about the demographic composition of persons living in the nation – including information about migration, citizenship, and mobility. For decades, the United States has relied on household survey questionnaires to collect data on migration and immigration status (Census Bureau 2002). Generally, the focus is on whether an individual has lived in that current location for more than one (or five) years, a date for their last move, citizenship status, and year of naturalization. To date, the collection of this information via survey vehicles has been sufficient for general statistical reporting on immigrants living in the U.S.; however, very few studies have examined the extent to which individuals answer these sensitive questions accurately, how inclusion of these questions affects overall response rates, or how item nonresponse on these questions compares to other questions.

In this paper, we study the quality of self-reported citizenship questions by comparing responses in the American Community Survey (ACS), the Census, the Survey of Income and Program Participation (SIPP), and administrative records on citizenship from the Social Security Administration. There are now multiple survey and administrative sources of data to study immigration and citizenship status. We examine the strengths and weaknesses of these sources for the development of future statistics on citizenship status. We focus on both the accuracy and completeness in all options. The alternatives we consider for constructing a count of resident citizens are the following: (A) no change in current data collection, combined with small area estimation using the ACS and administrative citizenship data sources, (B) add a citizenship question to the 2020 Census, (C) obtain citizenship status from administrative records for the entire 2020 Census population, and (D) combine alternatives (B) and (C). Factors to consider when evaluating these alternatives include the quality of the data sources, comprehensiveness and biases in data coverage, cost, and the effects on the quality of the 2020 full population count. We analyze each of these aspects.

We find that discrepancies between survey-collected citizenship data and administrative records are more extensive than discrepancy estimates from previous research. The degree to which persons who are noncitizens in administrative records self-report being citizens in surveys is greater for non-Hispanics than Hispanics. Most of the people with these discrepancies report being citizens from birth or naturalized long ago, regardless of ethnicity. The discrepancy patterns imply that the ACS estimate of the noncitizen share of the population is lower than comparable estimates based on currently available administrative records.

The remainder of the paper is structured as follows. Section 2 provides general background and history of the current issue. Section 3 documents the coverage of survey and administrative record citizenship data. The quality of the data from survey and administrative record sources is analyzed in Section 4. Section 5 contains regression analyses of item response and data quality. Section 6 estimates the effects of inclusion of a citizenship question on survey response rates. Estimates of the citizenship question's effects on the cost and quality of the 2020 Census in general are provided in Section 7. Forecasts of the number of people for whom citizenship is sourced by the 2020 Census

citizenship question, administrative records, and model imputation when using each of the alternatives are given in Section 8. Section 9 concludes.

2. Background

2.1 History of Citizenship Data Collection through Household Surveys and Censuses

The Census Bureau has collected and preserved citizenship data since 1820 via historical full count censuses, household surveys, and administrative records (AR), but the practice of asking citizenship and migration-related questions on censuses has varied over time. The 1820 and 1830 Censuses asked for a tally of the total number of non-naturalized foreigners in the household. The 1870 Census asked citizenship status of all male persons aged 21 and older (Census Bureau 2002). The federal government did not ask citizenship status during the 1880 Census, but reintroduced it in the 1890 Census, and the question stayed on full-count Census questionnaires through 1950. The 1950 Census was the last full-count Census to ask the citizenship status of every resident in the U.S. if he or she reported a foreign birthplace (Census Bureau 2002).

While the 1960 Census did not ask about citizenship throughout the country, it was reintroduced on the long form (which sampled approximately one-in-six households across the country) in the 1970 Census and remained on the long form until 2000 (Census Bureau 2002). The question never reappeared on the short form after 1950. After the 2000 Census, citizenship data collection moved to the American Community Survey (ACS), which replaced the Census long form. The ACS collects responses from approximately 1.6 percent of households annually (American Community Survey 2016a, American Community Survey 2016b).[2]

Since the advent of the long form and continuing with the ACS, the Census Bureau has focused Census enumeration on obtaining only the data necessary for a concise and condensed full-population count (Weinberg 2011). It also prioritizes the collection of data mandated by Public Law 94-171 (PL94), which instructs the Census Bureau to cooperate with state redistricting offices in support of their efforts to redraw legislative districts in compliance with the Constitution, Supreme Court, and the 1965 Voting Rights Act. The questionnaire asks only the core demographic, race, ethnicity, and housing questions, not including citizenship.

2.2 The Citizen Voting Age Population by Race and Ethnicity (CVAP) Table

On December 12, 2017, the Census Bureau received a request from the Department of Justice to include a citizenship question on the 2020 Census of Population and Housing (Department of Justice 2017). The request prompted the Census Bureau to conduct a study of the feasibility and best options for meeting this request. This paper summarizes the technical analysis conducted for alternative options for obtaining citizenship data for the entire population to produce the Citizen Voting Age Population by Race and Ethnicity (CVAP) table at the census block level. CVAP is

[2] We calculate this number using American Fact Finder (AFF) Tables B98001 and B25001.

currently produced at the census block-group level using estimates from the five-year American Community Survey (ACS) data.

Since 1975, the Census Bureau has provided population estimates by detailed geography to support redistricting under Public Law 94-171 (PL94). For the 2000 Census, the Citizen Voting Age Population (CVAP) estimates, tabulated at the block-group level, were produced from the long form citizenship question. Since 2011, the CVAP estimates have been tabulated annually at the block-group level from the most recent 5-year ACS data. The 2011 publication was based on the 2005-2009 ACS surveys. These data were released in the same time frame as the 2010 PL94 redistricting estimates.[3] The redistricting data must be released before April 1st of the year following a census under the authority of 13 U.S.C. Section 141.

The difficulty in integrating these two tables for redistricting and enforcement of the Voting Rights Act was cited by the Department of Justice in its December 12, 2017 letter. The Department of Justice requested block-level citizen voting-age population estimates by the U.S. Office of Management and Budget (OMB)-approved race and ethnicity categories[4] directly from the 2020 Census of Population and Housing, which would require the addition of a citizenship question directly onto the full count 2020 Census enumeration form.

2.3 Prior Research on Citizenship Data Quality

We build on past research on Census citizenship data quality. Prior studies have suggested that citizenship is inaccurately estimated in Census Bureau surveys. Passel and Clark (1997) document that the 1990 Census and 1996 Current Population Survey (CPS) estimates of the number of naturalized persons are much higher than the numbers from Immigration and Naturalization Services (INS) administrative data.[5] The study suggests that about 75 percent of those who report having lived in the U.S. fewer than five years and being naturalized citizens probably are not citizens, at least at the time of the survey. Furthermore, one-third of longer-resident Central American and Mexican origin individuals who self-reported naturalization were probably not citizens at the time of the survey. These discrepancies were attributed to incorrect reporting, possibly because respondents were confused about their status or had an incentive to misreport it to enumerators and interviewers.

Camarota and Capizzano (2004) conducted focus groups with over 50 field representatives (FRs) for the Census 2000 Supplemental Survey (a pilot for the ACS). FRs reported that foreign-born respondents living in the country illegally or from countries where there is distrust in government were less likely to participate. Some foreign-born respondents failed to list all household members. FRs suspected that some foreign-born respondents misreported citizenship status, and they

[3] For more information, see: https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.html and https://www.census.gov/rdo/data/2010_census.html .
[4] See Office of Management and Budget (1997).
[5] This comes from Van Hook and Bachmeier's (2013) summary of Passel and Clark (1997).

believed this was due to "recall bias, a fear of the implications of certain responses or a desire to answer questions in a socially desirable way."

More recently, Van Hook and Bachmeier (2013) compared 2010 ACS and Office of Immigration Statistics (OIS) naturalizations data, finding that the ACS produced higher naturalization estimates than OIS for those residing in the U.S. less than five years, as well as for longer-resident Mexican-origin persons. Several papers have studied the effects of state immigration laws on the number and locational choices of immigrants (see, for example, Amuendo-Dorantes and Lozano 2014 and 2015, Bohn et al. 2014, Ellis et al. 2014, Good 2013, and Orrenius and Zavodny 2016). They have generally found reductions in the immigrant population after the introduction of these laws.[6] Deterioration in survey data quality during periods of stronger immigration enforcement could help explain the measured reductions. We contribute to the literature on Census citizenship data quality by directly linking Census and household survey data to administrative records. We not only examine the quality of survey-collected citizenship data, but also the effect of including a citizenship question on the quality of other data via their consequences for response rates and nonresponse follow-up.

3.1 Survey Coverage

In addition to the full count Census of Population and Housing that collects a limited amount of information on the entire population once every ten years, the Census Bureau also collects information on individuals and households in both legally-mandated and sponsored (reimbursable) surveys. These surveys collect more detailed demographic, social, and economic characteristics of people living in the United States, including information on citizenship status and migration variables.

The Census Bureau currently conducts four surveys that ask citizenship questions. The American Community Survey (ACS), the Current Population Survey (CPS), the American Housing Survey (AHS), and the Survey of Income and Program Participation (SIPP) all collect data on citizenship status. The universe for citizenship questions on these surveys is all persons living in the household. The ACS, CPS, SIPP, and AHS distinguish between citizens born in the United States, those born in U.S. territories, those born abroad to U.S. citizen parents, and those of foreign nativity but naturalized. Additionally, the SIPP asks about more nuanced naturalizations, including becoming a citizen through one's own or a spouse's military service or via adoption by U.S. citizen parents. [7]

To assess the citizenship coverage of existing Census Bureau survey data, we link all of the household surveys measuring citizenship status to the 2010 Census. The person-level linkage to

[6] For more information, see https://www.troutman.com/files/FileControl/89dad504-6be0-4335-aa1a-35a433102d63/7483b893-e478-44a4-8fed-f49aa917d8cf/Presentation/File/Survey%20of%20state%20and%20federal%20laws%20requiring%20E-Verify.pdf and table 1 in Orrenius and Zavodny (2016) for the list of states with mandatory E-Verify laws.

[7] This information is from the Master Demographic Pilot Feasibility Study.

the 2010 Census is based on the Protected Identification Key or PIK (the Census Bureau's internal unique person identifier) appended to person records using the Person Identification Validation System (PVS). To implement the record linkage, we first compiled an unduplicated list of individuals surveyed by the Census Bureau in Title 13 mandated surveys (ACS and SIPP) and reimbursable surveys (CPS[8] and AHS[9]). We link this unduplicated list of individuals to the 2010 Census (see Appendix Table A1).

Household surveys linked to the 2010 Census contain self-reported citizenship status for 44.6 million people, or 14.4 percent of the 2010 Census population. Of these, 43.1 million report being citizens (see Appendix Table A2). We conclude that the population coverage from existing survey data is a relatively small share of the total population, consistent with the sampling rates of these surveys.

Figure 1 Panel A shows item nonresponse in the 2016 ACS for sex, age, and citizenship.[10] We show nonresponse rates for the full sample, as well as for select subgroups by race/ethnicity and relationship to the householder.[11,12] Sex has the lowest nonresponse rates across the entire sample, as well as within subgroups with all recording less than 1 percent nonresponse, except for nonrelatives. Nonresponse rates for age are higher, and for some subgroups it has the highest level of nonresponse among the three items shown here. This is true for non-Hispanic white, non-Hispanic black, reference person, and relative of the reference person.[13] Hispanics and non-Hispanic other race[14] have higher rates of nonresponse for citizenship than for sex or age, providing some preliminary evidence that these groups could be disproportionately impacted by the addition of citizenship on the 2020 Census questionnaire.

[8] The CPS is sponsored by the Department of Labor's Bureau of Labor Statistics.

[9] The AHS is sponsored by the Department of Housing and Urban Development.

[10] Appendix Table A3 shows item nonresponse rates for questions on the 2000 Census short form and the 2010 Census. We choose sex and age as benchmarks, since they are on the Census questionnaire. As shown in Appendix Table A3, item allocation rates (including both nonresponses and responses that are edited) are higher for many ACS questions than for sex, age, or citizenship, but they are not being considered for inclusion on the Census questionnaire and are thus less relevant.

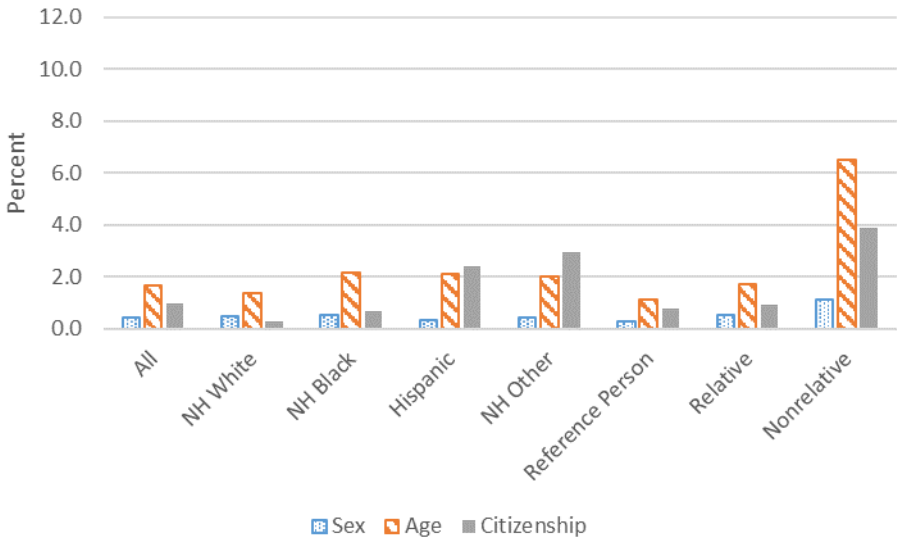
[11] Throughout the paper, we show results not only by citizenship, but also by race and ethnicity for two main reasons. The CVAP data provide counts not just by citizenship, but also race and ethnicity, so differential effects on race/ethnic groups from adding a citizenship question are relevant. In addition, our administrative record noncitizen measure has incomplete coverage (it does not cover noncitizens without SSNs), while a significant percentage of noncitizens without SSNs are Hispanic (Bond et al., 2014). Thus, to some extent the Hispanic category captures noncitizens excluded from the measured noncitizen category.

[12] The householder, also referred to as the reference person or person 1, is the first person listed on the household roster. The reference person typically is the primary or sole respondent to the survey. The relative and nonrelative categories are based on the person's relationship to the householder. The relative category includes husband or wife, biological son or daughter, adopted son or daughter, stepson or stepdaughter, brother or sister, father or mother, grandchild, parent-in-law, son-in-law or daughter-in-law, other relative, unmarried partner, and foster child. The nonrelative category includes roomer or boarder, housemate or roommate, and other nonrelative.

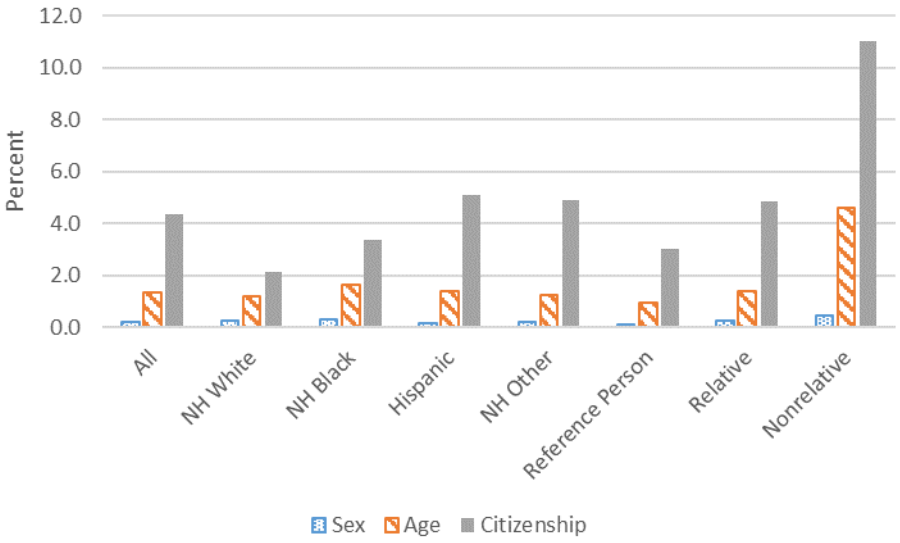
[13] We treat all persons in group quarters as reference persons. The results are qualitatively similar if group quarters are excluded.

[14] Non-Hispanic other race includes non-Hispanic Asian, non-Hispanic American Indian and Alaskan Native, non-Hispanic Native Hawaiian and Other Pacific Islander, and non-Hispanic two or more races.

Figure 1. American Community Survey (ACS) Nonresponse, 2016



Panel A. Item Nonresponse



Panel B. Item Nonresponse for Census Numident-Identified Noncitizens

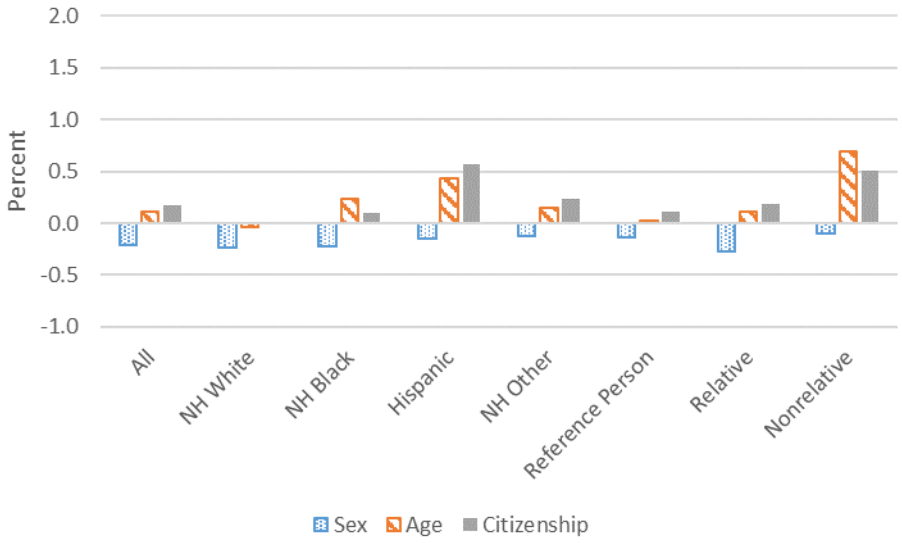
Source: American Community Survey (ACS) and Census Numident, 2016.

Given item nonresponse to the citizenship question as shown in Figure 1 Panel A, we are particularly interested in understanding the potential sensitivity of response specifically for noncitizens. Figure 1 Panel B shows the same information as Panel A, restricted to those

individuals who are identified as noncitizens in the Census Numident,[15] meaning that administrative records show their status as noncitizen. Panel B illustrates the heightened sensitivity associated with collecting citizenship data for noncitizens through surveys. Item nonresponse to the citizenship question is particularly high for nonrelative household members, where one-in-ten do not have a citizenship response in the ACS.

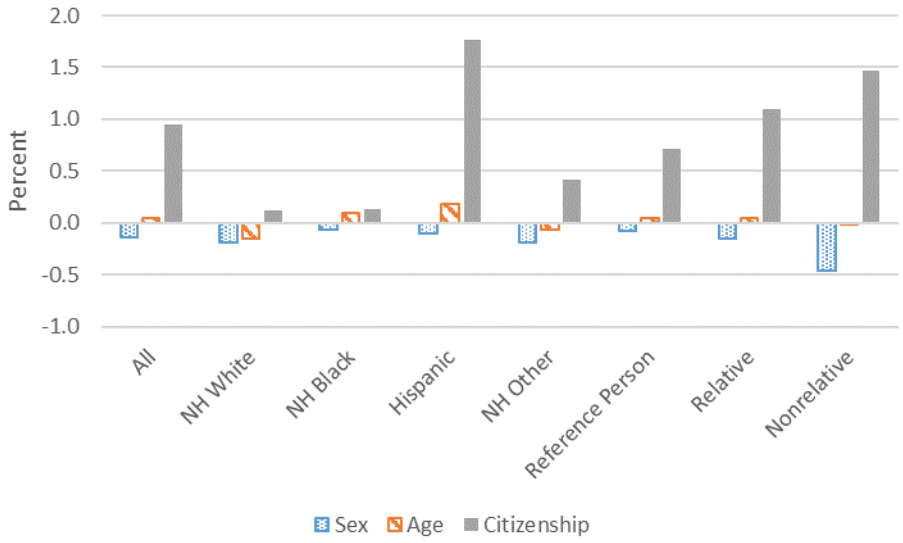
Next, we study whether nonresponse rates have been changing over time. Figure 2 has the same layout as Figure 1. It displays the difference in item nonresponse rates between the 2013 and 2016 ACS for the indicated variable.[16] A positive value indicates an increase in the item nonresponse rate, while a negative value indicates a decrease in the same rate. Figure 2, Panel A reports the difference in rates for the entire survey population as well as subgroups (see also Appendix Table A3 for the rates in the 2000 and 2010 Census short forms). Notice that item nonresponse rates for sex have gone down over time. However, item nonresponse for age and citizenship have increased, and, in particular, the increase in citizenship item nonresponse is largest for Hispanics and nonrelatives.

Figure 2. Difference in American Community Survey (ACS) Item Nonresponse between 2013 and 2016



Panel A. Difference in Item Nonresponse

[15] The Census Numident, which contains all Social Security card applications, is currently the Census Bureau's most complete and reliable administrative record source of citizenship data. For more details, see Section 3.2.
[16] Appendix Table A5 shows citizenship item nonresponse rates in 2013 and 2016 separately for mail-in and internet responses.



Panel B. Difference in Census Numident-Identified Noncitizen Item Nonresponse

Source: American Community Survey (ACS) and Census Numident, 2013 and 2016.

Note: Administrative record noncitizens make up 6.7 percent of the overall 2016 ACS sample.

Figure 2 Panel B shows the same differenced rates, but for those who are identified as noncitizens in the Census Numident. The trends over time are relatively similar for sex and age, with minimal changes. However, item nonresponse to the ACS citizenship question increased for all noncitizen groups, rising by 1.5 percentage points for nonrelatives and 1.8 percentage points for Hispanics. Hispanics, nonrelatives, and noncitizens are particularly sensitive to answering the citizenship question in the ACS, and that sensitivity has increased in recent years.

Table 1 shows break-off rates for the 2016 ACS internet self-responses (ISR) separately by question screen. Using this table, we examine which questions are subject to higher break-off rates. Higher break-off rates indicate potentially sensitive items. They are used as an indicator to inform when the respondent might stop answering the rest of the questions on a survey (Census Bureau 2013). A break-off is the moment in time during which a respondent decides not to continue with the survey and leaves the on-line survey. Break-off rates are highest for Hispanics and lowest for non-Hispanic whites in all question screens. Citizenship-related questions have the most heterogeneous rates across race/ethnicity groups: the ratio of break-off rates for Hispanics versus non-Hispanic whites is much higher for year of entry and citizenship than any of the other question screens in the ACS, except for English proficiency (included in Table 1 for reference purposes). In contrast, financial and work-related questions are sensitive for all groups. This again suggests that citizenship-related questions are more sensitive for Hispanics.

Table 1. 2016 ACS Internet Self-Response Break-off Rates (%) by Screen

	Non-Hispanic White		Non-Hispanic Other		Hispanic	
	(%)	S.E.	(%)	S.E.	(%)	S.E.
Work Location	0.642	0.011	1.045	0.032	1.246	0.038
Place of Birth	0.448	0.009	0.766	0.026	0.961	0.039
Wage Amount	0.589	0.006	0.691	0.029	0.751	0.032
Work Last Week	0.257	0.006	0.407	0.010	0.597	0.024
Work for Wages	0.365	0.009	0.459	0.019	0.590	0.028
Type of Employee	0.221	0.007	0.367	0.011	0.399	0.026
Verify Income	0.198	0.007	0.263	0.016	0.368	0.021
Citizenship	0.035	0.002	0.268	0.016	0.363	0.026
Health Insurance	0.188	0.006	0.331	0.015	0.336	0.019
Highest Level of Education	0.167	0.005	0.257	0.015	0.298	0.019
Work Duties	0.143	0.005	0.223	0.015	0.266	0.020
Year of Entry into U.S.	0.022	0.002	0.119	0.009	0.260	0.021
Taxes	0.164	0.005	0.182	0.014	0.259	0.019
Interest, Dividends Income	0.209	0.006	0.179	0.013	0.242	0.020
Residence Last Year	0.104	0.004	0.182	0.014	0.232	0.016
English Proficiency	0.003	0.001	0.020	0.005	0.036	0.007
Total Non-Breakoff	90.52	0.040	85.93	0.109	82.41	0.145

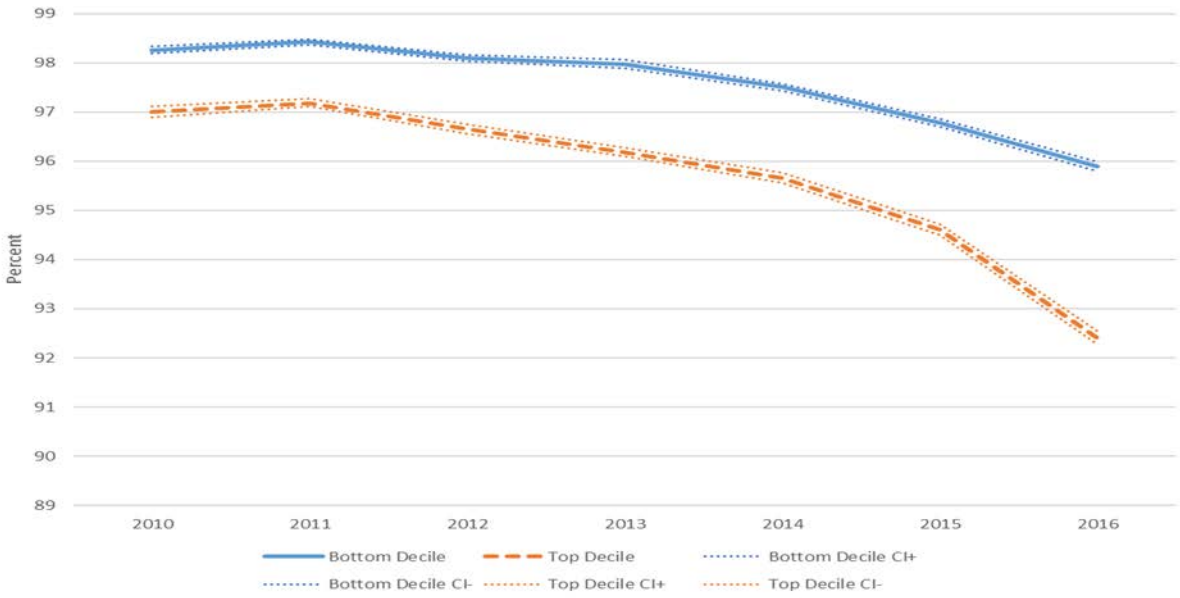
Source: 2016 ACS.

Notes: These are the top fifteen screens, sorted by Hispanic break-off rate. English proficiency and total non-breakoff are also included for reference. The rates are unweighted. The standard errors are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5.

Another alternative for measuring sensitivity of response is to examine the extent to which unit nonresponse changes. Unit nonresponse refers to a situation where no one in the household (or unit) responds to the survey. Figure 3 shows ACS unit nonresponse rates from 2010 to 2016 for housing units in the decile of tracts with the highest percent of noncitizens (25.5 percent noncitizens or more), and those in the decile of tracts that have the lowest percent of noncitizens (0.6 percent or less).[17] Tracts with noncitizen shares in the top decile have lower levels of unit response. In tracts with the highest concentrations of noncitizens, unit response rates have decreased over time and show a sharper drop between 2015 and 2016 than for units in tracts with the lowest concentrations of noncitizens.

[17] An internet response option was introduced to the ACS in 2013. Baumgardner, Griffin, and Raglin (2014) show that this was associated with an increase in self-response rates for economically advantaged groups and a decrease for economically disadvantaged groups, which could help explain the widening of the gap between these two tract groups in 2013. It cannot explain the further widening of the gap in 2016, however.

Figure 3. ACS Unit Response Rate by Tract-Level Share of Noncitizens



Source: American Community Survey (ACS), 2010-2016. The deciles of the distribution for noncitizen share of the tract population are 2011-2015 5-year ACS estimates.

Notes: The noncitizen share is 0.0 to 0.6 percent in the bottom decile and 25.5 to 100 percent in the top decile. The confidence intervals (CI) are at the 90 percent level, calculated via the successive differences replicate methodology, using 80 ACS replicate weights (see American Community Survey (2014)).

The data shown in this section provide preliminary evidence that unit nonresponse and citizenship item nonresponse rates are low in the population as a whole. The very low unit and item nonresponse rates among citizens and non-Hispanics mask increasingly higher noncitizen and Hispanic nonresponse rates, however.

3.2 Administrative Record Coverage

An alternative way to obtain citizenship information is to use data collected in the administration of government programs or by commercial data resellers. Respondent sensitivity to answering the question should be less of an issue with administrative sources, since proof of citizenship status is required to determine eligibility for a passport, a job, or government benefits. However, administrative data have incomplete coverage for other reasons, as discussed in this subsection.

Among the sources in Table 2, the Census Numident is the most complete and reliable administrative record source of citizenship data currently available to the Census Bureau. The Numident file is a record of individual applications for Social Security cards and certain subsequent transactions for those individuals. Unique, life-long Social Security Numbers (SSNs) are assigned to individuals based on these applications. In addition, a full record of all changes to the account information (such as change of name) is also maintained. To obtain an SSN, the

applicant must provide documented identifying information to the Social Security Administration (SSA). Through the "enumeration at birth" program, children can be issued an SSN when they are born.[18] Examples of data elements on a Numident record include name, date and place of birth, parents' names, and date of death. The SSA began requiring evidence of citizenship in 1972. Hence, citizenship data for more recently issued SSNs should be reliable as of the time of application.[19] SSA is not automatically notified when previously noncitizen SSN holders become naturalized citizens, however, so some naturalizations may be captured with a delay or not at all. To change citizenship status on an individual's SSN card, naturalized citizens must apply for a new card, showing proof of the naturalization (U.S. passport or certificate of naturalization).[20] Naturalized citizens wishing to work have an incentive to apply for a new card showing their U.S. citizenship, because noncitizen work permits expire, and the Numident is used in combination with U.S. Citizenship and Immigration Services (USCIS) data in the E-Verify program that confirms that job applicants are eligible to work.

Whether or not citizenship data are collected on the 2020 Census questionnaire, administrative records may be useful for editing and imputing the citizenship variable, when necessary.[21]

[18] A parent can apply for the infant's SSN at the hospital where the infant is born. Otherwise, applications for U.S.-born persons require an original or certified copy of a birth record (birth certificate, U.S. hospital record, or religious record before the age of five including the date of birth), which SSA verifies with the issuing agency, or a U.S. passport. Foreign-born U.S. citizen applications require certification of report of birth, consular report of birth abroad, a U.S. passport, a certificate of citizenship, or a certificate of naturalization. Noncitizen applications require a lawful permanent resident card, machine readable immigrant visa, arrival/departure record or admission stamp in an unexpired foreign passport, or an employment authorization document. See https://www.ssa.gov/ssnumber/ss5doc.htm. The enumeration at birth was rolled out starting in 1987, and 45 states, Puerto Rico, the District of Columbia, and New York City had signed agreements to offer it by 1991. Today over 90 percent of parents use this process in all 50 states plus Puerto Rico and the District of Columbia. See https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html.

[19] A detailed history of the SSN is available at https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html (Exhibit 1). For some categories of persons, the citizenship verification requirements started a few years later, but all were in place by 1978.

[20] For more information, see https://www.ssa.gov/ssnumber/ss5doc.htm.

[21] Data edits refer to updating data when there is a clear error either in data entry or in response. Imputations occur when the individual or household did not answer a survey or questions on a survey. They involve modeling a most likely response for that individual or household using other available data.

Table 2. Administrative Record (AR) Sources Currently Held and/or Under Negotiation for Acquisition

Administrative Records Data with Citizenship Info. Currently Held	Universe
Census Numident	National-level file of SSA transactions
HHS TANF	National Level (not full content for all states)
Alaska Permanent Fund	Alaska residents
Colorado Leap	Colorado low income energy assistance program
Some State SNAP/TANF	State-level program participants
Army	Active duty and retired soldiers and family members
Bureau of Prisons	Federal prison inmates
Commercial Files	Purchased data from data resellers
Administrative Records Data with Citizenship Info Under Negotiation for Acquisition	**Universe**
Department of Homeland Security United States Citizenship and Immigration Services	National-level file of Lawful Permanent Residents, Naturalizations
Department of Homeland Security United States Customs and Border Protection	National-level file of Customs and Border transaction data
Department of State Passport Services	National-level passport transaction data

Table 3 shows the coverage of the 2010 Census population by the 2010 Numident and ITINs.[22] Ninety-one percent of persons in the 2010 Census can be assigned a Protected Identification Key (PIK) by the Person Identification Validation System (PVS).[23] Once a PIK is assigned, the vast majority of records are matched to the 2010 Numident (98.2 percent in Table 3). Most of the PIKs associated with persons not in the 2010 Numident are derived from linkage to Individual Taxpayer Identification Numbers (ITIN), issued by the Internal Revenue Service to persons who do not have

[22] Rastogi and O'Hara (2012) used an earlier version of the crosswalk between the Numident and ITINs and the 2010 Census, and we show results using that version in Table A6. The enhanced crosswalk in Table 3 uses additional household and geospatial information to increase person linkage, and it has much greater coverage of ITINs. See Bond et al. (2014) for details.

[23] See NORC (2011) and Layne, Wagner and Rothhaas (2014) for details about the process used to assign and the quality of the PIKs used in data linkage at the Census Bureau.

and are not eligible to obtain SSNs, but are required to file a federal individual income tax return (4.3 million person links derived from ITINs vs. 804,000 person links that are not derived from ITINs). Among persons with nonmissing citizenship in the 2010 Numident, 91.3 percent are U.S. citizens.

Approximately 20.9 percent, or 57.6 million of the 2010 Numident records have missing citizenship status. Many older persons did not report citizenship when applying for an SSN, which was not required prior to 1972. Of these older persons with missing citizenship, 7.0 million have either passed away by 2017 or are likely to do so by 2020 (since they would be over 100 years old). Of the remaining 50.7 million persons with missing citizenship in the 2010 Numident, it becomes nonmissing for 5.8 million of them by 2017, nearly all switching to U.S. citizens. About 42.5 million of those still missing citizenship in 2017 were born in the U.S. We treat U.S.-born persons missing citizenship as administrative record citizens in our analysis.[24] This leaves just 2.5 million foreign-born persons with missing citizenship, some of whom could be noncitizens. In the analysis, we treat foreign-born persons with missing citizenship as having missing administrative record citizenship.

Appendix Table A7 shows that among persons who are missing citizenship, alive in 2017, and born after 1919, those who are foreign-born have a much lower propensity to be linked to the 2010 Census (36.3 percent vs. 74.5 percent for U.S.-born persons). Many of the foreign-born people missing citizenship in the Numident are presumably residing outside the U.S. and thus will not be counted in the 2020 Census.[25]

[24] Analysis in later sections of this paper labeled "initial assumptions" instead treats all persons with missing Numident citizenship values as AR citizens, whether they are U.S.- or foreign-born. This includes Table 6, Figures 10B, 11A, 12A, and 12C and Appendix Tables A8 and A9.

[25] An example is persons who received temporary work visas prior to when evidence of citizenship was required to receive an SSN and who have since returned to their home countries.

Table 3. Administrative Record (AR) Coverage of the 2010 Census

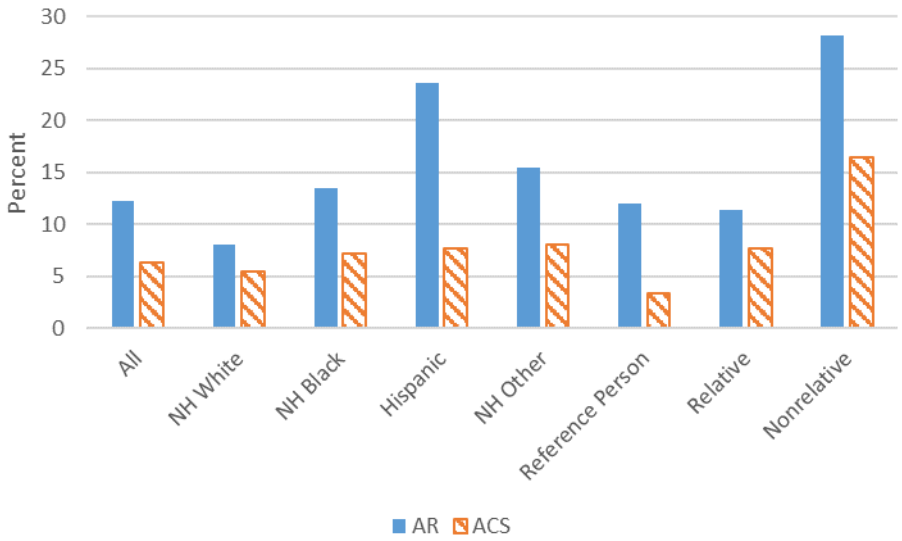
	Count	Percent of 2010 Census Population	Percent of Matched Sample
No PIK, not sent to PVS	10,260,000	3.3	
No PIK, failed in PVS	17,490,000	5.7	
PIK, but not in 2010 Numident, not an ITIN	804,000	0.3	
PIK, but not in 2010 Numident, is an ITIN	4,326,000	1.4	1.5
2010 Numident U.S. Citizen	199,300,000	64.6	71.1
2010 Numident Noncitizen	18,970,000	6.1	6.8
2010 Numident Missing Citizenship	57,620,000	18.7	20.6
Of which:			
Alive in 2017, born after 1919	50,670,000	16.4	18.1
Of which:			
2017 Numident U.S. Citizen	5,678,000	1.8	2.0
2017 Numident Noncitizen	70,500	0.0	0.0
2017 Numident Missing Citizenship	44,920,000	14.5	16.0
Of which:			
U.S.-born	42,460,000	13.8	15.2
Foreign-born	2,464,000	0.8	0.9
Total	308,745,538	100.00	100.00

Source: 2010 Census Unedited File (CUF) and 2010 and 2017 Census Numident Files.

Notes: The 2010 Census Numident File is used for all calculations with "Numident" in the label. The 2017 Census Numident File is used to calculate the number alive in 2017 and born after 1919 and the foreign-born share of them. PVS is the Person Identification Validation System used to assign PIKs. PIK is Protected Identification Key, which is a unique person identifier.

Figure 4 shows the share of persons in the 2016 ACS for whom administrative record citizenship status is not available, as well as the ACS citizenship allocation rate (including both item nonresponse and edits to original responses; i.e., the share of persons for whom the value tabulated is not the respondent's answer). The missing data rates are higher for administrative records (AR) than the ACS, and both sources' rates are higher for minorities and nonrelatives. The variability in coverage is higher for AR than the ACS.

Figure 4. Percent without Administrative Record or ACS Citizenship in 2016



Source: American Community Survey (ACS) and Census Numident, 2016.

Note: For the ACS this is the citizenship item allocation rate, which includes both item nonresponses and edited values.

As shown in Appendix Table A8, the percent of persons in the ACS who cannot be linked to citizenship in AR increases from 8.5 to 10.9 percent between 2010 and 2016. Note that the linkage between the ACS and administrative data from the SSA Numident and IRS ITIN tax filings depends on two factors: (a) the quality of the personally identifiable information (PII) on the ACS response and (b) whether the ACS respondent is in the SSN/ITIN universe.

With respect to the quality of the PII on the ACS, there may be insufficient information on the ACS due to item nonresponse to allow a successful match using the production record linkage system. There may also be more than one record in the Numident or ITIN IRS tax filings that matches the person's PII. Finally, there may be a discrepancy between the PII provided to the ACS and the PII in the administrative records.

Alternatively, the person may not be in the Numident or ITIN IRS tax filing databases, because they are out of the universe for those administrative systems. This happens when the person is a citizen without an SSN, or when the person is a noncitizen who has not obtained an SSN or ITIN.

Very few of the unlinked cases are due to insufficient PII in the ACS or multiple matches with administrative records. The vast majority of unlinked ACS persons have sufficient PII, but fail to match any administrative records sufficiently closely. This means that most of the nonmatches are because the ACS respondent is not in the administrative record universe.

The incidence of ACS persons with sufficient PII but no match with administrative records increased between 2010 and 2016. One contributing factor is that the number of persons linked to ITIN IRS tax filings in 2016 was only 35 percent as large as in 2010,[26] suggesting that either fewer

[26] This percentage uses survey weights. Unweighted, it is 39 percent.

of the noncitizens in the 2016 ACS had ITINs, or more of them provided PII in the ACS that was inconsistent with their PII in IRS records.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than currently held administrative records. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the citizenship status distribution of the sampled non-respondents is statistically the same as that of respondents with similar related characteristics. They might not actually be similar, however, even when selecting the allocation of citizenship status using basic characteristics. For example, Hispanics who respond to the survey might be different from Hispanics who do not respond in various characteristics (including immigration status). Additionally, our unit and item nonresponse analysis in Section 3.1 above casts serious doubt on this assumption, suggesting that those who do not respond to either the entire ACS or the citizenship question on the ACS are not statistically similar to those who do. In particular, their responses to the citizenship question would not be well predicted by the answers of those who did respond.

To reduce the AR coverage gaps, the Census Bureau is considering the possibility of acquiring access to several other national citizenship-related files listed in Table 2. United States Customs and Immigration Services (USCIS) administrative records on naturalizations and lawful permanent residents (LPR), and Customs and Border Protection transaction records on border entries can partially address the weaknesses of the Numident. Through preliminary project development discussions with USCIS, we were informed that USCIS records provide up-to-date information since 2001 (and possibly back to 1988, but with incomplete records prior to 2001). These will fill some gaps for naturalized citizens, lawful permanent residents, and persons with extended visa applications without SSNs, as well as naturalized citizens who did not inform SSA about their naturalization. These data do not cover naturalizations occurring before 1988, and they miss some between 1988 and 2000. USCIS records do not always cover children under 18 at the time a parent became a naturalized U.S. citizen. These children automatically become U.S. citizens under the Child Citizenship Act of 2000. The USCIS receives notification of some, but not all, of these child naturalizations. Others inform the U.S. government of their U.S. citizenship status by applying for U.S. passports, which are less expensive than the application to notify the USCIS. USCIS visa applications list people's children but the information may not be in electronic form.

U.S. passport administrative records available from the State Department can help plug the gaps for child naturalizations, missing status on the Numident, and out-of-date citizenship information on the Numident. Since U.S. citizens are not required to have a passport, however, these records will also have coverage gaps.

The acquisition of these sources would also improve record linkage for noncitizens by allowing the construction of a supplementary record linkage master list for such people, who are currently only in scope for receiving a PIK if they apply for and receive either an SSN or ITIN. Improved record linkage would not only facilitate greater use of administrative record citizenship data, but it could also permit other uses of these administrative records in 2020 Census operations to lower costs and raise quality. Noncitizens are a hard-to-count population (as evidenced by the lower ACS

unit response rates in tracts with more noncitizens in Figure 3), making having reliable administrative records on them particularly valuable.

If the Census Bureau were to obtain each of these files, the remaining AR citizenship data gaps would include the following categories:

1. U.S. citizens from birth with no SSN or U.S. passport. They will not be processed by the production record linkage system used for the 2020 Census, because their PII won't be found in the PVS reference files.

2. U.S. citizens from birth born outside the U.S., who do not have a U.S. passport, and either applied for an SSN prior to 1974 and were 18 or older, or applied before the age of 18 prior to 1978. These people will be assigned PIKs, but none of the administrative sources discussed above will reliably generate a U.S. citizenship variable.

3. U.S. citizens who were naturalized prior to 2001 and did not inform SSA of their naturalization, because they originally applied for an SSN after they were naturalized, and it was prior to when citizenship verification was required for those born outside the U.S. (1974). These people either already had an SSN when they were naturalized, and they didn't inform SSA about the naturalization, or they never applied for an SSN. The former group has inaccurate data in the Numident. The latter group will not be assigned a PIK.

4. U.S. citizens who were automatically naturalized if they were under the age of 18 when their parents became naturalized in 2000 or later, and they did not inform USCIS or receive a U.S. passport. Note that such persons would not be able to get an SSN with U.S. citizenship on the card without either a U.S. passport or a certificate from USCIS. These people will also not be assigned a PIK.

5. Lawful permanent residents (LPR) who received that status prior to 2001 and either do not have an SSN, or they applied for an SSN prior to when citizenship verification was required for those born outside the U.S. (1974). The former group will not be found in the PVS reference files. The latter group has inaccurate data in the Numident.

6. Noncitizen, non-LPR, residents who do not have an SSN or ITIN and who did not apply for a visa extension. These persons will not be found in PVS.

7. Persons with citizenship information in administrative data, but the administrative and Census data cannot be linked due to missing or discrepant PII.

It is uncertain whether Census Bureau household survey data could reliably fill the above gaps when their person record cannot be assigned a PIK or when they have a PIK but the administrative record lacks up-to-date citizenship information. Persons in Category 6 have a strong incentive to provide an incorrect survey answer, if they answer at all, due to concerns about the data being used for enforcement.[27] Presumably a significant, but unknown, fraction of persons without PIKs are in

[27] Title 13, U.S.C. prohibits the use of Census data for enforcement purposes, but respondents may still have this concern.

Category 6. Distinguishing these people from the other categories of persons without PIKs is inherently inexact, because there is no feasible method of independently verifying their citizenship status.

4. Data Reliability

To assess the reliability of citizenship data, we compare the responses to the 2000 Census long form and 2010 and 2016 ACS citizenship questions with the administrative record (AR) citizenship variable (from the 2002, 2010, and 2016 Numidents and ITINs for the latter two years).[28] Since previous studies suggest that Census survey-AR discrepancies are greater for Hispanics, and the CVAP tables show citizen counts by race/ethnicity and voting age, we show discrepancies separately by race/ethnicity and the voting-age population (age 18 and over). Appendix Tables A8 and A9 show a full set of results for all three years, while the discussion in this section focuses on the 2016 comparison.

Discrepancies between AR and ACS citizenship could be due to several causes: (1) Linkage errors result in the administrative records not matching to the right people in the ACS. The relative discrepancy rates would vary depending on whether AR citizens or noncitizens have more linkage errors. One might expect unrelated persons in the household to have more linkage errors than relatives of reference persons, since PII quality is likely to improve with familiarity. (2) AR incorrectly report that the person is a citizen. This would appear as AR citizen-ACS noncitizen discrepancies. (3) AR are out of date, missing some naturalizations captured by the ACS. This would show up as AR noncitizen-ACS citizen discrepancies.[29] (4) The respondent does not know the person's citizenship status and guesses wrong. This is most plausible for unrelated persons and least so for the reference person. (5) The respondent misunderstands the question and answers incorrectly, despite actually knowing the citizenship status. It is not clear whether this would lead to more AR citizen-ACS noncitizen or AR noncitizen-ACS citizen discrepancies, but it should not vary across reference person, related persons, and unrelated persons. (6) The respondent knows the person's citizenship status and misreports it. Here the reference person may have a harder time justifying item nonresponse (implying (s)he does not know her/his own citizenship), so the way to keep from attracting attention is to say (s)he is a U.S. citizen. When asked about others, the respondent can more easily say (s)he does not know. This factor is likely to be more relevant when people have heightened concerns that the data will be used for immigration enforcement.

Of the candidate reasons (1) through (3) relevant for administrative records, linkage errors (reason 1) would be the most difficult to overcome. If linked to the wrong people, even perfect administrative records will produce inaccurate statistics. Though improvements can be made to record linkage methods, the linkage quality also depends on the quality of PII supplied by the sources being linked. In contrast, the acquisition of more timely administrative record sources

[28] The 2002 Numident is the closest available Numident to the 2000 Census.

[29] Note that as the Census Bureau receives more administrative record sources of citizenship data, the probability that the administrative records are incorrect should fall.

should reduce missing naturalizations problems (reason 3). The use of additional administrative record sources can also help illuminate instances where currently held administrative records are more likely to be incorrect (reason 2).[30]

Guessing wrong (reason 4) and misunderstanding the question (reason 5) would reduce precision (i.e., increase statistical variability), but it is not clear that either would result in biased estimates. In contrast, intentional misreporting (reason 6) is likely to result in reduced accuracy (more bias), since citizens and noncitizens may have different incentives to misreport status. Of these three reasons, the extent of intentional misreporting is most likely to vary across geographical areas and over time, depending on the degree of concern about personal security.

Figure 5 Panel A shows that a remarkably high 99.6 percent of U.S. citizens (according to administrative records) report being U.S. citizens in the 2016 ACS.[31] This suggests that when AR report the person is a citizen, (s)he is actually a citizen, and reason (2) is not an important factor. The discrepancy rate is higher for Hispanics (2.0 percent) and other minorities (1.3 percent) than for non-Hispanic white individuals. The discrepancy rate is higher for nonrelatives than relatives of the respondent, and for relatives than reference persons, consistent with the reference person knowing other people's status less well than his/her own.

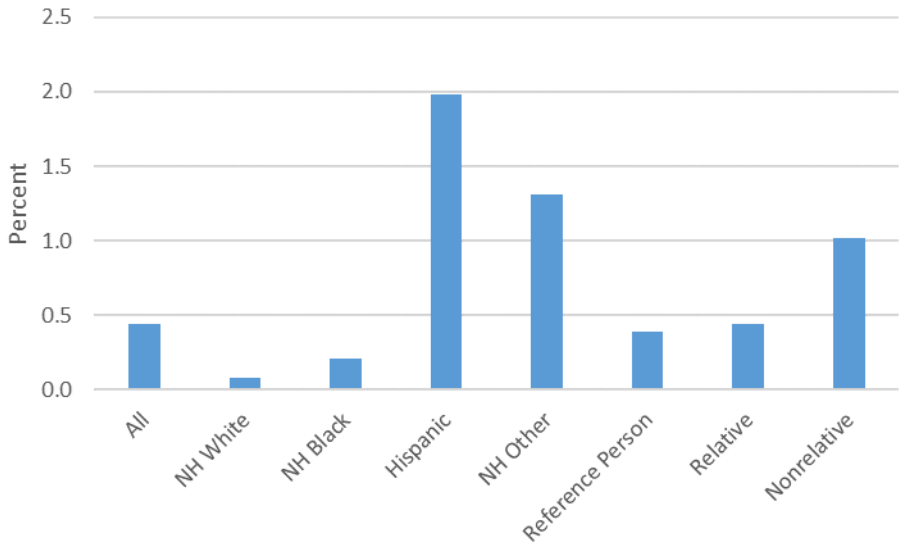
Discrepancy rates are higher for those individuals identified as U.S. noncitizens in administrative records: 37.6 percent report being U.S. citizens in the ACS, as shown in Figure 5 Panel B. This implies that ACS estimates of the U.S. citizen population are higher than they would be if one were to use currently available administrative records.[32] The ordering of rates across groups is reversed compared to the AR citizen-ACS noncitizen rates. Here non-Hispanic white individuals have the highest discrepancy rate and Hispanic individuals the lowest. This means that the difference between ACS citizen and AR citizen population estimates is greatest for non-Hispanic white individuals and lowest for Hispanic individuals. This contrasts with Van Hook and Bachmeier's (2013) conclusion based on aggregates that self-reported naturalizations by persons of Mexican origin are most likely to be incorrect.[33,34]

The AR noncitizen-ACS citizen discrepancy rate is highest for the reference person, followed by relatives and then nonrelatives. This pattern is not a clear outcome of out of date administrative

[30] For example, if a person is a foreign-born citizen in one administrative record source, but other administrative records and the survey response each say the person is a noncitizen, one might have more confidence in selecting noncitizen than when having only the first administrative record source and the survey response.
[31] This is even higher than the agreement rate for sex in the 2010 Census vs. the Numident, which is 99.4 percent. See Rastogi and O'Hara (2012).
[32] Note that since we are unable to compare records that are missing in one or both sources, the estimates provided in this section may understate the difference between the ACS estimate of the U.S. citizen population and the true value, especially since most unauthorized persons (other than the small fraction with ITINs) are missing AR citizenship data here.
[33] Hispanics make up the largest number of AR noncitizen-ACS citizen persons (2.6 million), compared to 2.5 million non-Hispanic other minorities, 1.7 million non-Hispanic whites, and 800,000 non-Hispanic blacks, which may be why previous studies' analysis of aggregated data find the largest administrative record-survey differences to be among Hispanics. But the discrepancy rate is more relevant for evaluating quality than the absolute number of discrepancies.
[34] According to 2016 1-year ACS data in American Factfinder Table S0201 (American Community Survey 2016c), 63.2 percent of Hispanics are of Mexican origin.

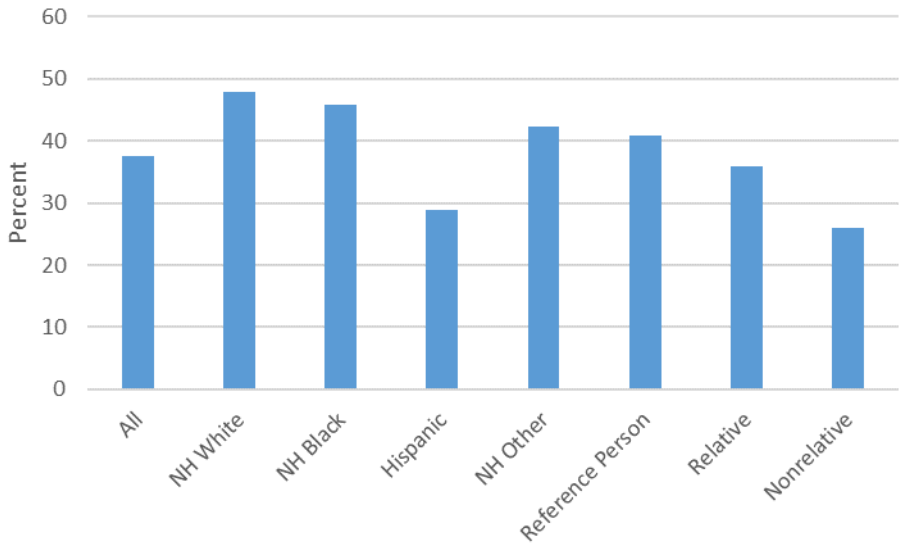
records (reason 3), lack of knowledge about others' status (reason 4), or misunderstanding the question (reason 5). Recall that citizenship item nonresponse is highest for nonrelatives and lowest for reference persons (see Figure 1). This suggests respondents behave differently when asked about their own status versus that of others. It may be easier for respondents to say they do not know the status of someone else (particularly a nonrelative) than their own status. They thus misreport their own status (reason 6), while they say they do not know the status of others.

Figure 5. Administrative Records-ACS Survey Response Citizenship Agreement



Panel A. Percent of Administrative Record Citizens who respond as 2016 ACS Noncitizens

Notes: Administrative record citizens make up 81.1 percent of the overall 2016 ACS sample, 90.1 percent for non-Hispanic white, 81.5 percent of non-Hispanic black, 60.2 percent of Hispanic, 62.5 percent of non-Hispanic other race, 81.1 percent of reference persons, 82.1 percent of relatives, and 64.8 percent of nonrelatives. See Appendix Table A10.



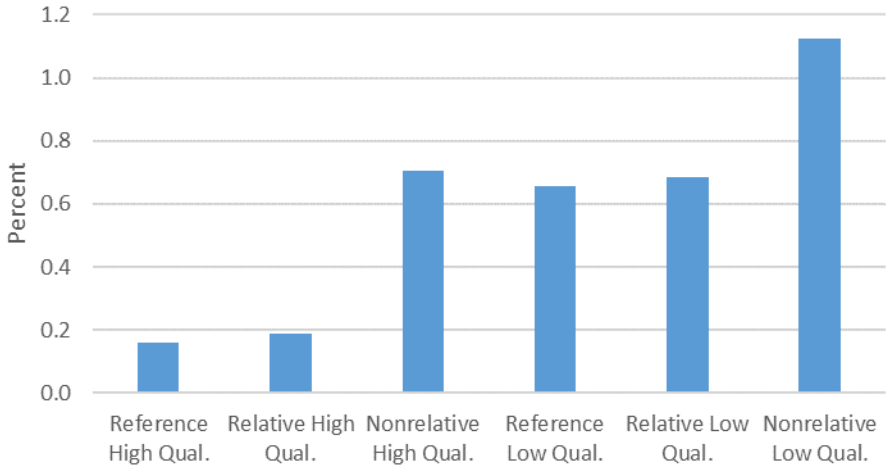
Panel B. Percentage of Administrative Record Noncitizens who respond as 2016 ACS Citizens

Notes: Administrative record noncitizens make up 6.7 percent of the overall 2016 ACS sample, 1.9 percent for non-Hispanic white, 5.1 percent of non-Hispanic black, 16.2 percent of Hispanic, 22.0 percent of non-Hispanic other race, 6.9 percent of reference persons, 6.5 percent of relatives, and 7.1 percent of nonrelatives. See Appendix Table A10.

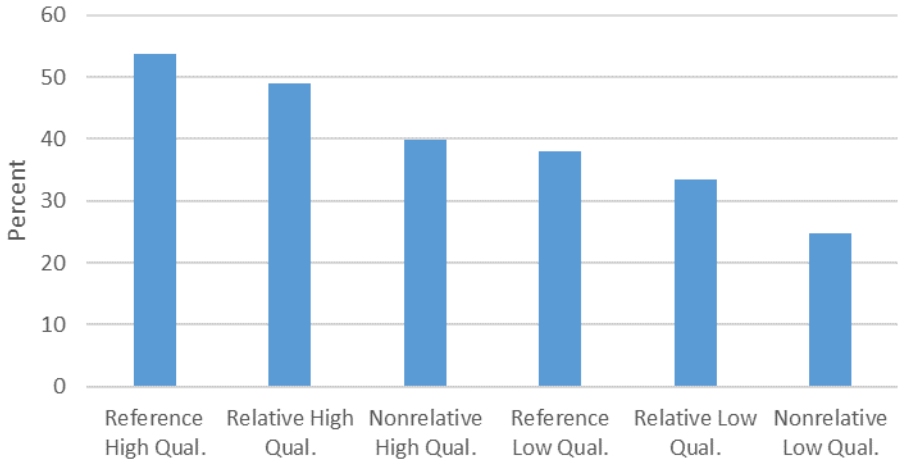
Source: American Community Survey (ACS) 1-year file and Census Numident, 2016.

We show the AR citizen-ACS noncitizen and AR noncitizen-ACS citizen discrepancies separately for higher- and lower-quality linkages and by reference person vs. relative vs. nonrelative categories in Figure 6. For AR citizen-ACS noncitizen discrepancies, the rates are lowest for the reference person and highest for nonrelatives, likely due to people being able to report their own PII more accurately than that of others. Records with high-quality links have lower discrepancy rates, consistent with linkage errors being a contributing factor to these discrepancies. The patterns reverse for AR noncitizen-ACS citizens. Higher-quality linked records actually have higher discrepancy rates, so linkage errors (reason 1) do not appear to explain the AR noncitizen-ACS citizen discrepancies. This pattern holds regardless of the type of person the reference person is responding about (oneself, a relative, or a nonrelative).

Figure 6. Quality of the Citizenship Question Responses by Relation to Reference Person and Higher- vs. Lower-Quality Linkage



Panel A. AR Identifies as a Citizen and 2016 ACS Identifies as a Noncitizen



Panel B. AR Identifies as a Noncitizen and 2016 ACS Identifies as a Citizen

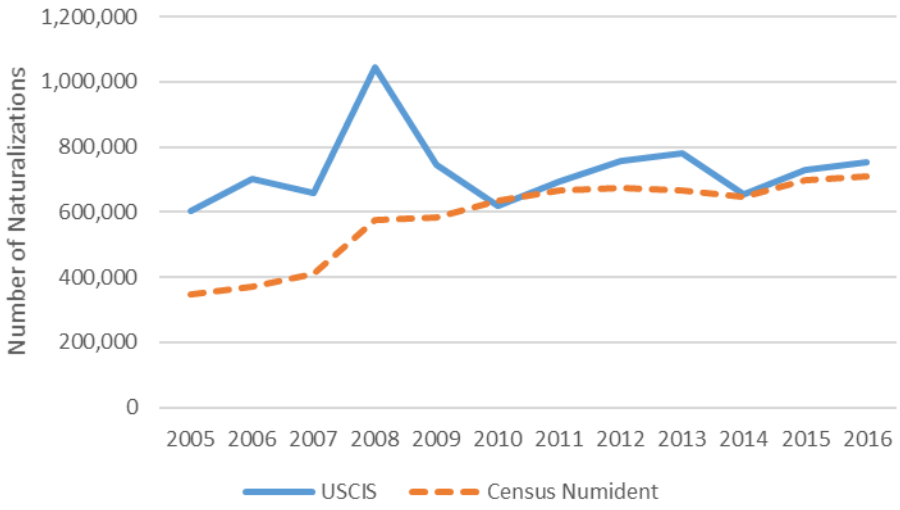
Source: American Community Survey (ACS) and Census Numident, 2016.

Notes: High-quality linkage is defined as having an above-median linkage confidence score on the first linking attempt (pass), and lower-quality is all others. The weighted sample shares of the ACS are 18.1 percent for reference person high-quality linkage, 23.9 percent for relative high-quality linkage, 0.6 percent for nonrelative high-quality linkage, 20.3 percent for reference person low-quality linkage, 33.8 percent for relative low-quality linkage, and 3.2 percent for nonrelative low-quality linkage. See Appendix Table A11.

To evaluate further the hypothesis that AR are out of date (reason 3), we make comparisons to USCIS statistics. In the AR-ACS citizenship status comparison above, we estimate 7,605,000 persons are AR noncitizens-ACS citizens. This is equivalent to the Numident missing all the

naturalizations reported by USCIS back to 2007, plus some of 2006. Figure 7 shows the annual number of persons who first entered the Numident as noncitizens and switch to U.S. citizenship in each particular year, as well as the number of naturalizations according to USCIS statistics.[35] USCIS reports significantly more naturalizations prior to 2010, but there is little difference subsequently. This suggests that if the main reason for the discrepancies were out-of-date Numident citizenship, the Numident would have to be missing many naturalizations that occurred long ago.

Figure 7. Estimated Annual Naturalizations in Census Numident Data versus USCIS Statistics

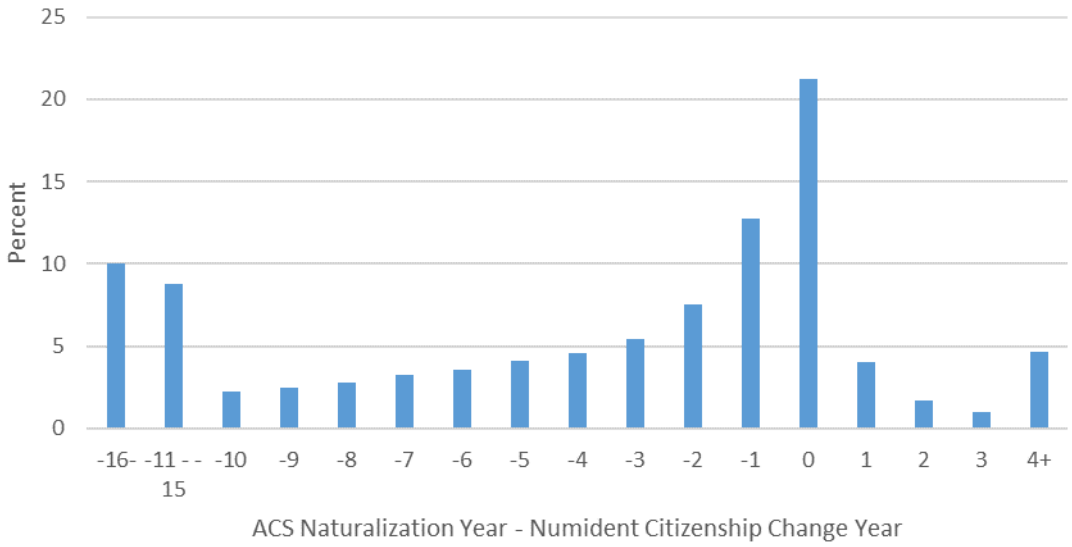


Source: USCIS Immigration Yearbooks and 2017 Census Numident.

We compare the ACS naturalization year and the year when citizenship switched to U.S. citizen in the Numident among persons with naturalized citizen status in both sources in Figure 8.[36] For 67.4 percent of these persons, the ACS naturalization year is earlier than the Numident citizenship change year, and 33.1 percent have an ACS naturalization year that is more than five years prior. Just 11.3 percent have a later ACS naturalization year. This is consistent with tardy notification to SSA about naturalizations.

[35] The Numident switches do not include persons who did not have an SSN prior to being naturalized. According to USCIS officials, the percentage of persons naturalized in 2014 who did not previously have an SSN is 0.33 percent, and it is 0.40 percent in 2015, suggesting that this type of Numident omission is negligible, at least recently.
[36] The Numident citizenship change year is the year when citizenship changed from noncitizen to citizen in the data.

Figure 8. Difference between ACS Naturalization and Numident Citizenship Change Years



Source: American Community Survey (ACS) and Census Numident, 2016. The sample is persons who are naturalized citizens in both sources, and the ACS citizenship value is as reported by the respondent.

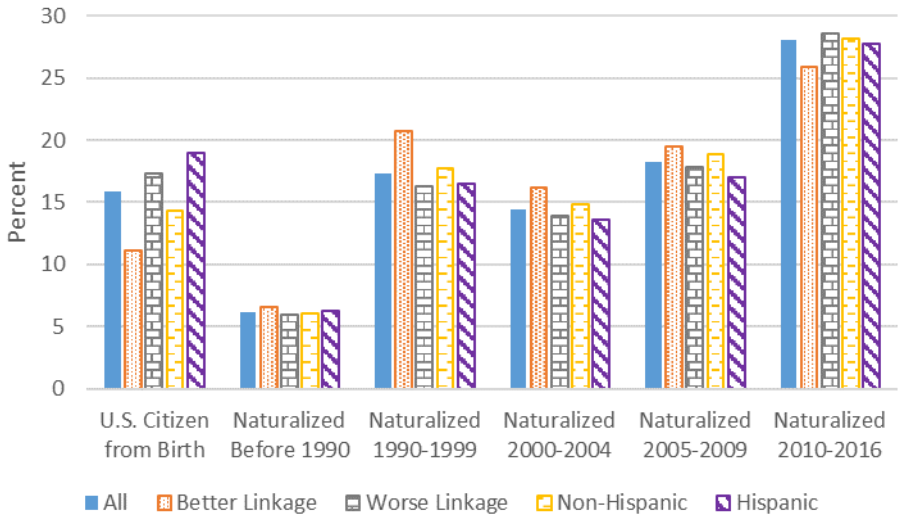
Figure 9 shows the distribution of AR noncitizen-ACS citizens by naturalization year. Approximately 15.9 percent report being citizens from birth, which, if true, would mean that the Numident is not out of date for these people, but incorrect from the first SSN application. This possibility seems unlikely, given that proof of citizenship status must be presented to SSA when applying for an SSN, whereas the ACS citizenship response is not checked. A third of the ACS-reported naturalizations (2.1 million) occurred between 2010 and 2016, while the total gap between USCIS naturalizations and Numident switches from noncitizen to citizen between 2010 and 2016 is several times less than that, at 288,000.

Figure 9 shows that the AR noncitizen-ACS citizen naturalization distributions are very similar for Hispanics and non-Hispanics. The results are contrary to Van Hook and Bachmeier's (2013) finding that citizenship misreporting by persons saying they were naturalized more than five years ago primarily occurs among persons of Mexican origin, and Passel and Clark's (1997) finding that it is among those of Mexican or Central American origin.

We also explore whether the AR noncitizen-ACS citizen naturalization distributions vary with linkage quality. One might expect that if linkage quality is driving the discrepancies, then persons with higher quality links would be recently naturalized, reflecting out-of-date Numident data. In contrast, more of the persons with low quality links would be ACS citizens from birth or naturalizations long ago, since the Numident and ACS records could be for different people, and the Numident should be less likely to be out of date for citizens from birth and earlier naturalizations. Figure 9 does show a higher share of ACS citizens from birth among those with

lower quality links, but also for more recent naturalizations. This is further evidence that linkage errors are probably not an important explanation for these discrepancies.

Figure 9. Distribution of ACS Citizenship Receipt Timing for Administrative Record Noncitizen-ACS Citizens by Linkage Quality and Ethnicity



Source: American Community Survey (ACS) and Census Numident, 2016.

As a final data quality check, we calculate the 2016 ACS citizenship distribution for persons with ITINs. Though only noncitizens may have ITINs, 6.6 percent say they were born citizens, and 11.1 percent report being citizens in the ACS (see Appendix Table A12).

5. Item Response and Data Quality Regression Analysis

We estimate multivariate regressions predicting item response in Table 4 and AR-ACS discrepancies in Table 5. The item response and citizenship status disagreement regressions test whether the associations shown above are statistically significant and robust to inclusion of controls. These analyses also provide an opportunity to study other potentially relevant factors. The item response regressions are estimated separately for AR citizens, AR noncitizens, and those missing AR citizenship. The item response variables are equal to one if there is a response for the item (whether it was later edited or not), and zero otherwise. The ACS noncitizen-AR citizen dependent variable is equal to one if the person is an as-reported noncitizen in the ACS and an AR citizen, and it is zero if both sources say the person is a citizen. Analogously, the ACS citizen-AR noncitizen dependent variable is equal to one if the person is an as-reported citizen in the ACS and an AR noncitizen, and it is zero if both sources say the person is a noncitizen. The last specification in Table 5 investigates determinants of the difference between the ACS naturalization year and the year in which the status changed to citizen in the Numident among persons who were noncitizens in their first SSN application.

Besides relationship to the reference person, we include several other factors that theoretically could drive differences observed in both survey response and data quality. These include demographic characteristics such as sex, race/ethnicity, log one plus age, and its square. We also include socioeconomic characteristics such as educational attainment, working in the last week, and searching for a job in the last four weeks. Educational attainment is classified as less than high school diploma (base category), at least high school but less than a bachelor's degree, bachelor's degree, and graduate degree. Time since entry to the U.S. and reference person English language variables are included, since these variables may influence item response and discrepancies in citizenship status reporting. For our analysis, those variables are log of one plus the number of years since entering the U.S. (or since birth if born in the U.S.) and its square[37] and English language ability for those speaking another language at home (speaking only English at home is the base category). We include an indicator for better or worse quality person linkage, since it may also drive differences in survey response and data quality. An indicator for whether the response is via mail or internet (i.e., without participation by an interviewer) vs. a personal or telephone interview. According to Camarota and Capizzano (2004), item nonresponse rates are lower in in-person interviews, and foreign-born persons are more likely to take the survey via personal interview, so controlling for mode could be particularly important when comparing the behavior of citizens and noncitizens.

The associations highlighted in Figures 1-6 above are robust to inclusion of other variables and are highly statistically significant.[38] Item nonresponse and ACS noncitizen-AR citizen discrepancy rates are higher for nonrelatives, but the ACS citizen-AR noncitizen propensity is much lower, again consistent with reference persons misreporting their own citizenship, but not reporting that of others at all, especially nonrelatives. Like nonrelatives, Hispanics have a lower propensity to provide citizenship, a higher propensity to have ACS noncitizen-AR citizen discrepancies, and a lower propensity to have ACS citizen-AR noncitizen discrepancies. Better linkage is strongly associated with ACS citizen-AR noncitizen discrepancies, inconsistent with the hypothesis that these discrepancies are driven by linkage errors.

Now turning to factors not investigated in previous sections, labor market activity is positively associated with having a citizenship answer; especially for AR noncitizens (see Table 4). However, as Table 5 shows, working is also associated with both types of citizenship status disagreements, particularly ACS citizen-AR noncitizen. Reference persons who speak another language at home have a higher propensity to respond about sex, especially when their English language ability is less strong. This is also true for AR citizens for the citizenship question, but when asked to report about AR noncitizens, those speaking another language at home have much lower citizenship item response rates. Those speaking English less well also have a higher propensity to report ACS noncitizen when the person they are responding about is an AR citizen, perhaps reflecting misunderstanding of the question. However, the reference person's English language ability is positively associated with ACS citizen-AR noncitizen discrepancies, again suggesting that

[37] In cases where the person came to live in the U.S. more than once, respondents are instructed to give the latest year.
[38] In results not shown here, we also estimate item response regressions with the full sample, regardless of AR citizenship status. The patterns are similar to those described in this paragraph, except that Hispanics have higher propensity to have item response for age in the full sample.

misunderstanding the question is an important factor behind ACS noncitizen-AR citizen, but not ACS citizen-AR noncitizen discrepancies. Responding without the participation of an interviewer results in lower item response (except for age for AR noncitizens), consistent with Camarota and Capizzano (2004), and this effect is particularly strong for citizenship item response among AR noncitizens. ACS noncitizen-AR citizen discrepancies are more prevalent with interviewer participation, but ACS citizen-AR noncitizen discrepancies are much less prevalent. Interviewers may develop a rapport that encourages noncitizens to truthfully respond to what is a sensitive question for them.[39] It could also be more difficult psychologically for a respondent to misreport to another person than when they fill out a questionnaire on their own.

As shown in Table 4, the associations with citizenship item response tend to be several times stronger for AR noncitizens than for citizens, with those missing AR citizenship falling in between the other two categories. Such differences are much more muted for sex and age. This again highlights the nonrandom nature of citizenship item nonresponse.

[39] This effect may be weaker in the Census than in the ACS, however, since ACS interviewers have much more experience than most Census enumerators.

Table 4. Item Response Regressions

	Sex Item Response			Age Item Response			Citizenship Item Response		
	AR Citizen	AR Noncitizen	AR Missing	AR Citizen	AR Noncitizen	AR Missing	AR Citizen	AR Noncitizen	AR Missing
Relative	-0.159	-0.079	-0.759	-0.234	-0.224	-4.446	-0.057	-0.480	-0.106
	(0.007)	(0.016)	(0.039)	(0.013)	(0.060)	(0.123)	(0.010)	(0.082)	(0.066)
Nonrelative	-0.455	-0.309	-1.146	-2.353	-3.509	-9.533	-1.141	-7.395	-4.808
	(0.035)	(0.072)	(0.084)	(0.080)	(0.307)	(0.300)	(0.047)	(0.390)	(0.200)
Non-Hispanic	-0.136	-0.160	-0.003	-0.142	-0.227	-0.225	-0.122	-3.092	-0.979
African Amer.	(0.014)	(0.050)	(0.082)	(0.029)	(0.143)	(0.247)	(0.012)	(0.171)	(0.078)
Hispanic	0.128	0.002	0.147	0.033	0.075	2.068	-0.391	-4.432	-1.692
	(0.013)	(0.030)	(0.069)	(0.032)	(0.103)	(0.210)	(0.024)	(0.140)	(0.119)
Other Non-	0.050	0.038	0.230	-0.100	-0.108	1.229	-0.177	-2.320	-1.885
Hispanic	(0.017)	(0.028)	(0.072)	(0.034)	(0.092)	(0.230)	(0.031)	(0.129)	(0.152)
Worked in	0.174	0.073	0.694	0.334	0.149	1.872	0.915	8.687	3.773
Last Week	(0.008)	(0.024)	(0.037)	(0.017)	(0.081)	(0.132)	(0.013)	(0.141)	(0.088)
Searched for	0.045	0.017	0.668	0.457	0.466	3.834	0.769	7.414	3.494
Job	(0.020)	(0.046)	(0.063)	(0.033)	(0.126)	(0.252)	(0.016)	(0.185)	(0.114)
English Very	0.116	0.101	0.690	0.084	0.068	1.823	0.087	-1.036	-0.580
Well	(0.014)	(0.029)	(0.064)	(0.035)	(0.088)	(0.224)	(0.028)	(0.133)	(0.129)
English Well	0.141	0.050	0.703	0.306	0.074	3.044	0.390	-1.688	-0.892
	(0.023)	(0.034)	(0.073)	(0.052)	(0.102)	(0.247)	(0.056)	(0.159)	(0.191)
English Not	0.125	-0.006	0.523	0.056	-0.148	1.728	0.475	-2.115	-0.441
Well	(0.024)	(0.041)	(0.082)	(0.073)	(0.128)	(0.270)	(0.070)	(0.191)	(0.190)
English Not	0.117	0.070	0.599	-0.179	-0.155	3.178	0.571	-1.241	0.846
At All	(0.035)	(0.036)	(0.072)	(0.143)	(0.189)	(0.272)	(0.122)	(0.236)	(0.178)
Better	1.022	0.338	2.502	1.384	1.193	9.002	0.127	1.766	2.078
Linkage	(0.010)	(0.019)	(0.061)	(0.015)	(0.040)	(0.122)	(0.008)	(0.115)	(0.125)
Mail or Internet	-0.967	-0.449	-2.703	-0.083	0.708	-3.527	-0.397	-5.923	-2.329
Response	(0.010)	(0.024)	(0.068)	(0.019)	(0.073)	(0.156)	(0.011)	(0.122)	(0.092)
Weighted Obs.	264,700,000	21,910,000	39,950,000	264,700,000	21,910,000	39,950,000	264,700,000	21,910,000	39,950,000
Unweighted Obs.	4,418,000	280,000	558,000	4,418,000	280,000	558,000	4,418,000	280,000	558,000

Source: American Community Survey (ACS) and Census Numident, 2016. Notes: These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. Standard errors are clustered by household. The base categories are reference person for relationship, non-Hispanic white for race/ethnicity, speaks only English at home for English ability, and in-person or phone interview for response mode. We also include educational attainment (less than high school, high school but less than bachelor's degree, bachelor's degree, and graduate degree), log of one plus age and its square, and log of one plus the number of years in the U.S. and its square, but do not report them here.

Table 5. Citizenship Status and Naturalization Year Disagreement Regressions

	ACS Noncitizen-AR Citizen	ACS Citizen-AR Noncitizen	ACS – Numident Natural. Year
Relative	0.028	-0.753	-0.343
	(0.011)	(0.215)	(0.068)
Nonrelative	0.571	-5.461	-0.852
	(0.045)	(0.613)	(0.282)
Non-Hispanic	-0.137	2.744	0.683
African Amer.	(0.013)	(0.546)	(0.128)
Hispanic	0.621	-16.00	1.129
	(0.030)	(0.417)	(0.104)
Other Non-	-0.327	0.755	0.144
Hispanic	(0.034)	(0.376)	(0.093)
Worked in	0.398	1.992	0.631
Last Week	(0.015)	(0.260)	(0.095)
Searched for	0.302	-0.620	0.136
Job	(0.029)	(0.542)	(0.157)
English Very	-0.452	1.983	0.517
Well	(0.031)	(0.373)	(0.096)
English Well	0.114	1.063	0.712
	(0.081)	(0.426)	(0.107)
English Not	1.461	-4.927	0.997
Well	(0.113)	(0.480)	(0.129)
English Not	3.391	-8.282	1.656
At All	(0.260)	(0.592)	(0.210)
Better	0.060	4.586	0.006
Linkage	(0.009)	(0.308)	(0.067)
Mail or Internet	-0.262	3.810	0.365
Response	(0.012)	(0.285)	(0.077)
Weighted Obs.	250,300,000	20,220,000	6,407,000
Unweighted Obs.	4,165,000	254,000	89,000

Source: American Community Survey (ACS) and Census Numident, 2016.

Notes: These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. Standard errors are clustered by household. The base categories are reference person for relationship, non-Hispanic white for race/ethnicity, speaks only English at home for English ability, and in-person or phone interview for response mode. We also include educational attainment (less than high school, high school but less than bachelor's degree, bachelor's degree, and graduate degree), log of one plus age and its square, and log of one plus the number of years in the U.S. and its square, but do not report them here.

The last specification of Table 5 shows that the ACS naturalization-Numident citizenship change gap is larger when reporting for a relative or especially a nonrelative, which could indicate lack of respondent knowledge about others' naturalization years.[40] Lack of English language ability is associated with a smaller gap between the ACS and Numident years, suggesting that misunderstanding the question is not an important explanatory factor. Employed people have smaller gaps, reflecting the incentive to promptly tell SSA about the naturalization to facilitate their employment eligibility verification.

[40] Since very few observations have Numident citizenship change years before the ACS naturalization year, a positive coefficient generally means a smaller gap.

6. Effect of Citizenship Question on Unit Self-Response Rates

To forecast the effect of adding a citizenship question to the 2020 Census, we compare mail response rates in the 2010 Census and the 2010 American Community Survey (ACS) for the same housing units. By comparing the self-response behavior of the same housing unit across two surveys, we control for the household's propensity to self-respond to mandatory Census Bureau household surveys in general.

The Census Bureau randomly selected a sample of households to receive the ACS questionnaire in 2010. The questionnaire included 75 questions and asked individuals to report their citizenship status. These households also received the full-count Census questionnaire in the same year, a list of 10 questions that did not include citizenship. We focus on Census housing units that received both questionnaires by mail. In the 2010 Census, these are the housing units from the initial mailing that did not have the questionnaire returned as Undeliverable as Addressed (UAA) and which were not classified as a vacant or delete (meaning uninhabitable or cannot be found). We define a 2010 Census self-response as a returned questionnaire from the first mailing that is not blank. For the 2010 ACS, a self-response is a mail response, also from the first contact mailing.

The presence of a citizenship question is not the only potential reason why a household may be less inclined to self-respond to the ACS than the Census. Census self-response is bolstered by a media campaign and intensive community advocacy group support, and the ACS questionnaire involves much greater respondent burden (OMB 2008, OMB 2009). To distinguish the citizenship question effect, we compare the actual ACS-Census difference in response rates for households that are likely to be more sensitive to the citizenship question to the ACS-Census difference for households less likely to be sensitive to the question. We assume that any reduction in self-response to the ACS vs. the Census for households unsensitive to the citizenship question is due to factors other than the presence of a citizenship question. We use two ways to divide the sample into sensitive and non-sensitive groups. The first is to define the sensitive group as households where at least one person is an AR noncitizen and has been assigned to this housing unit in Rastogi and O'Hara's (2012) administrative records person-address crosswalk (AR noncitizen households), and the less sensitive group is households where all of the persons assigned to the address are AR citizens (AR all-citizen households).[41] AR citizenship status is established using the 2010 Numident and ITINs, as described in Section 3.2.[42] The choice of noncitizens as the sensitive group is motivated by the results in Section 3.1 that AR noncitizens have much higher item nonresponse rates for the citizenship question, both relative to their nonresponse rates for other demographic questions and compared to other people for citizenship. The use of an independent source for where noncitizens are located avoids the potential problem that households with noncitizens may be less likely to provide PII on household members, preventing linkage to

[41] Here we impose a restriction that all household members have nonmissing AR citizenship for the less sensitive group, but we do not impose that restriction on the sensitive group.

[42] The initial definition of citizenship (treating all persons in the Numident but with missing citizenship as citizens) is used for this first set of groups. In the second set of groups, U.S.-born persons with missing citizenship in the Numident are treated as citizens, while foreign-born persons with missing citizenship in the Numident are treated as missing AR citizenship.

their AR citizenship data. The remaining noncitizen households where AR linkage is done may be relatively more cooperative, potentially biasing the results.

We examine a second set of groups for several reasons. We would like to project the citizenship self-response effect forward in time, since population characteristics associated with this effect may be changing. No administrative records person-place crosswalk is available after 2010, however, so we instead use the ACS household roster to define which people are living in the household.[43] AR noncitizens are probably not the people most sensitive to a citizenship question, since most of them are legal residents. Those lacking an SSN should presumably be even more sensitive to a citizenship question, so the AR noncitizen definition may exclude much of the sensitive population.[44] In our second dichotomy the less sensitive group is "AR & ACS all-citizen households", those households where all persons reported in the ACS to be living in the household at the time of the survey are AR citizens, and all are self-reported as being citizens in the ACS as well. The more sensitive group is "all other households", including those households where some residents are both AR citizens and self-reported citizens but at least one is not; there is a mismatch between the survey report and administrative record response; or citizenship status is not reported in one or both sources. We assume AR & ACS all-citizen households are less sensitive to a citizenship question than all other households, since they have demonstrated a willingness to provide citizenship status answers for all household members, those answers are consistent with administrative records and thus likely truthful responses,[45] and citizens presumably have less to fear about revealing their status than noncitizens. In comparison to others, more of this group's reluctance to self-respond to the ACS should be due to reasons other than the citizenship question, such as unwillingness to answer a longer questionnaire. Note that if some of the reluctance by AR & ACS all-citizens households to self-respond is due to the citizenship question in the ACS, then our analysis will underestimate the citizenship question unit self-response effect.

The sample size for the second set of groups is significantly larger than that for the first set of groups, because the first set excludes households where no persons are AR noncitizens at the address, but at least one person assigned to that address by administrative records cannot be linked to the Numident.

Table 6 displays unweighted 2010 Census and ACS response rates for the AR all-citizen households and AR noncitizen household groups. The self-response rate is higher for the 2010 Census than for the ACS for both citizenship categories, presumably reflecting the higher burden of the ACS. The all-citizen response rate is greater than the noncitizen rate in each survey, suggesting that noncitizen households have a lower participation rate in general. Most important for this study is understanding how the difference in self-response rate across groups varies

[43] Another reason to use the survey household roster rather than the AR crosswalk is that the AR crosswalk often places people in different locations. Rastogi and O'Hara (2012) report that among the 279.2 million persons in the 2010 Census who could be assigned a PIK, 27.2 percent are assigned to an address in the AR crosswalk that differs from their Census address.

[44] This is consistent with Camarota and Capizzano (2004), who say field representatives reported that illegal immigrants were less likely to respond than other foreign-born persons. Illegal immigrants are ineligible for SSNs.

[45] As shown in Section 4 above, when an administrative record shows that someone is a citizen, the ACS response is nearly always citizen as well, giving us a high degree of confidence that the person truly is a citizen.

between the 2010 Census and ACS. While the self-response rate for citizen households is 13.8 percentage points lower in the ACS than in the 2010 Census, the self-response rate for households with at least one noncitizen is 18.9 percentage points lower for the ACS than the self-response rate to the 2010 Census, which is a 5.1 percentage point difference between the two categories.

Table 6. Comparison of 2010 ACS to 2010 Census Response Rates with Initial Assumptions

	Self-Response Rate (%)		Difference
	2010 ACS	2010 Census	
Households with at least one AR noncitizen	52.6	71.5	-18.9
	(0.21)	(0.19)	(0.26)
AR all-citizen households	66.1	79.9	-13.8
	(0.05)	(0.04)	(0.06)
Difference-in-differences			-5.1
			(0.26)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. All persons in the 2010 Numident that are missing citizenship are treated as citizens here. Robust standard errors are in parentheses, calculated from regressions. The estimates are unweighted. Around 5.9 percent of the households have at least one noncitizen. The sample size is 929,000. DRB clearance number CBDRB-2017-CDAR-001.

Using survey weights can facilitate comparisons of results across years, since sampling can change, and we would like to be able to project results forward in time. We thus display weighted response rates in Table 7, now both for the first and second sets of groups. As expected, the restriction to being a citizen in both the AR and ACS results in higher self-response rates in the AR & ACS all-citizen household group compared to the AR all-citizen household group. The response rates for the two noncitizen groups differ little from each other. The difference-in-differences estimate for the first set of groups increases to 8.9 percentage points compared to the unweighted gap in Table 6. It is three percentage points higher (11.9) across the second set of groups.

Table 7. Comparison of 2010 ACS to 2010 Census Response Rates (Weighted)

	Self-Response Rate (%)		Difference
	2010 ACS	2010 Census	
Households with at least one AR noncitizen	42.4	62.1	-19.7
	(0.32)	(0.18)	(0.26)
AR all-citizen households	62.0	72.8	-10.8
	(0.34)	(0.11)	(0.24)
Difference-in-differences			-8.9
			(0.35)
All other households	42.0	62.7	-20.7
	(0.32)	(0.14)	(0.25)
AR & ACS all-citizen households	65.6	74.4	-8.9
	(0.33)	(0.11)	(0.24)
Difference-in-differences			-11.9
			(0.34)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors for the self-response rates and differences are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5. The difference-in-differences (*DiD*) standard errors (*SE*) are calculated as $DiD\ SE = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$, where the two estimates (*Est*) are the 2010 Census – 2010 ACS differences for the two groups. The estimates use ACS housing unit weights. 88.2 percent of households are in the AR all-citizen household group vs. 11.8 percent in the households with at least one AR noncitizen group. 74.9 percent are in the AR & ACS all-citizen household group vs. 25.1 percent are in the all other households group. The number of observations is 1,418,000.

The larger decline in self-response rates for the AR noncitizen household and all other households groups may not actually be due to greater sensitivity. Other characteristics besides citizenship status could be associated with lower ACS self-response, and the AR noncitizen household and all other households groups could have a higher propensity to have such characteristics. To explore this possibility, we perform Blinder-Oaxaca decompositions (Blinder 1973 and Oaxaca 1973).[46]

Households may belong to one of two groups $G \in (S, U)$, where the *S* group is thought to be potentially sensitive to a citizenship question, while the *U* group is not. We set the self-responses $R_{G_iACS_t}$ and $R_{G_iCensus_t}$ equal to 100 if household *i* in group *G* self-responds in year *t* to the ACS and Census, respectively, and zero otherwise.[47] The difference between the survey responses is

$$\Delta R_{G_it} = R_{G_iACS_t} - R_{G_iCensus_t} \quad (1)$$

The vector of predictors *X* includes household size and reference person characteristics (sex, race/ethnicity, age, educational attainment, household income, working in the last week, job search

[46] This method was initially developed to study the extent to which the gender wage gap is due to different distributions of characteristics associated with wages by gender (explained variation) vs. differing behavior across gender for a given set of characteristics (unexplained variation). The unexplained variation is usually attributed to discrimination, but it also captures any effects of differences in unobserved variables.

[47] We use 100 for response so that the results are expressed in percentages.

in the last four weeks, and English language ability among those speaking a language other than English at home). β contains the slope parameters and intercept, and ε is an error term with mean zero.

We estimate OLS models for each household group

$$\Delta R_{S_{it}} = X'_{S_{it}}\beta_{S_t} + \varepsilon_{S_{it}} \quad (2)$$

$$\Delta R_{U_{it}} = X'_{U_{it}}\beta_{U_t} + \varepsilon_{U_{it}} \quad (3)$$

The difference-in-differences in expected self-response rates across the two surveys for the two groups *S* and *U* in year *t* is

$$\Delta\Delta R_{SU_t} = E\left(\Delta R_{S_t}\right) - E\left(\Delta R_{U_t}\right) \quad (4)$$

We decompose this as follows:

$$\Delta\Delta R_{SU_t} = \left[E\left(X_{S_t}\right) - E\left(X_{U_t}\right)\right]'\beta_{U_t} + \left[E\left(X_{S_t}\right)'\left(\beta_{S_t} - \beta_{U_t}\right)\right] \quad (5)$$

The first term (explained variation) applies the coefficients for the unsensitive group to the difference between the expected value of the sensitive group's predictors and those of the unsensitive group. The second (unexplained variation) is the difference between the expected value of the sensitive group's predictors applied to the sensitive group's coefficients and the same predictors applied to the unsensitive group's coefficients. The interpretation that the unexplained variation represents the citizenship question effect is dependent on the assumption that there are no unobserved variables relevant to the difference-in-differences in self-response across the two surveys.

Table 8 shows the results of the Blinder-Oaxaca decomposition for the two sets of groups. In the AR all-citizen vs. AR noncitizen comparison, virtually all the difference-in-differences is explained by differences in predictors across the two groups. Thus, it appears that the larger fall in self-response to the ACS vs. the Census for AR noncitizen households is not due to sensitivity to the citizenship question, but rather that AR noncitizen households have a greater propensity to have other characteristics that are associated with lower ACS self-response. In contrast, about half (6.1 percentage points) of the difference-in-differences for the AR & ACS all-citizen vs. all other household comparison is unexplained, suggesting that the larger drop-off in ACS self-response for all other households is partly due to sensitivity to the citizenship question. Appendix Table A13 shows the regression coefficients for equations (2) and (3), and the explained variation and unexplained variation coefficients for each predictor are shown in Appendix Table A14.

Table 8. Blinder-Oaxaca Decomposition of Comparison of Predicted 2010 ACS to 2010 Census to Response Rates by Households Citizenship Type

	2010 ACS – 2010 Census
Households with at least	-19.7
one AR noncitizen	(0.13)
AR all-citizen households	-10.8
	(0.12)
Difference-in-differences	-8.9
	(0.09)
Explained	-8.7
	(0.11)
Unexplained	-0.2
	(0.13)
All other households	-20.7
	(0.12)
AR & ACS all-citizen households	-8.9
	(0.12)
Difference-in-differences	-11.9
	(0.07)
Explained	-5.8
	(0.14)
Unexplained	-6.1
	(0.16)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The number of observations is 1,418,000.

To see how changes in predictors over time affect the magnitude of the unexplained variation (*UV*) in the decomposition, we apply the coefficients from the 2010 models to the predictors in the 2016 ACS

$$UV_{2016} = E\left(X_{S_{2016}}\right)' \beta_{S_{2010}} - E\left(X_{S_{2016}}\right)' \beta_{U_{2010}} \quad (6)$$

Table 9 shows that the unexplained variation is still insignificant for the AR all-citizen vs. AR noncitizen comparison. It is of a similar magnitude in 2016 as in 2010 (5.8 percentage points vs. 6.1) for the AR & ACS all-citizen vs. all other household comparison. Note that this does not capture changes over time in the degree of sensitivity to a citizenship question for a housing unit with a fixed set of characteristics. That would require estimating models on fresher data of surveys with and without a citizenship question for the same households.

Table 9. Comparison of Predicted 2016 ACS to 2010 Census Response Rates for AR Noncitizen and All Other Households with Their Own vs. All-Citizen Models

	2016 ACS – 2010 Census
Model\Sample	AR noncitizen household sample
AR noncitizen household model	-19.7
	(0.47)
AR all-citizen household model	-20.5
	(0.34)
Difference-in-differences	0.8
	(0.58)
Model\Sample	All other household sample
All other household model	-21.7
	(0.33)
AR & ACS all-citizen household model	-15.9
	(0.39)
Difference-in-differences	-5.8
	(0.51)

Source: 2016 ACS 1-year file and 2016 Numident.

Notes: 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors for the 2010 Census – 2016 ACS response differences are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5. The difference-in-differences (*DiD*) standard errors (*SE*) are calculated as $DiD\ SE = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$, where the two estimates (*Est*) are the 2010 Census – 2016 ACS differences for the two groups. The estimates use ACS housing unit weights. 28.6 percent are in the all other households group in 2016. The standard errors are in parentheses. They are the standard errors of the model predictions, based on the bootstrapped regressions in Appendix Table A12 that use 80 ACS replicate weights. The number of observations is 163,000 for the AR noncitizen household sample and 477,000 for the all other household sample.

Though suggestive, these exercises and the ones performed below are not perfect laboratories for studying the self-response effect of inclusion of a citizenship question on the 2020 Census. The ACS contains 75 questions, so any one question is unlikely to stand out, whereas an added question will be more visible in the 2020 Census questionnaire, which contains just 10 other questions.[48] Thus, we would ideally want to compare response rates on a short questionnaire without a citizenship question to one adding just the citizenship question. Second, the level of concern about using citizenship data for enforcement purposes may be very different in 2020 than it was in 2000 or 2010, so a more recent test would be preferable. These factors suggest the estimated effect on self-response from the exercise in Table 9 is conservative.

[48] A preferable test would be a randomized control trial (RCT) comparing self-response rates where some households are randomly chosen to have an 11-question Census questionnaire with a citizenship question (the treated group), and a randomly chosen set of control households receive a 10-question Census questionnaire without citizenship.

As robustness checks we do similar exercises below with the 2000 Census and the 2014 Survey of Income and Program Participation (SIPP), a longitudinal survey that follows the same individuals over time. Unlike the decennial census and the ACS, individuals respond for themselves in the SIPP. The 2000 Census long form (sent to one of every six housing units, selected randomly) contained a citizenship question among many other additional questions, while the short form (sent to the remaining housing units) did not. As in the first set of groups above, we divide housing units into those with all citizens and those with at least one noncitizen, based on citizenship data from the 2002 Numident for persons enumerated at those housing units in the 2000 Census.[49] As with the 2010 ACS and Census exercises, Table 10 shows that self-response rates are higher in the short form than the long form, and they are higher in households with all citizens. The short- vs. long-form difference in response rates is greater for households with at least one noncitizen by 3.3 percentage points, again consistent with the possibility that households with noncitizens are more sensitive to the inclusion of citizenship questions.

Table 10. 2000 Census Long Form and Short Form Analysis

	Self-response rate (%)		
Households by Citizen	Long Form	Short Form	Difference
At Least One Noncitizen	62.5	71.0	-8.5
	(0.017)	(0.016)	(0.023)
All Citizens	76.1	81.3	-5.2
	(0.005)	(0.004)	(0.006)
Difference	13.6	10.3	-3.3
	(0.017)	(0.016)	(0.024)

Source: 2000 Census short and long forms.

Notes: These are weighted using housing unit weights. The number of short forms is 105.5 million, and the number of long forms is 16.4 million. The definition of self-response is mail response here. Robust standard errors are in parentheses, generated from weighted regressions of response on an interaction of the household citizenship status with short form. The standard errors for the differences are calculated as $SE(Est_1 - Est_2) = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$.

Longitudinal data provide another means for understanding response sensitivity to questions of citizenship. Using the 2014 Survey of Income and Program Participation (SIPP) longitudinal panel waves 1 and 2, we show how nonresponse changes from Wave 1 to Wave 2 for noncitizen respondents, as well as for households with at least one noncitizen. The first row in Table 11 shows nonresponse rates for noncitizens from the 2014 Survey of Income and Program Participation (SIPP) Waves 1 and 2. Noncitizens made up around 6 percent of the 2014 SIPP survey in Wave 1. The proportion of noncitizens in Wave 2 decreased slightly, implying that noncitizens were more

[49] To be classified as a housing unit with all citizens in this exercise, all persons must be linked to the Numident. A housing unit can be classified as having at least one noncitizen if there is at least one person linked to the Numident who is a Numident noncitizen, whether or not all the other persons in the housing unit could be linked to the Numident or not.

likely to leave the survey due to attrition or other factors than citizens. In addition, the rate of nonresponse among those households with at least one noncitizen increased from Wave 1 to Wave 2, from 7.9 percent to 8.5 percent. While noncitizens were more likely to drop out of the survey, those who stayed were more likely to live in households where at least one member did not respond. These data provide additional hints of the potential future impact to nonresponse for noncitizens in surveys that ask about citizenship status.

Table 11. Noncitizens and Nonresponse in the 2014 Survey of Income and Program Participation

	Wave 1		Wave 2	
	(%)	(se)	(%)	(se)
Noncitizens	6.1	(0.144)	5.7	(0.174)
At least one member in the noncitizen household did not respond	7.9	(0.473)	8.5	(0.537)

Source: 2014 SIPP, Waves 1 and 2

Notes: Citizenship status refers to status in Wave 1. The standard errors are clustered in Wave 2. These estimates are run on the internal run 16 version of the 2014 SIPP.

7. Effects of Citizenship Question on Nonresponse Follow-up Costs and Enumeration Quality

A drop in the self-response rate from adding a citizenship question in Alternatives B (obtaining citizenship from the 2020 Census only) and D (obtaining citizenship from the 2020 Census and administrative records) results in increased costs in the Nonresponse Follow-up (NRFU) operation and affects the quality of the population count. Households deciding not to self-respond because of the citizenship question are likely to refuse to cooperate with enumerators coming to their door in NRFU, resulting in the use of neighbors as proxy respondents on their behalf. [50] As shown in Table 12, Mule (2012) reports that the correct enumeration rate is 27.1 percentage points lower for proxies than mail in self-responses based on data from the 2010 Census Coverage Measurement (CCM) survey. The person linkage rate is 62.9 percentage points lower for proxies than for mail in self-responses in the 2010 Census, according to Rastogi and O'Hara (2012). Both these studies provide suggestive evidence that proxies supply poor quality individual demographic and socioeconomic characteristic information about the person on behalf of whom they are responding.

[50] A proxy response is a response about the household by someone outside the household, such as a neighbor or property manager. The enumerator will seek a proxy response for households that don't mail back their Census questionnaire or give an in-person interview after several attempts.

Table 12. Enumeration Quality in Mailout/Mailback and Nonresponse Follow-up (NRFU) Proxy Responses

	Mailout/Mailback Response	NRFU Proxy
Correct Enumerations	97.3	70.2
Erroneous Enumerations	2.5	6.7
Whole-Person Census Imputations	0.3	23.1
Person Linkage Rate	96.7	33.8

Source: Mule (2012) for correct enumerations, erroneous enumerations, and whole-person Census imputations, and Rastogi and O'Hara (2012) for the person linkage rate.

We provide two sets of estimates, the first based on our initial assumptions (in parentheses), and a second based on revised assumptions. The main changes in the revised assumptions are an expansion of the group of housing units considered potentially sensitive to a citizenship question and the estimated percentage of them who will not respond to a questionnaire due to the presence of a citizenship question (5.8 percent in Table 9 vs. 5.1 percent in Table 6).

Using these estimates as well as the data in Table 12, we can develop cautious estimates of the data quality and cost consequences of adding the citizenship question to the enumeration form. We assume that all-citizen households are unaffected by the change and that an additional 5.8 percent (5.1 percent) of households that possibly have noncitizens go into NRFU because they do not self-respond.[51] We expect 320 million persons in 126 million occupied households in the 2020 Census.[52] Based on a combination of administrative records from the 2016 Numident and ITINs and the 2016 ACS, we estimate that 28.6 percent (9.8 percent) of all households could potentially contain at least one noncitizen. Combining these assumptions implies an additional 2,090,000 households (630,000 households) and 6.5 million persons (1.6 million persons) in NRFU.[53] If the NRFU data for those households have the same quality as the average NRFU data in the 2010 Census, then the result would be 561,000 (139,000) fewer correct enumerations, of which 185,000 (46,000) are additional erroneous enumerations and 376,000 (93,000) are additional whole-person census imputations. This analysis assumes that during the NRFU operations a cooperative member of the household supplies data 79.0 percent of the time, and 21.0 percent receive proxy responses. If all of these new NRFU cases go to proxy responses instead,[54] the result would be 1,750,000

[51] Recall that the initial estimate is based on households with at least one AR noncitizen, which is only a fraction of the housing units in the all other households category, which also includes persons with missing citizenship in AR or the ACS or citizenship values that conflict between AR and the ACS.

[52] We assume 10 million residents of group quarters. Group quarters are not included in either mailout/mailback or NRFU operations, and here we assume no effect of a citizenship question on their enumeration.

[53] The initial assumption here is that average household size for households with at least one noncitizen is the same as the forecast for all households in the 2020 Census (2.54 persons). The revised assumption is that average household size for all other households is the same as its average in the 2016 ACS, 3.1 persons.

[54] If a household declines to self-respond due to the citizenship question, we suspect it would also refuse to cooperate with an enumerator coming to their door, resulting in a need to use a proxy.

(432,000) fewer correct enumerations, of which 272,000 (67,000) are erroneous enumerations, and 1,477,000 (365,000) are whole-person census imputations.[55] The number of persons who are linkable to administrative records would fall by 4.1 million (1 million).

Our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 2,090,000 households (630,000 households), or approximately 1.66 percentage points (0.5 percentage points). We currently estimate that for each percentage point increase in NRFU, the cost of the 2020 Census increases by approximately $55 million. Accordingly, the addition of a question on citizenship could increase the cost of the 2020 Census by at least $91.2 million ($27.5 million). It is worth stressing that this cost estimate is a lower bound. Our estimate of $55 million for each percentage point increase in NRFU is based on an average of three visits per household. We expect that many more of these noncitizen households would receive six NRFU visits.

8. Distribution of 2020 Citizenship Data Sources by Collection Method

Figures 10-12 provide forecasts of how many U.S. residents in the 2020 Census acquire their citizenship data from survey responses, administrative records, and model-based imputation methods in Alternatives B, C, and D. Once again we provide forecasts based on initial and revised assumptions, with initial forecasts in parentheses.[56] A reduction in self-response rates and increase in proxy responses from adding the citizenship question in Alternatives B and D is likely to affect the number of persons with survey responses for citizenship. As shown above, reference persons are much less likely to answer the citizenship question for nonrelatives in the household than for themselves, so they may be even less likely to answer it for neighbors. In order to obtain a range of estimates based on best and worst case scenarios, Figure 10 Panel A and Figure 12 Panels A and B assume that proxies report citizenship at the same rate as they do in the 2010 ACS relative to all persons in the 2010 ACS,[57] while Figure 10 Panels B and C and Figure 12 Panels C and D assume none of the proxies report citizenship.

We begin with the estimated 2020 Population of 330 million, the total number of persons we expect to count in the 2020 Census. Under Alternative B with complete citizenship data from proxy

[55] These enumeration errors may not be avoidable simply by spending more money on fieldwork. Once a household decides not to cooperate, it may not be possible to obtain an accurate enumeration no matter how many times an enumerator knocks on their door.

[56] In addition to the differences between the initial and revised assumptions mentioned in Section 7, two others are relevant here. One is that the initial assumptions classify foreign-born persons with missing citizenship in the Numident and without an ITIN as AR citizens, while the revised assumptions classify them as having missing AR citizenship. A second is that instead of showing the difference in the AR linkage rate with and without a citizenship question in the 2020 Census as an increase in the AR linkage rate in Alternative C, the revised assumptions show it as a decrease in the AR linkage rate in Alternative D.

[57] Within 2010 ACS households that have NRFU proxy responses in the 2010 Census, the nonmissing citizenship rate is 96.7 percent, vs. 97.1 percent for all ACS households. We apply this proxy to total sample ratio to the 93.7 percent nonmissing citizenship rate in the 2016 ACS to get an estimated 92.9 percent nonmissing citizenship rate for proxies in 2020.

responses, 309.1 million citizenship responses are obtained from the Census. Applying the missing citizenship rate of 6.3 percent in the 2016 ACS, we expect 20.9 million to have missing data for the citizenship question, either because the respondent skipped the question, or because a proxy response in nonresponse follow-up (NRFU) did not deliver information on that question. Citizenship is imputed using models for these 20.9 million persons.[58] With no citizenship data from proxy responses, the number of citizenship responses drops to 290 million (294.6 million), with 40 million (35.4 million) modeled.[59] The accuracy of this imputation system is unknown at this time. As discussed above, the imputation will be challenging due to the fact that nonresponse is highly correlated with citizenship.

Under Alternative C, we expect to link 289.6 million (295.0 million) to administrative records containing citizenship data, applying the linkage rate for the 2016 ACS to currently available administrative records.[60] The remaining 40.4 million (35.0 million) will have citizenship imputed using models based on the variables common to the linked and non-linked portions of the data. At this time, the accuracy of that imputation system is not known, but it would be based on the administrative record citizenship variable, so it would not be subject to the biases caused by survey citizenship reporting issues.

Of the 309.1 million who provide valid responses to the Census citizenship question in Alternative D, we expect to link 269.6 million (272.5 million) records to the administrative data.[61] Of these, the vast majority, 260.9 million (263.0 million), will have administrative record and Census responses that agree (applying the 2016 ACS-AR agreement rate of 96.8 percent), and since the agreement is with the same administrative record system as in Alternative C, these people will have the same citizenship status under either alternative. Of the 269.6 million (277.4 million) linked Census responses with a valid answer to the 2020 Census question, we expect the administrative record and the Census response to disagree for 8.7 million (9.7 million). These are the persons for whom we have two choices: (1) accept the Census questionnaire answer or (2)

[58] General imputation models develop a response for those who did not respond using all available relevant data.

[59] Based on the analysis in Table 9, under our revised assumptions we project 6.5 million additional proxy responses due to the citizenship question, of which an estimated 840,000 already have missing citizenship (applying the allocation rate of 13.0 percent from the 2016 ACS among persons who do not both report being citizens and are AR citizens). This is in addition to an estimated 14.5 million proxy responses in 2020 without a citizenship question, of which an estimated 1,030,000 already have missing citizenship (applying the 2016 ACS citizenship item allocation rate of 6.3 percent among all ACS-AR citizenship groups, adjusted by the ratio of the 2010 ACS citizenship allocation rate for 2010 Census proxy respondents (3.3 percent) to the 2010 ACS citizenship allocation rate for the whole 2010 ACS sample (2.9 percent)). Note that the proxy responses that are anticipated to occur in 2020 regardless of presence of a citizenship question may happen in households containing people in any ACS-AR citizenship group, whereas the additional proxies due to the citizenship question are assumed to come from housing units where people are not in the group with both ACS and AR citizen responses.

[60] As discussed in Section 7, our initial estimate of the effect of a citizenship question on the number of linkable persons is 1 million, and the revised estimate is 4.1 million. Our initial estimate adds 1 million to the number of linked persons when no citizenship question is included in the questionnaire. We incorporate the change in the number of linkable persons as a reduction in AR linkage in Alternative D for our revised estimate, as discussed below.

[61] When applying the 2016 ACS linkage to administrative record citizenship rate, the estimate is 273.4 million persons with linked citizenship. Of the 4.1 million anticipated reduction in linkage due to the citizenship question in our revised estimate, about 3.9 million are applied to the group with observed 2020 citizenship, as 93.7 percent of persons are anticipated to have observed 2020 citizenship (applying the missing citizenship rate in the 2016 ACS).

replace the questionnaire answer with the administrative answer. If we do the former, all of these cases will differ from the Alternative C answer. The estimated direct response is U.S. citizen for 7.6 million (7.7 million) of these persons, compared to 1.1 million (2.0 million) in the administrative records. Use of direct responses for those with disagreement would result in a projected 6.5 million (5.7 million) more U.S. citizens than when using administrative records.[62]

Continuing with Alternative D, we would process the 20.9 million responses where we did not get a valid answer to the Census citizenship question as in Alternative C. This would result in 16.0 million (16.6 million) persons for whom we expect to find an answer in the administrative records, and 4.9 million (4.3 million) for whom we would use a modeled answer.[63] The models would be developed using the same methods as in Alternative C, but not the same input data, because of the change in response behavior associated with asking the citizenship question.

When 2020 citizenship is observed in Alternative D, but the record cannot be linked to administrative data, we would accept the survey response for an expected 39.5 million (31.7 million) people. The number of persons whose records can be linked to administrative data is lower by 4.1 million (10.7 million) in Alternative D than in Alternative C due to poorer linkage quality from proxy responses, which would have been self-responses without a citizenship question (see Table 10). This captures the negative effect of inclusion of the citizenship question on the ability to use administrative data for citizenship.

When we assume that none of the proxy responses report citizenship, the number where 2020 citizenship is observed falls to 289.5 million (294.6 million) in Alternative D, just as in Alternative B. 263.4 million (272.5 million) of these are linked to administrative record citizenship, 255.6 million (263.0 million) of those answers agree between sources, and 7.8 million (9.5 million) disagree. The direct response for the latter group is U.S. citizen for 6.8 million (7.5 million) vs. 1.0 million (2.0 million) U.S. citizens in administrative records, leading to a 5.8 million (5.6 million) higher count of U.S. citizens if direct responses are used.

Of the 26.6 million (22.2 million) persons for whom 2020 citizenship is observed, but the record cannot be linked to administrative data, we estimate that about 560,000 (500,000) noncitizens will respond as citizens, based on the AR noncitizens reporting as ACS citizens share of the 2016 ACS (2.3 percent in the initial estimates and 2.1 percent in the revised estimates).

These results show that there is a tendency for persons missing citizenship in one source to also be missing it in the other. Among persons with observed 2020 Census citizenship in Figure 12 Panel D, 90.8 percent have AR citizenship, while only 55.5 percent of those without 2020 Census citizenship have AR citizenship. Of those with AR citizenship, 92.2 percent have 2020 Census citizenship, but just 59.9 percent of those without AR citizenship have 2020 Census citizenship. The correlated missingness reduces the coverage gain from using multiple sources. Only 22.2 million persons' citizenship values can be covered by AR among those without 2020 Census

[62] To put this in context, the 2016 ACS estimates that 22.5 million U.S. residents are noncitizens, or 7.0 percent of the population.

[63] Here we apply the remainder of the anticipated 4.1 million reduction in linkage to administrative record citizenship due to increased proxy response to the group for which 2020 citizenship is not observed.

citizenship, whereas AR coverage would be 34.6 million if the missingness correlation were zero. Analogously, just 26.6 million persons missing AR citizenship have 2020 Census citizenship, vs. 39.0 million if the correlation were zero.

Across the three alternatives, the data for at least 255.6 million (263.0 million) persons would be identical, and it would be identical for at least 276.9 million (284.3 million) between alternatives C and D. If the administrative record response is used when the cases disagree, then the data for alternatives C and D would agree for 285.6 million (294.0 million) linked cases.

Alternative C results in more persons with modeled citizenship responses, while Alternative D has fewer imputations. If no proxy respondents report citizenship, then Alternative B has about the same number of imputations as Alternative C, but otherwise its level is in between that of Alternatives C and D.

As mentioned above, the estimated reduction in self-response due to the inclusion of a citizenship question is based on a comparison of a long 2010 ACS questionnaire to a short 2010 Census questionnaire. The visibility of the citizenship question may be more prominent when added to a short questionnaire, resulting in a larger reduction in self-response than what we have estimated here. If the assumption that all proxy responses result in citizenship item nonresponse is accurate, every additional person without Census citizenship will have to have modeled citizenship in Alternative B. With Alternative D, fewer of the additional nonresponses will be modeled, as some can be linked to administrative record citizenship data. The option to use administrative records in Alternative D thus partially mitigates the citizenship question self-response effect.

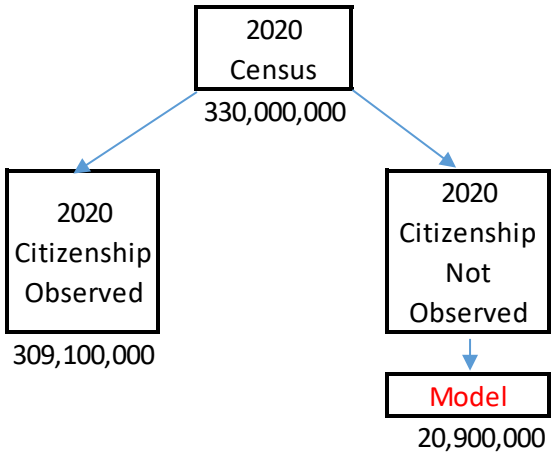
These estimates are based on currently available administrative record citizenship data and linkage capability. The Census Bureau may obtain several additional sources by 2020 and develop better linkage, in which case administrative record coverage may be higher than that shown here. This would lead to fewer imputations in Alternative D and especially Alternative C. The number of imputations in Alternative C is not much higher than in Alternative B, so even a small improvement in administrative record citizenship data coverage would lead to a lower imputation rate in Alternative C than B. Alternative D's advantage in coverage over Alternative C would shrink, though it is unlikely to vanish completely.

A key question when comparing Alternatives C and D is whether the data quality is higher for the 2020 Census or for imputed values for the persons with imputations in Alternative C and observed 2020 Census data in Alternative D. Survey citizenship data exhibit a markedly higher U.S. citizen share compared to administrative records for persons with both sources, but it is unknown whether that tendency also applies to persons without links to administrative records.

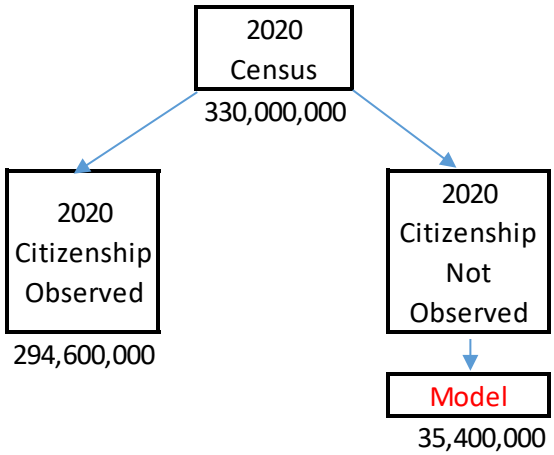
A second question is what data source(s) to use when administrative records and the survey response disagree in Alternative D. Citizenship status is verified via documentation from the issuing government agencies in the administrative records data, but not in the survey, and the analysis in Section 4 above exhibits patterns suggesting that the survey responses are more often inaccurate when they disagree. On the other hand, using administrative records when the sources disagree would mean that the survey response contribution to the citizenship statistics would be minor – it would only be necessary for persons without linked administrative record citizenship

data. The 2020 Census citizenship data is the sole source for 8.1 percent (6.7 percent) of persons in Figure 12 Panel D (Panel C), and this share could be smaller if administrative record coverage improves or survey coverage is lower than estimated. It could be difficult to justify burdening respondents with this question if needed for only a small fraction of the population.

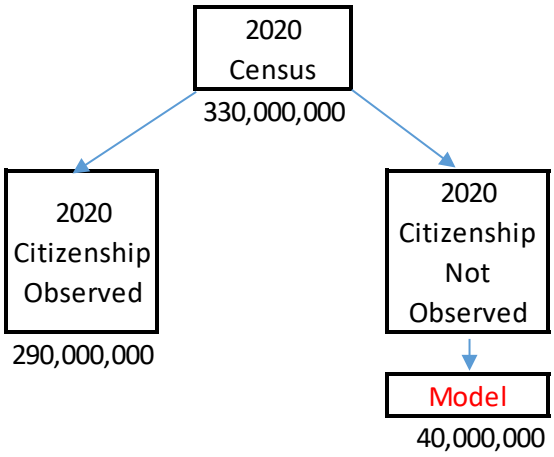
Figure 10. Alternative B



Panel A. Alternative B, Proxies Report Citizenship

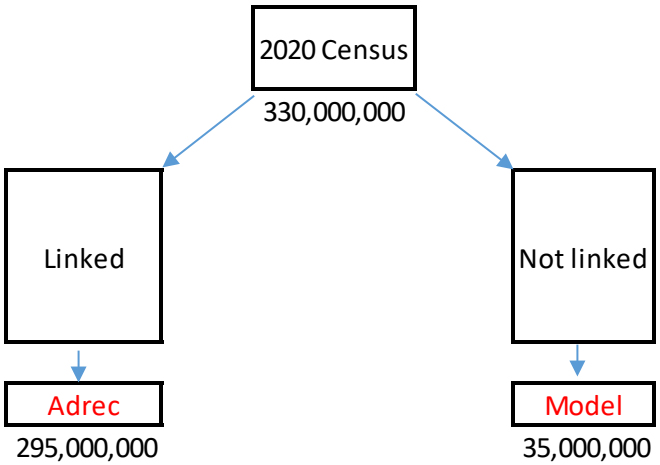


Panel B. Alternative B, Proxies Don’t Report Citizenship, Initial Assumptions

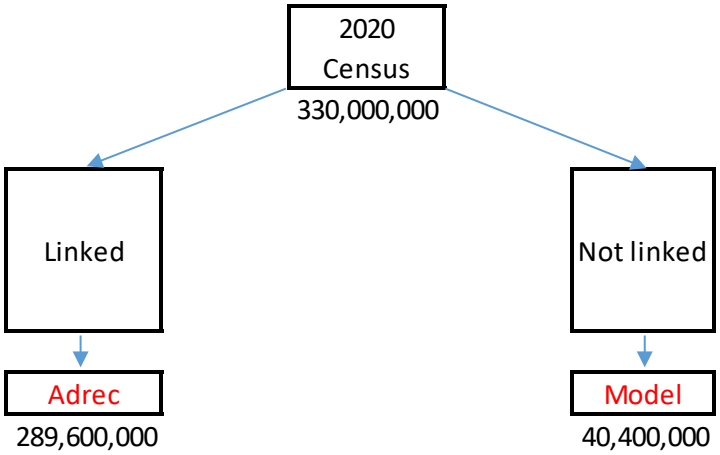


Panel C. Alternative B, Proxies Don’t Report Citizenship, Revised Assumptions

Figure 11. Alternative C

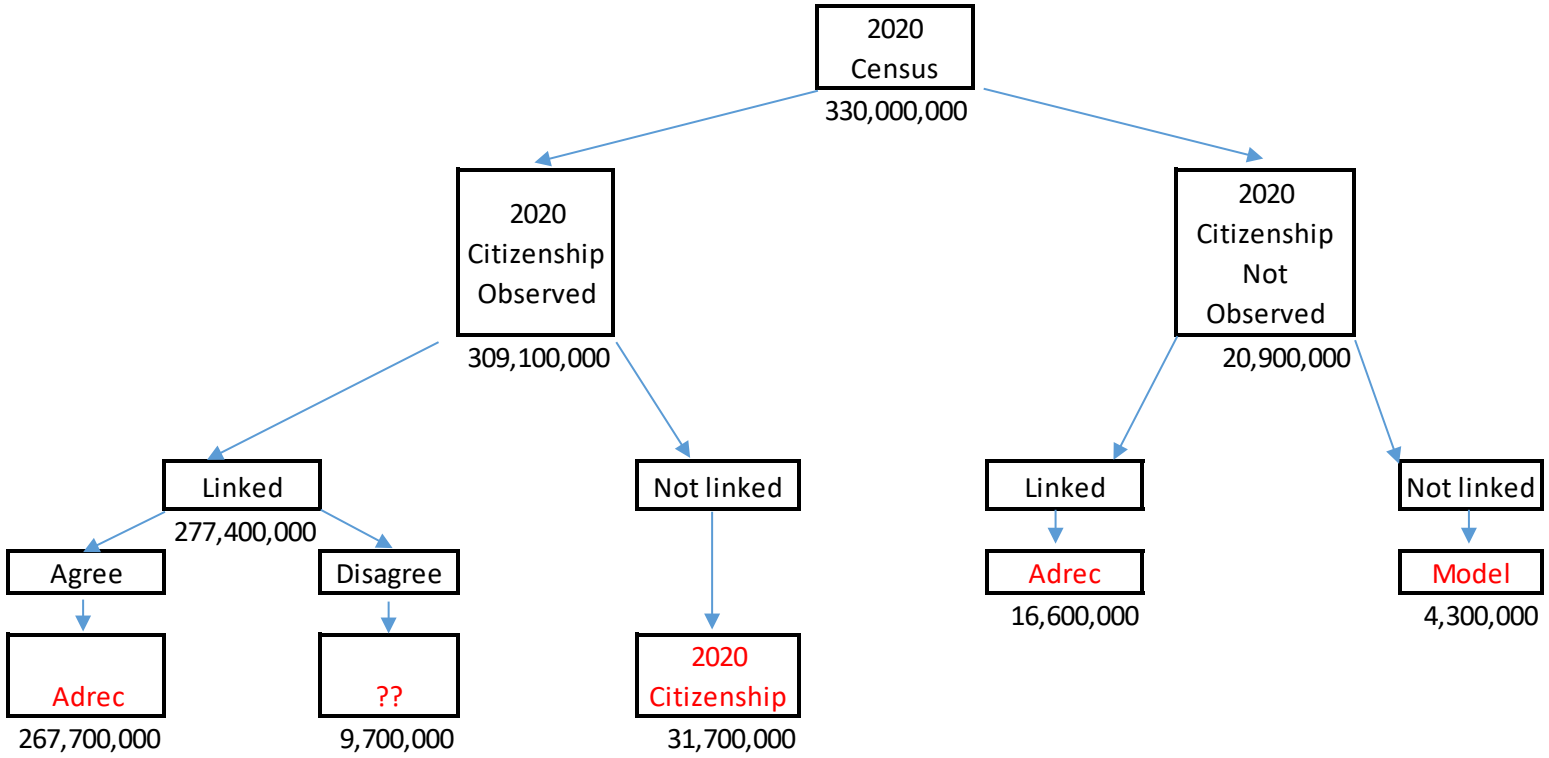


Panel A. Initial Assumptions

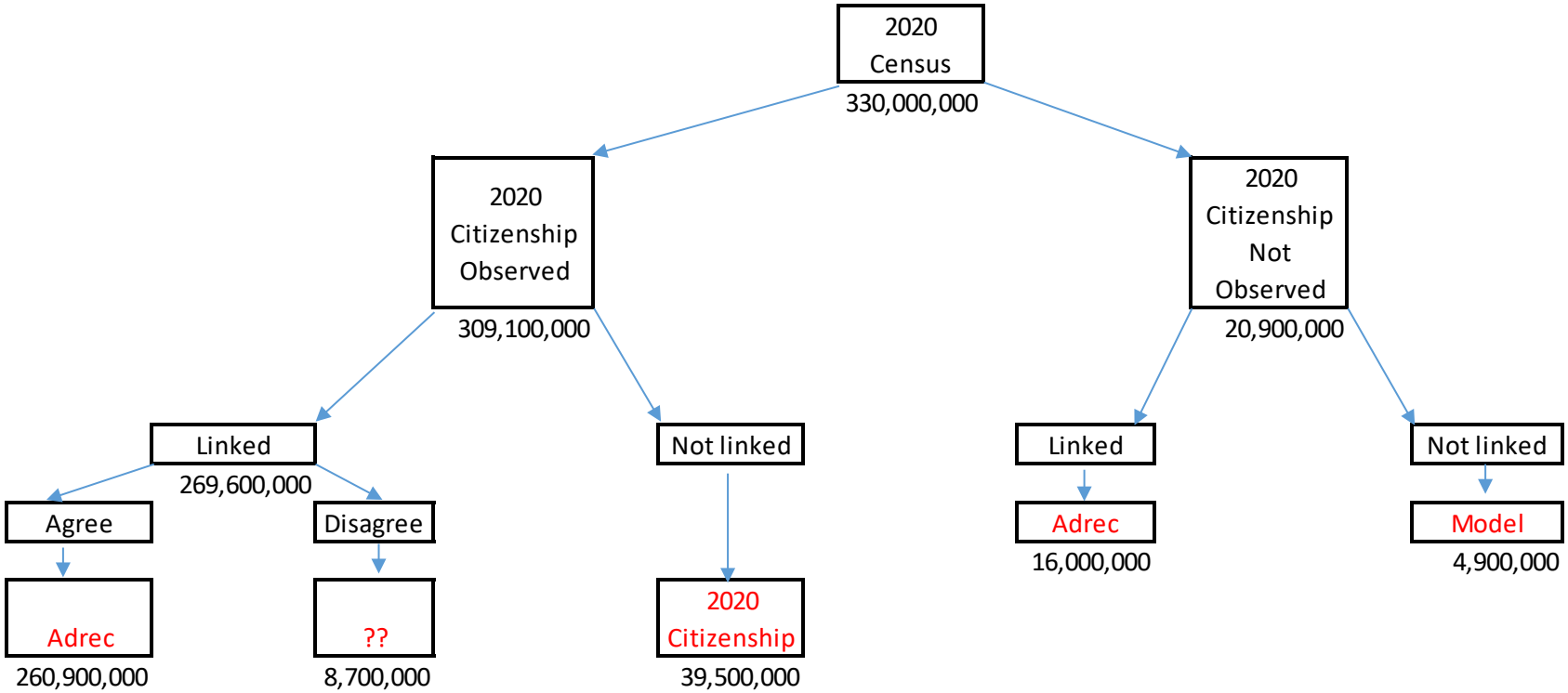


Panel B. Revised Assumptions

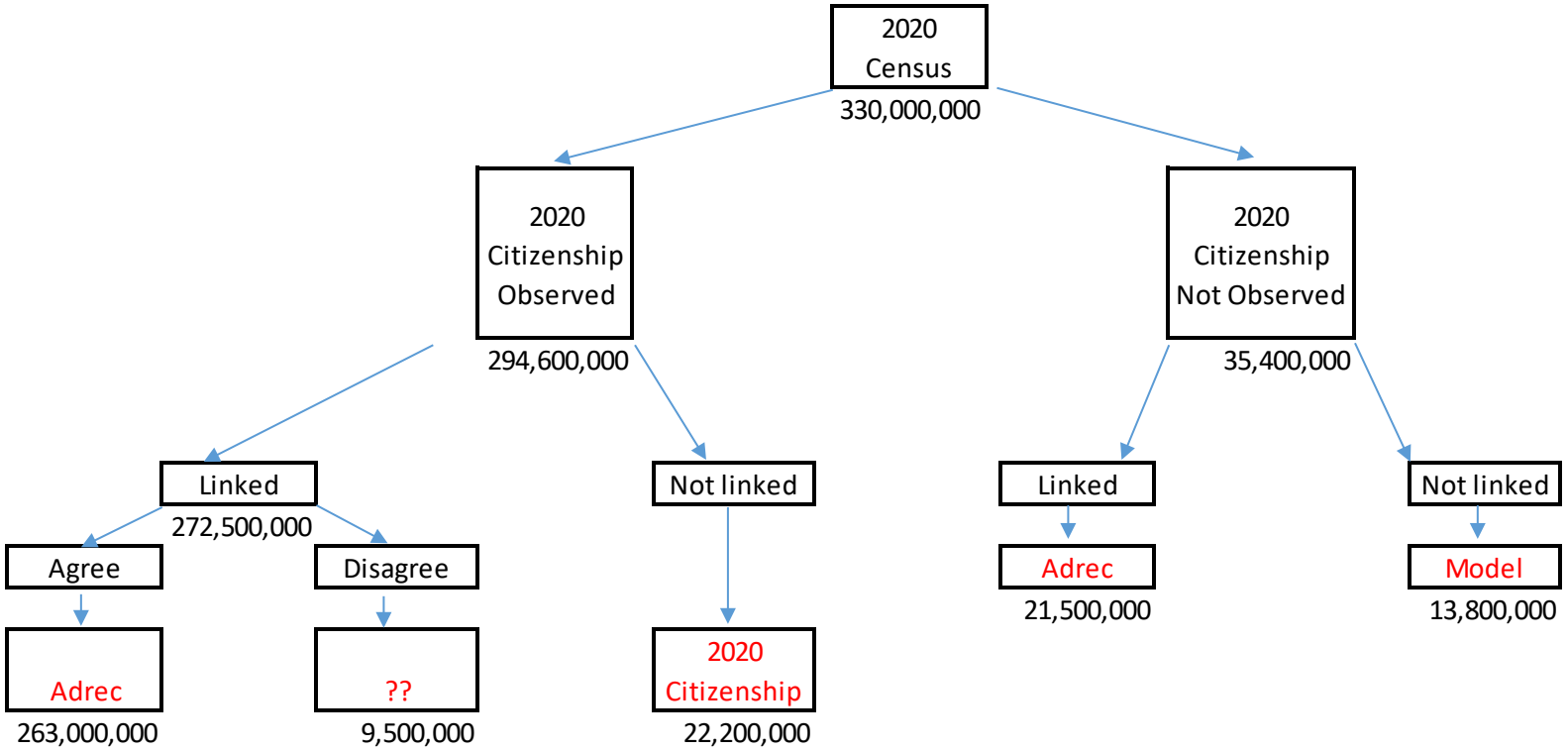
Figure 12. Alternative D



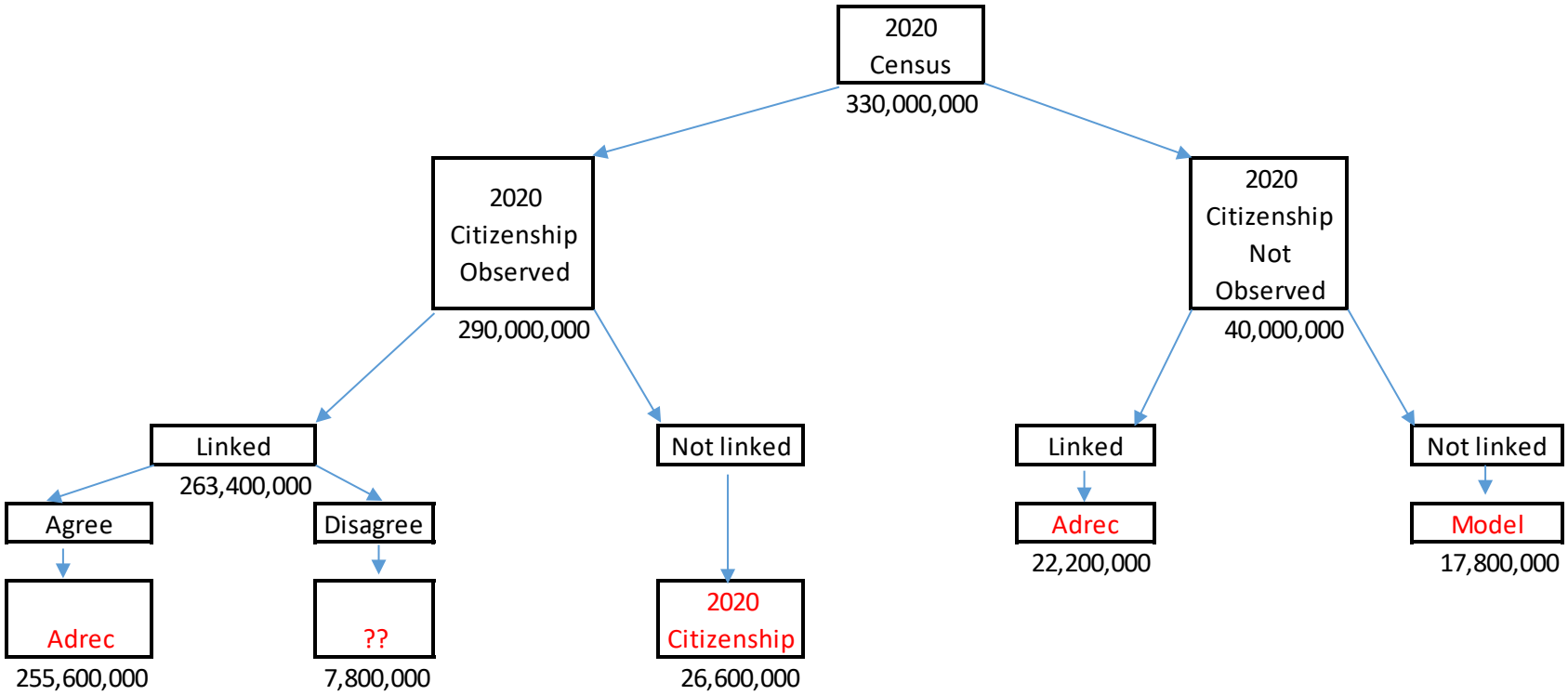
Panel A. Alternative D, Proxies Report Citizenship, Initial Assumptions



Panel B. Alternative D, Proxies Report Citizenship, Revised Assumptions



Panel C. Alternative D, Proxies Don't Report Citizenship, Initial Assumptions



Panel D. Alternative D, Proxies Don't Report Citizenship, Revised Assumptions

9. Conclusion

This paper analyzes general issues of data quality in self-reported citizenship data and examines the coverage and quality of survey-collected and administrative records data available to produce block-level estimates of the Citizen Voting Age Population (CVAP). Our descriptive and regression analyses suggest that many noncitizens misreport their own citizenship on the American Community Survey (ACS), and, in many cases, they do not provide it at all for other noncitizens in the household. The evidence also suggests some naturalized persons either do not notify the Social Security Administration (SSA) about their change in citizenship status or they do so with delay. This potential weakness in SSA data illustrates the desirability of obtaining more timely and complete citizenship data from the U.S. Customs and Immigration Services (USCIS), Customs and Border Protection (CBP), and the State Department. Addressing survey misreporting would be more difficult, however. In the absence of 100 percent complete, accurate, and up to date administrative records, one cannot rule out the possibility that the self-reported citizenship status is correct. Conceptually, it would be challenging to decide which answer to use when sources conflict. Asking respondents to provide proof of citizenship status could reduce misreporting, but this would significantly increase respondent burden and the cost of administering the survey, and it could result in additional unit nonresponse.

This paper's examination of several Census Bureau surveys with and without citizenship questions suggests that households that may contain noncitizens are more sensitive to the inclusion of citizenship in the questionnaire than all-citizen households. The implication is that adding a citizenship question to the 2020 Census would lead to lower self-response rates in households potentially containing noncitizens, resulting in more nonresponse follow-up (NRFU) fieldwork, more proxy responses, and a lower-quality population count.

References

American Community Survey, 2014, "Variance Estimation," in *American Community Survey Design Methodology (January 2014)* version 2.0, Chapter 12, https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch12_2014.pdf, accessed on June 27, 2018.

American Community Survey, 2016a, "Unweighted Housing Unit Sample," Table B98001 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on June 27, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

American Community Survey, 2016b, "Housing Units," Table B25001 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on June 27, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

American Community Survey, 2016c, "Selected Population Profile in the United States," Table S0201 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on July 31, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_1YR_S0201&prodType=table.

American Community Survey, 2018a, "Item Allocation Rates," accessed July 9, 2018, https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/.

American Community Survey, 2018b, "Item Allocation Rates Definitions," accessed July 9, 2018, https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html.

Amuedo-Dorantes, C., and F. Lozano, 2014, "Piecemeal Immigration Enforcement and the New Destinations of Interstate Undocumented Migrants: Evidence from Arizona," mimeo, Pomona College and San Diego State University.

Amuedo-Dorantes, C., and F. Lozano, 2015, "On the Effectiveness of SB1070 in Arizona," *Economic Inquiry*, 53, pp. 335-51. https://doi.org/10.1111/ecin.12138

Baumgardner, Stephanie K., Deborah H. Griffin, and David A. Raglin, 2014, "The Effects of Adding an Internet Response Option to the American Community Survey," 2014 American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-21.

Blinder, Alan S., 1973, "Wage Discrimination: Reduced Form and Structural Estimates," The Journal of Human Resources, 8, pp. 436-455. https://doi.org/10.2307/144855.

Bohn, Sarah, Magnus Lofstrom, and Steven Raphael, 2014, "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *Review of Economics and Statistics*, 96(2), pp. 258-269. https://doi.org/10.1162/rest_a_00429

Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara, 2014, "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey," Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference, https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-08.html (cited on July 3, 2018)

Camarota, Steven, and Jeffrey Capizzano, 2004, "Assessing the Quality of Data Collected on the Foreign-Born: An Evaluation of the American Community Survey," accessed July 31, 2018, http://www.copafs.org/seminars/evaluation_of_american_community_survey.aspx.

Census Bureau, 2002, "Measuring America: The Decennial Censuses from 1790 to 2000," Washington DC, accessed June 27, 2018, https://www.census.gov/history/pdf/measuringamerica.pdf.

Census Bureau, 2013, "Use of Paradata to Assess the Quality and Functionality of the American Community Survey Internet Instrument," 2012 American Community Survey Research and Evaluation Report Memorandum Series #ACS12-RER-26-R1/DSSD 2010 American Community Survey Memorandum Series #ACS12-MP-04-R1, Washington DC, accessed July 5, 2018, https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_01.pdf.

Department of Justice, 2017, "Request to Reinstate Citizenship Question on 2020 Census Questionnaire," letter accessed on June 27, 2018, https://www.documentcloud.org/documents/4340651-Text-of-Dec-2017-DOJ-letter-to-Census.html.

Ellis, M., R. Wright, M. Townley, and K. Copeland, 2014, "The Migration Response to the Legal Arizona Workers Act," *Political Geography*, 42, 46-56. https://doi.org/10.1016/j.polgeo.2014.06.001

Good, M., 2013, "Do Immigrant Outflows Lead to Native Inflows? An Empirical Analysis of the Migratory Responses to US State Immigration Legislation," *Applied Economics*, 45, pp. 4275-97. https://doi.org/10.1080/00036846.2013.786802

Layne, Mary, Deborah Wagner, and Cynthia Rothhaas, 2014, "Estimating Record Linkage False Match Rate for the Person Identification Validation System," CARRA Working Paper #2014-02.

Mule, Thomas, 2012, "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, Washington, DC, accessed on July 6, 2018, https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g01.pdf.

NORC, 2011, "Final Report: Assessment of the U.S. Census Bureau's Person Identification Validation System," University of Chicago: Bethesda, MD.

Oaxaca, Ronald, 1973, "Male-Female Wage Differentials in Urban Labor Markets," International Economic Review, 14, pp. 693-709. https://doi.org/10.2307/2525981.

Office of Management and Budget, 1997, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity," *Federal Register*, 62 FR 58782, Washington DC, accessed on June 28, 2018, https://www.federalregister.gov/documents/1997/10/30/97-28653/revisions-to-the-standards-for-the-classification-of-federal-data-on-race-and-ethnicity.

Office of Management and Budget, 2009, "The American Community Survey," OMB Control No. 0607-0810/ICR Reference No. 200910-0607-005, Washington DC, accessed on July 5, 2018, https://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=200910-0607-005#.

Office of Management and Budget, 2008, "2010 Census," OMB Control No. 0607-0919/ICR Reference No. 200808-0607-003, Washington DC, accessed on July 5, 2018, https://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=200808-0607-003#.

Orrenius, Pia, and Madeline Zavodny, 2016, "Do State Work Eligibility Verification Laws Reduce Unauthorized Immigration?" *IZA Journal of Migration*, 5(5). https://doi.org/10.1186/s40176-016-0053-3

Passel, Jeffrey S., and R.L. Clark, 1997, "How Many Naturalized Citizens Are There: An Assessment of Data Quality in the Decennial Census and the Current Population Survey," Paper presented at the 1997 Annual Meeting of the Population Association of America.

Rastogi, Sonya, and Amy O'Hara, 2012, "2010 Census Match Study," 2010 Census Planning Memoranda Series No. 247.

Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, Census Bureau: Washington DC, accessed July 6, 2018, https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf.

Van Hook, Jennifer, and James D. Bachmeier, 2013, "How Well Does the American Community Survey Count Naturalized Citizens?" *Demographic Research*, 29(1), pp. 1-32. https://doi.org/10.4054/DemRes.2013.29.1

Weinberg, Daniel H., 2011, "Management Challenges of the 2010 U.S. Census," Census Bureau Working Paper: Washington, DC, accessed on June 27, 2018, https://www.census.gov/history/pdf/weinberg-managing2010.pdf.

Appendix I: Tables & Figures

Table A1. Linkage Rates to the 2010 Census by Household Survey and Year, 2000 to 2015

	ACS			SIPP			CPS			AHS		
	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)
2000							33,500	27,500	0.8114			
2001	1,301,000	1,097,000	0.8432	46,500	38,500	0.8328	44,500	37,500	0.8381			
2002	1,110,000	948,000	0.8539				71,000	61,000	0.8564			
2003	1,225,000	1,060,000	0.8652				68,500	59,000	0.8635			
2004	1,222,000	1,072,000	0.8770	86,000	76,000	0.8847	62,000	54,000	0.8716	6,700	5,700	0.8397
2005	4,068,000	3,609,000	0.8870				61,000	53,500	0.8783			
2006	4,259,000	3,792,000	0.8904				78,000	69,000	0.8853			
2007	4,159,000	3,754,000	0.9026				77,500	69,000	0.8909			
2008	4,123,000	3,774,000	0.9155	94,500	84,500	0.8927	75,500	68,000	0.9021			
2009	4,068,000	3,759,000	0.9241				77,500	70,500	0.9059			
2010	4,206,000	3,868,000	0.9195				77,000	70,500	0.9115			
2011	4,571,000	4,205,000	0.9200				77,000	70,000	0.9068	298,000	270,000	0.9047
2012	5,096,000	4,615,000	0.9055				75,500	67,000	0.8869			
2013	4,801,000	4,275,000	0.8903				75,000	65,500	0.8747	133,500	115,500	0.8683
2014	5,014,000	4,385,000	0.8745	66,000	57,500	0.8717						
2015	4,966,000	4,273,000	0.8605							173,500	146,500	0.8455

Source: 2010 Census and Master Demographics, U.S. Census Bureau.
Notes: Household survey data unweighted. All counts have been rounded.

Table A2. Citizenship in Household Surveys Linked to the 2010 Census by Demographics

	Household Surveys Linked to 2010 Census							2010 Census	
	Noncitizen		Citizen		Missing		Total		
	N	(%)	N	(%)	N	(%)	(%)	N	(%)
Total Population	1,523,000		43,090,000		1,192,000		**100.0**	308,745,538	**100.0**
								Coverage	***14.4***
Sex									
Female	785,000	1.7	22,380,000	48.9	613,000	1.3	**51.9**	157,000,000	**50.8**
Male	738,000	1.6	20,710,000	45.2	579,000	1.3	**48.1**	151,800,000	**49.2**
Race									
White	729,000	1.6	35,320,000	77.1	837,000	1.8	**80.5**	227,200,000	**73.6**
Black	127,000	0.3	4,157,000	9.1	173,000	0.4	**9.7**	40,400,000	**13.1**
American Indian, Aleut Eskimo	15,000	0.0	562,000	1.2	16,000	0.0	**1.3**	4,007,000	**1.3**
Asian or Pacific Islander	364,000	0.8	1,688,000	3.7	93,000	0.2	**4.7**	16,770,000	**5.4**
Other	287,000	0.6	1,358,000	3.0	74,500	0.2	**3.8**	20,400,000	**6.6**
Ethnicity									
Hispanic/Spanish	675,000	1.5	4,046,000	8.8	198,000	0.4	**10.7**	50,480,000	**16.4**
Non-Hispanic/Spanish	848,000	1.9	39,040,000	85.2	994,000	2.2	**89.3**	258,300,000	**83.7**

Source: 2010 Census and Master Demographics, U.S. Census Bureau.

Notes: The household survey data are unweighted. The reported population total is the official count from the 2010 Census. All other counts have been rounded. CBDRB-2017-CDAR-001.

Table A3. Item Nonresponse Rates for 2000 and 2010 Short Form Person Questionnaires

	Relationship	Sex	Age	Hispanic Origin	Race	Tenure
2000	1.3	1.1	3.7	3.1	2.9	4.1
2010	1.5	1.5	3.5	3.9	3.3	4.5

Source: Rothhaas, Lestina, and Hill (2012) Tables.

Notes: Rothhaas, Lestina, and Hill (2012) state "the INR rate is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item (i.e. the respondent did not provide an answer to the item). For INR, missing values are included in the rates, but inconsistent responses (i.e. incompatible with other responses) are considered non-missing responses."

Table A4. ACS Item Allocation Rates

Item	2010	2013	2016
Overall housing allocation rate occupied and vacant housing units	5.2	5.6	4.9
Overall person allocation rate total population	5.8	8.4	9.5
Vacancy status vacant housing units	2.9	3.5	3.9
Tenure occupied housing units	1.2	1.3	1.2
Units in structure occupied and vacant housing units	1.5	1.5	1.5
Year moved in occupied housing units	3.4	3	3
Month moved in occupied housing units into which households move in the last two years	0.7	0.7	0.7
Year built occupied and vacant housing units	16.2	17.1	18.2
Lot size occupied and vacant single family and mobile homes	4.2	3.9	3.9
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4.4	4.2	4
Business on property occupied and vacant single family and mobile homes	3	2.4	**
Number of rooms occupied and vacant housing units	5.2	5.5	5
Number of bedrooms occupied and vacant housing units	4.3	4.6	5.5
Running water occupied and vacant housing units	2	2.1	2.4
Flush toilet occupied and vacant housing units	2	2.2	**
Bathtub or shower occupied and vacant housing units	2	2.2	2.6
Sink with a faucet occupied and vacant housing units	2	2.2	2.6
Stove or range occupied and vacant housing units	2.5	2.8	3.1
Refrigerator occupied and vacant housing units	2.7	2.9	3.2
Telephone occupied housing units	1.1	1.2	1.5

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Number of vehicles occupied housing units	1.3	1.4	1.2
Heating fuel, occupied housing units	3.3	3.4	3.4
Monthly electricity cost occupied housing units	7.3	8.2	8.1
Monthly gas cost occupied housing units	9.8	9.9	9.6
Yearly water and sewer cost occupied housing units	8.1	8.8	8.5
Yearly other fuel cost occupied housing units	10.6	8.3	7.3
Yearly food stamp recipiency household occupied housing units	1.3	1.7	1.7
Yearly real estate taxes owner-occupied housing units	16.3	18.5	16.7
Yearly property insurance owner-occupied housing units	23.2	25.6	23.9
Mortgage status owner-occupied housing units	2.1	2.5	2.2
Monthly mortgage payment owner-occupied housing units with a mortgage payment	10.7	12.4	10.5
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	(X)	6.9	6.2
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	(X)	7.4	6.8
Second mortgage owner-occupied housing units	3.4	3.7	3.2
Home equity loan owner-occupied housing units	4.2	4.3	3.7
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	17.9	21.7	23.3
Property value owner-occupied housing units and vacant housing units for sale	12.3	12.9	11.6
Yearly mobile home costs Occupied mobile homes and other units	19.9	21.5	21.7
Monthly condominium fee owner-occupied housing units	0.7	0.8	0.8
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	9.3	9.8	10.5

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2	2.1	2.1
Desktop/laptop/notebook computer occupied housing units	**	3.2	1.3
Handheld computer/smart mobile phone occupied housing units	**	3.3	**
Tablet or other portable wireless computer occupied housing units	**	**	1.6
Smartphone occupied housing units	**	**	1.6
Other computer occupied housing units	**	3.7	1.7
Household has internet access occupied housing units	**	4.4	3.3
Dial-up internet service occupied housing units with internet access	**	5.7	3.8
DSL internet service occupied housing units with internet access	**	5.7	**
Cable modem internet service occupied housing units with internet access	**	5.7	**
Fiber-optic internet service occupied housing units with internet access	**	5.7	**
Cellular data plan (formerly mobile broadband) occupied housing units with internet access	**	26.7	7.6
Satellite internet service occupied housing units with internet access	**	5.7	3.8
High speed internet service occupied housing units with internet access	**	**	3.8
Some other internet service occupied housing units with internet access	**	5.7	3.8
Race total population	1.5	1.6	1.5
Hispanic origin total population	1.8	2.1	1.8
Sex total population	0.1	0.1	0.1
Age total population	1.3	1.6	1.7
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	3	4.8	5.3

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Married past 12 months total population 15 years and over, except those never married	4.7	6.6	6.9
Widowed past 12 months total population 15 years and over, except those never married	4.5	7	7.4
Divorced past 12 months total population 15 years and over, except those never married	4.5	7	7.4
Times married Total population 15 years and over, except those never married	5.1	7.8	8.1
Year last married total population 15 years and over, except those never married	11.4	13.3	13.5
Place of birth total population	6.5	8.6	9.1
Citizenship total population	2.7	5.2	6
Year of naturalization total population naturalized citizens	16.6	22.5	22.5
Year of entry total population not born in U.S.	10.3	13.2	14.8
Speaks another language at home total population 5 years and over	3.4	5.9	6.8
Language spoken total population 5 years and over who speak another language at home	5.7	7	8.3
English ability total population 5 years and over who speak another language at home	4	5.9	7.1
School enrollment total population 3 years and over	3.7	6	6.7
Grade level attended total population 3 years and over enrolled	6	8.9	10.2
Educational attainment total population 3 years and over	5.6	8	8.5
Field of degree total population 25 years and over with a bachelor's degree or higher	9.8	12.4	13.5
Mobility status total population 1 year and over	4	6.5	7.2

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Migration state/foreign county total population 1 year and over movers	7.1	11.3	13.2
Migration county total population 1 year and over movers within U.S.	8.3	12.5	14.6
Migration minor civil division total population 1 year and over movers within U.S.	8.4	12.1	14.2
Migration place total population 1 years and over movers within U.S.	8.8	12.9	15
Health insurance through employer/union total population	6.2	9	10.7
Health insurance purchased directly total population	6.9	9.7	11.3
Health insurance through Medicare total population	5.2	8.1	9.5
Health insurance through Medicaid total population	7.9	10.5	12.2
Health insurance through TRICARE total population	8.1	10.8	12.5
Health insurance through VA total population	8.1	10.7	12.3
Health ins. Through Indian Health Service total population	8.5	11.1	12.8
Visual difficulty total population	3.4	6.1	7.1
Hearing difficulty total population	3.2	5.9	6.8
Physical difficulty total population 5 years and over	3.5	6.7	7.5
Difficulty remembering total population 5 years and over	3.5	6.7	7.5
Difficulty dressing total population 5 years and over	3.5	6.7	7.5
Difficulty going out total population 16 years and over	3.4	6.5	7.3
Grandchildren living in home noninstitutionalized population 30 years and over	0.9	1	1.1
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	12	15.7	17.7

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	14.9	16.1	17.2
Fertility status female total population 15-50	3.7	6.7	7.8
Veteran status total population 17 years and over	3.8	6.8	7.3
Periods of military service total population 17 years and over on active duty now or previously	6.3	9.3	9.7
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	3.9	6.6	6.8
Employment status recode noninstitutionalized population 16 years and over	0.7	0.2	0.2
When last worked noninstitutionalized population 16 years and over	5.1	8.1	8.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	6.9	9.7	10.6
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	7.7	10.8	11.9
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	6.3	10.4	11.8
Place of work county noninstitutionalized population 16 years and over at work last week	7	11	12.5
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	2.1	3.3	3.6
Place of work place noninstitutionalized population 16 years and over at work last week	7.6	11.6	13.1
Transportation to work noninstitutionalized population 16 years and over at work last week	5.7	8.8	9.6
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	6.8	9.9	10.9

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	12.8	18.5	20.2
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	9.7	13.3	14.5
Class of worker total population 16 years and over who worked in the last 5 years	7.2	10.7	11.7
Industry total population 16 years and over who worked in the last 5 years	7.8	11.4	12.7
Occupation total population 16 years and over who worked in the last 5 years	8.1	11.8	13.4
Wages/salary income total population 15 years and over	16	19	19.1
Self-employment income total population 15 years and over	5.9	9.3	10.5
Interest, dividends, etc. income total population 15 years and over	8.8	12.6	15.2
Social security or railroad retirement total population 15 years and over	8.9	12.3	14.5
Supplemental security income total population 15 years and over	6.7	10.3	12.7
Public assistance total population 15 years and over	6.8	10.5	13.2
Retirement income total population 15 years and over	7.5	11.1	13.6
Other income total population 15 years and over	7.4	10.8	13.2
Some or all income allocated total population 15 years and over	22.4	25.3	28.4

Source: American Community Survey (ACS) 1-year files in 2010, 2013, and 2016.

Notes: Item allocation includes nonresponses and responses that were edited. See ACS (2018a and 2018b) for more information about ACS item allocation rates. ** Item was not asked in this year. (X) Some instances where no response to this question was required were incorrectly tallied as allocations, overstating the true level of item allocation required. The incorrect rates have been removed.

Table A5. Citizenship Item Allocation Rate by Response Mode, 2013-2016

	Mail-in Response				Internet Self-Response			
	2013		2016		2013		2016	
	(%)	s.e.	(%)	s.e.	(%)	s.e.	(%)	s.e.
NH White	6.1	(0.023)	6.3	(0.024)	6.2	(0.019)	6.2	(0.018)
NH Black	12.3	(0.090)	12.6	(0.100)	12.3	(0.100)	13.1	(0.091)
NH Asian/NHPI	10.3	(0.126)	12.7	(0.151)	9.4	(0.083)	9.6	(0.075)
NH Other	8.4	(0.143)	8.4	(0.154)	10.0	(0.128)	10.2	(0.114)
Hispanic/Latino	11.8	(0.080)	12.3	(0.088)	13.0	(0.078)	15.5	(0.071)

Source: 2013 & 2016 ACS 1-year files.

Note: Item allocation includes nonresponses and responses that were edited.

Table A6. Administrative Record (AR) Coverage of the 2010 Census, Using Initial AR-Census Crosswalk

	Count	Percent of Decennial Population	Percent of Matched Sample
No PIK, not sent to PVS	10,370,000	3.4	
No PIK, failed in PVS	19,200,000	6.2	
PIK, but not in Numident, not ITIN	8,900	0.0	
PIK, but not in Numident, is ITIN	1,567,000	0.5	
Blank Citizenship	57,910,000	18.8	20.9
U.S. Citizen	200,400,000	64.9	72.2
Noncitizen	19,270,000	6.2	6.9
Total	308,745,538	100.00	100.00

Source: 2010 Numident and initial administrative record-2010 Census crosswalk.

Note: This is the crosswalk used by Rastogi and O'Hara (2012).

Table A7. Percent Linked to 2010 Census among 2017 Numident Records with Missing Citizenship

	Foreign-Born	U.S.-Born
Percent Linked to 2010 Census	36.3	74.5
Total	6.8 million	57.0 million

Source: 2010 Census and 2017 Numident

Notes: These are persons in the 2017 Numident with missing citizenship, born after 1919, and with no date of death. Our preliminary analysis reported 6.6 million foreign-born persons, which excluded some relevant records.

Table A8 Panel A: Citizenship Agreement between 2000 Census Long Form and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=42,580,000)		
Census Citizen	98.8	29.9	71.6	93.0
Census Noncitizen	0.9	66.4	23.8	6.2
Census Missing	0.3	3.7	4.6	0.8
Percent by AR Cat.	86.9	5.4	7.7	100.0
		Non-Hispanic White (N=31,690,000)		
Census Citizen	99.4	31.8	92.4	97.9
Census Noncitizen	0.4	65.8	5.9	1.8
Census Missing	0.2	2.4	1.7	0.3
Percent by AR Cat.	93.4	1.8	4.8	100.0
		Non-Hispanic Black (N=4,543,000)		
Census Citizen	99.3	36.3	92.5	96.1
Census Noncitizen	0.4	59.2	5.4	3.3
Census Missing	0.2	4.5	2.1	0.6
Percent by AR Cat.	85.3	4.0	10.8	100.0
		Hispanic (N=4,534,000)		
Census Citizen	94.3	25.6	35.9	69.3
Census Noncitizen	4.7	69.7	55.2	27.3
Census Missing	1.0	4.7	9.0	3.3
Percent by AR Cat.	60.6	19.1	20.3	100.0
		Non-Hispanic Other Race (N=1,821,000)		
Census Citizen	93.4	33.3	53.1	71.2
Census Noncitizen	5.1	63.7	37.0	26.0
Census Missing	1.4	3.0	9.9	2.9
Percent by AR Cat.	59.2	29.3	11.6	100.0
		Reference Person (N=16,450,000)		
Census Citizen	98.7	32.9	76.9	94.0
Census Noncitizen	0.9	63.3	19.3	5.3
Census Missing	0.4	3.9	3.8	0.8
Percent by AR Cat.	89.4	5.5	5.1	100.0
		Relative (N=24,980,000)		
Census Citizen	98.9	28.6	71.7	92.9
Census Noncitizen	0.8	68.0	24.3	6.4
Census Missing	0.2	3.4	4.1	0.7
Percent by AR Cat.	86.4	5.3	8.3	100.0
		Nonrelative (N=1,153,000)		
Census Citizen	97.2	20.4	58.6	80.4
Census Noncitizen	1.9	72.0	31.3	15.6
Census Missing	0.9	7.6	10.1	4.1
Percent by AR Cat.	63.7	7.4	28.9	100.0
		Age 18+ (N=31,260,000)		
Census Citizen	98.5	30.3	67.1	91.8
Census Noncitizen	1.1	65.5	26.8	7.2
Census Missing	0.4	4.1	6.1	1.1
Percent by AR Cat.	86.2	6.5	7.2	100.0

Source: 2000 Census Long Form and 2002 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A8 Panel B: Citizenship Agreement between 2010 ACS and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=4,520,000)		
ACS Citizen	96.9	32.7	74.8	90.2
ACS Noncitizen	0.8	63.2	17.5	6.9
ACS Missing	2.4	4.1	7.7	2.9
Percent by AR Cat.	83.9	7.5	8.5	100.0
		Non-Hispanic White (N=3,152,000)		
ACS Citizen	97.8	42.4	87.4	96.1
ACS Noncitizen	0.2	53.9	4.0	1.5
ACS Missing	2.0	3.7	8.6	2.4
Percent by AR Cat.	92.4	2.0	5.7	100.0
		Non-Hispanic Black (N=434,000)		
ACS Citizen	96.3	40.1	85.0	92.4
ACS Noncitizen	0.5	54.5	6.5	3.8
ACS Missing	3.2	5.4	8.6	3.8
Percent by AR Cat.	86.2	5.2	8.6	100.0
		Hispanic (N=609,000)		
ACS Citizen	93.9	23.8	61.1	72.7
ACS Noncitizen	2.9	72.4	32.9	23.4
ACS Missing	3.2	3.9	6.0	3.9
Percent by AR Cat.	59.8	21.5	18.8	100.0
		Non-Hispanic Other Race (N=326,000)		
ACS Citizen	93.7	41.3	59.0	77.3
ACS Noncitizen	2.8	54.3	31.8	18.5
ACS Missing	3.5	4.4	9.1	4.2
Percent by AR Cat.	65.7	25.4	8.8	100.0
		Reference Person (N=1,770,000)		
ACS Citizen	97.7	37.2	80.8	91.9
ACS Noncitizen	0.7	59.9	14.1	6.2
ACS Missing	1.6	2.9	5.1	1.9
Percent by AR Cat.	85.4	7.7	6.9	100.0
		Relative (N=2,504,000)		
ACS Citizen	96.5	30.6	75.3	89.7
ACS Noncitizen	0.7	64.7	16.4	6.9
ACS Missing	2.8	4.6	8.2	3.4
Percent by AR Cat.	84.0	7.5	8.5	100.0
		Nonrelative (N=102,000)		
ACS Citizen	94.1	20.0	53.0	77.0
ACS Noncitizen	1.9	72.2	34.8	16.6
ACS Missing	4.0	7.8	12.2	6.4
Percent by AR Cat.	66.0	9.4	24.6	100.0
		Age 18+ (N=3,505,000)		
ACS Citizen	97.0	33.1	71.7	89.1
ACS Noncitizen	0.9	62.9	20.9	8.2
ACS Missing	2.1	4.0	7.3	2.7
Percent by AR Cat.	82.6	9.0	8.4	100.0

Source: 2010 American Community Survey (ACS) and 2010 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A8 Panel C: Citizenship Agreement between 2016 ACS and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=5,255,000)		
ACS Citizen	93.8	34.7	70.4	87.3
ACS Noncitizen	0.7	57.6	17.7	6.4
ACS Missing	5.5	7.7	11.9	6.3
Percent by AR Cat.	82.4	6.7	10.9	100.0
		Non-Hispanic White (N=3,579,000)		
ACS Citizen	95.0	44.7	81.8	93.1
ACS Noncitizen	0.2	48.8	4.4	1.4
ACS Missing	4.8	6.6	13.7	5.5
Percent by AR Cat.	91.1	1.9	7.0	100.0
		Non-Hispanic Black (N=495,000)		
ACS Citizen	93.2	42.0	82.5	89.3
ACS Noncitizen	0.4	49.7	5.8	3.6
ACS Missing	6.4	8.4	11.7	7.2
Percent by AR Cat.	82.3	5.1	12.6	100.0
		Hispanic (N=732,000)		
ACS Citizen	90.5	26.6	58.0	73.1
ACS Noncitizen	2.7	65.2	32.3	19.3
ACS Missing	6.8	8.2	9.7	7.7
Percent by AR Cat.	62.0	16.2	21.8	100.0
		Non-Hispanic Other Race (N=449,000)		
ACS Citizen	90.3	39.1	54.2	74.6
ACS Noncitizen	2.4	53.5	32.2	17.3
ACS Missing	7.3	7.4	13.6	8.1
Percent by AR Cat.	65.8	22.0	12.2	100.0
		Reference Person (N=2,037,000)		
ACS Citizen	96.7	39.1	71.6	90.6
ACS Noncitizen	0.7	56.4	20.1	6.1
ACS Missing	2.7	4.5	8.3	3.2
Percent by AR Cat.	85.5	7.3	7.2	100.0
		Relative (N=2,789,000)		
ACS Citizen	92.3	32.5	68.4	86.0
ACS Noncitizen	0.7	58.2	18.6	6.3
ACS Missing	6.9	9.3	13.0	7.7
Percent by AR Cat.	83.5	6.5	10.0	100.0
		Nonrelative (N=135,000)		
ACS Citizen	85.3	21.5	52.4	71.8
ACS Noncitizen	1.5	61.5	23.6	11.8
ACS Missing	13.2	17.0	23.9	16.4
Percent by AR Cat.	65.6	7.1	27.3	100.0
		Age 18+ (N=4,178,000)		
ACS Citizen	94.3	34.8	68.0	86.6
ACS Noncitizen	0.9	57.7	20.3	7.6
ACS Missing	4.8	7.6	11.7	5.8
Percent by AR Cat.	81.0	8.2	10.7	100.0

Source: 2016 American Community Survey (ACS) and 2016 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A9. Citizenship Agreement Rates (%) Between Census Surveys and Administrative Records

	Consistent	Inconsistent	Missing in One or Both Sources	Consistent, Conditional on Nonmissing	Inconsistent, Conditional on Nonmissing
2000 Census	89.4	2.4	8.2	97.4	2.6
2010 ACS	86.1	3.1	10.8	96.5	3.5
2016 ACS	81.1	2.9	15.9	96.5	3.5

Source: 2000 Census long form, 2002 Census Numident, 2010 American Community Survey (ACS), 2010 Census Numident, 2016 ACS, and 2016 Census Numident.

Notes: These are weighted percentages. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth. The original estimate for inconsistent in the 2000 Census was 2.3 percent, for consistent in the 2010 ACS was 86.0 percent, consistent conditional on nonmissing in the 2010 ACS was 96.4 percent, inconsistent conditional on nonmissing in the 2010 ACS was 3.6 percent, and consistent in the 2016 ACS was 81.2 percent.

Table A10. AR Citizen and Noncitizen Percentages of the 2016 ACS by Race/Ethnicity and Relationship to Reference Person

	AR Citizens	AR Noncitizens
All	81.1	6.7
Non-Hispanic White	90.1	1.9
Non-Hispanic Black	81.5	5.1
Hispanic	60.2	16.2
Non-Hispanic Other Race	62.5	22.0
Reference Person	81.1	6.9
Relative	82.1	6.5
Non-Relative	64.8	7.1

Source: 2010 ACS 1-year file

Notes: These are weighted percentages. The omitted category is persons missing AR citizenship.

Table A11. Percentages of the 2016 ACS Sample by Relationship to Reference Person and Record Linkage Quality

	High-Quality Linkage	Low-Quality Linkage
Reference Person	18.1	20.3
Relative	23.9	33.8
Non-Relative	0.6	3.2

Source: 2016 ACS 1-year file

Notes: These results are weighted. This excludes persons missing AR citizenship.

Table A12. 2016 ACS Citizenship Distribution for ITINs

	Percent of All ITINs
U.S. Citizens	11.1
Born Citizens	6.6

Source: 2016 ACS 1-year file

Note: These results are weighted.

Table A13. Comparison of 2010 ACS and 2010 Census Response Rates: Regressions by Household Citizenship Type

	AR all-citizen households	AR noncitizen households	AR & ACS all-citizen households	All other households
Log Household Size	-3.184	-8.237	-0.4762	-7.185
	(0.1476)	(0.5100)	(0.1737)	(0.3067)
Log Household Size Squared	-0.0998	1.565	-1.929	2.944
	(0.0899)	(0.2304)	(0.1159)	(0.1423)
Female	-6.665	-6.687	-6.263	-8.167
	(0.0557)	(0.1809)	(0.0578)	(0.1289)
Non-Hispanic African Amer.	-10.53	-8.422	-11.48	-3.573
	(0.1143)	(0.3952)	(0.1275)	(0.2064)
Hispanic	-7.532	-20.55	-7.145	-14.07
	(0.1585)	(0.2962)	(0.1640)	(0.2123)
Other Non-Hispanic	0.8338	-0.1256	0.4897	4.129
	(0.1809)	(0.2904)	(0.2162)	(0.2158)
Age 25-34	-4.052	-3.101	-4.658	-1.380
	(0.2078)	(0.5207)	(0.2602)	(0.3202)
Age 35-44	-9.122	-4.746	-9.582	-3.653
	(0.2117)	(0.5048)	(0.2704)	(0.3181)
Age 45-54	-11.83	-6.676	-12.26	-5.095
	(0.2418)	(0.5313)	(0.3082)	(0.3113)
Age 55-64	-12.78	-5.792	-13.20	-5.395
	(0.2715)	(0.5466)	(0.3530)	(0.3315)
Age 65+	-13.06	-4.225	-13.76	-3.617
	(0.3121)	(0.6672)	(0.4051)	(0.3857)
High School	0.7658	-1.195	1.641	-1.866
	(0.1055)	(0.2641)	(0.1097)	(0.1828)
Bachelor's Degree	3.864	2.383	5.116	0.1112
	(0.1197)	(0.3549)	(0.1262)	(0.2316)
Graduate Degree	7.685	7.098	8.448	6.310
	(0.1330)	(0.3923)	(0.1387)	(0.2661)
HH Income $1-$25,000	-1.854	-1.525	-2.249	-1.537
	(0.3130)	(0.9868)	(0.3665)	(0.5480)
HH Income $25,001-$50,000	-2.759	-1.995	-3.002	-3.304
	(0.3158)	(0.9549)	(0.3604)	(0.5348)
HH Income $50,001-$75,000	-3.093	-0.6062	-3.454	-2.555
	(0.3164)	(0.9907)	(0.3563)	(0.5494)
HH Income $75,001-$100,000	-3.037	-0.4054	-3.300	-2.435
	(0.3091)	(1.004)	(0.3590)	(0.5505)
HH Income $100,001+	-2.272	1.035	-2.499	-0.9051
	(0.3183)	(1.016)	(0.3672)	(0.5719)

Table A13. Continued

	AR all-citizen households	AR noncitizen households	AR & ACS all-citizen households	All other households
Worked in	2.741	1.204	3.027	-1.012
Last Week	(0.0644)	(0.2549)	(0.0742)	(0.1613)
Searched for	8.495	8.559	8.565	6.753
Job	(0.1282)	(0.3652)	(0.1357)	(0.2629)
Log Number of				
Years in	-11.17	-10.34	-14.09	-1.286
U.S.	(0.5538)	(0.5499)	(1.434)	(0.4572)
Log Number of				
Years in	2.845	1.997	3.315	-0.1304
U.S. Squared	(0.0904)	(0.1044)	(0.2063)	(0.0823)
English Very	0.9990	0.7404	0.7193	5.302
Well	(0.1669)	(0.2508)	(0.1760)	(0.1927)
English Well	3.823	0.4760	6.686	3.160
	(0.3037)	(0.3007)	(0.3449)	(0.2369)
English Not	-4.707	-7.014	-0.5008	-6.007
Well	(0.3595)	(0.3431)	(0.4334)	(0.3088)
English Not	-13.87	-14.00	-15.19	-13.50
At All	(0.6209)	(0.5355)	(1.070)	(0.4750)
Weighted Obs.	85,100,000	11,400,000	72,300,000	24,200,000
Unweighted Obs.	1,280,000	139,000	1,112,000	306,000

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: The 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The base categories are non-Hispanic white for race/ethnicity, less than high school for educational attainment, no income for household income, and speaks only English at home for English ability.

Table A14. Blinder-Oaxaca Decomposition Coefficients for Comparison of ACS 2010 and Census 2010 Self-Response Rates by Household Citizenship

	AR all-citizen vs. AR noncitizen households		AR & ACS all-citizen vs. all other households	
	Explained	Unexplained	Explained	Unexplained
Log Household Size	1.347	5.799	0.1479	6.758
	(0.0624)	(0.5861)	(0.0539)	(0.3311)
Log Household Size	0.0778	-2.687	1.108	-6.613
Squared	(0.0700)	(0.3702)	(0.0664)	(0.2305)
Female	-0.4591	0.0090	-0.2351	0.8355
	(0.0076)	(0.0761)	(0.0048)	(0.0619)
Non-Hispanic White	2.261	-0.6934	1.612	0.4970
	(0.0419)	(0.0480)	(0.0311)	(0.0566)
Non-Hispanic	-0.2343	-0.4469	-0.0307	-0.7966
African Amer.	(0.0042)	(0.0230)	(0.0030)	(0.0217)
Hispanic	1.196	4.487	0.6141	2.391
	(0.0420)	(0.0904)	(0.0280)	(0.0515)
Other Non-	-0.9869	-0.4652	-0.6266	-0.3870
Hispanic	(0.0258)	(0.0396)	(0.0202)	(0.0298)
Age Below 25	0.0184	0.1686	-0.0735	0.2663
	(0.0029)	(0.0160)	(0.0031)	(0.0156)
Age 25-34	-0.3132	0.7466	-0.1863	0.4588
	(0.0091)	(0.0536)	(0.0065)	(0.0402)
Age 35-44	0.0819	0.0026	0.0487	-0.0507
	(0.0107)	(0.0539)	(0.0068)	(0.0396)
Age 45-54	0.1199	-0.1936	0.0323	-0.3278
	(0.0036)	(0.0537)	(0.0017)	(0.0315)
Age 55-64	-0.2849	-0.3316	-0.1938	-0.3167
	(0.0069)	(0.0331)	(0.0059)	(0.0295)
Age 65+	-0.7552	-0.2910	-0.4319	-0.6361
	(0.0229)	(0.0218)	(0.0165)	(0.0354)
Below High School	0.4911	-0.2640	0.4301	-0.5491
	(0.0130)	(0.0609)	(0.0094)	(0.0338)
High School	-0.3682	0.4025	-0.1901	0.4193
	(0.0084)	(0.0735)	(0.0048)	(0.0557)
Bachelor's Degree	0.0169	0.0827	0.0341	0.4077
	(0.0013)	(0.0356)	(0.0018)	(0.0255)
Graduate Degree	-0.0978	-0.0594	-0.0036	-0.0646
	(0.0030)	(0.0320)	(0.0018)	(0.0217)
HH Income = $0	0.0017	0.0190	-0.0191	0.0117
	(0.0005)	(0.0099)	(0.0024)	(0.0096)
HH Income $1	0.0072	0.2648	0.0003	-0.0193
-$25,000	(0.0021)	(0.0567)	(0.0002)	(0.0453)
HH Income $25,001	0.0122	0.2196	0.0042	0.2364
-$50,000	(0.0017)	(0.0619)	(0.0007)	(0.0424)

Table A14. Continued

	AR all-citizen vs. AR noncitizen households		AR & ACS all-citizen vs. all other households	
	Explained	Unexplained	Explained	Unexplained
HH Income	0.0030	-0.1651	-0.0063	-0.0477
$50,000-$75,000	(0.0006)	(0.0470)	(0.0007)	(0.0288)
HH Income	-0.0032	-0.1245	-0.0060	-0.0278
$75,001-$100,000	(0.0005)	(0.0339)	(0.0007)	(0.0221)
HH Income	0.0003	-0.3579	-0.0001	-0.1981
$100,001+	(0.0003)	(0.0581)	(0.0007)	(0.0408)
Worked in	-0.3684	1.134	-0.0440	2.545
Last Week	(0.0089)	(0.1959)	(0.0024)	(0.1203)
Searched for	-0.1195	-0.0050	-0.0399	0.1230
Job	(0.0041)	(0.0286)	(0.0025)	(0.0198)
Log Number of.	-10.19	-2.434	-8.077	-42.62
Years in U.S	(0.5032)	(2.068)	(0.8219)	(4.835)
Log Years in	16.68	7.861	12.22	40.24
U.S. Squared	(0.5292)	(1.126)	(0.7610)	(2.490)
Only English	1.876	-0.2650	0.7321	-0.2700
	(0.1384)	(0.0675)	(0.1293)	(0.1763)
English Very	-0.9976	-0.3151	-0.4244	-1.183
Well	(0.0486)	(0.0873)	(0.0481)	(0.0699)
English Well	-1.270	0.4510	-1.010	0.3912
	(0.0491)	(0.0669)	(0.0360)	(0.0468)
English Not	0.3237	0.1948	-0.1211	0.5456
Well	(0.0444)	(0.0541)	(0.0393)	(0.0414)
English Not	0.6369	-0.0655	0.5055	-0.0864
At All	(0.0272)	(0.0377)	(0.0308)	(0.0358)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: The 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The base categories are non-Hispanic white for race/ethnicity, less than high school for educational attainment, no income for household income, and speaks only English at home for English ability. The number of observations is 1,418,000 (unweighted) and 96,540,000 (weighted).

Federal Register

Friday,

February 22, 2002

Part IX

Office of Management and Budget

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication

OFFICE OF MANAGEMENT AND BUDGET

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Republication

Editorial Note: Due to numerous errors, this document is being reprinted in its entirety. It was originally printed in the **Federal Register** on Thursday, January 3, 2002 at 67 FR 369–378 and was corrected on Tuesday, February 5, 2002 at 67 FR 5365.

AGENCY: Office of Management and Budget, Executive Office of the President.

ACTION: Final guidelines.

SUMMARY: These final guidelines implement section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106–554; H.R. 5658). Section 515 directs the Office of Management and Budget (OMB) to issue government-wide guidelines that ''provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies.'' By October 1, 2002, agencies must issue their own implementing guidelines that include ''administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency'' that does not comply with the OMB guidelines. These final guidelines also reflect the changes OMB made to the guidelines issued September 28, 2001, as a result of receiving additional comment on the ''capable of being substantially reproduced'' standard (paragraphs V.3.B, V.9, and V.10), which OMB previously issued on September 28, 2001, on an interim final basis.

DATES: *Effective Date:* January 3, 2002.

FOR FURTHER INFORMATION CONTACT: Brooke J. Dickson, Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20503. Telephone (202) 395–3785 or by e-mail to informationquality@omb.eop.gov.

SUPPLEMENTARY INFORMATION: In section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106–554; H.R. 5658), Congress directed the Office of Management and Budget (OMB) to issue, by September 30, 2001, government-wide guidelines that ''provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies * * *'' Section 515(b) goes on to state that the OMB guidelines shall:

''(1) apply to the sharing by Federal agencies of, and access to, information disseminated by Federal agencies; and

''(2) require that each Federal agency to which the guidelines apply—

''(A) issue guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by the agency, by not later than 1 year after the date of issuance of the guidelines under subsection (a);

''(B) establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency that does not comply with the guidelines issued under subsection (a); and

''(C) report periodically to the Director—

''(i) the number and nature of complaints received by the agency regarding the accuracy of information disseminated by the agency and;

''(ii) how such complaints were handled by the agency.''

Proposed guidelines were published in the **Federal Register** on June 28, 2001 (66 FR 34489). Final guidelines were published in the **Federal Register** on September 28, 2001 (66 FR 49718). The Supplementary Information to the final guidelines published in September 2001 provides background, the underlying principles OMB followed in issuing the final guidelines, and statements of intent concerning detailed provisions in the final guidelines.

In the final guidelilnes published in September 2001, OMB also requested additional comment on the ''capable of being substantially reproduced'' standard and the related definition of ''influential scientific or statistical information'' (paragraphs V.3.B, V.9, and V.10), which were issued on an interim final basis. The final guidelines published today discuss the public comments OMB received, the OMB response, and amendments to the final guidelines published in September 2001.

In developing agency-specific guidelines, agencies should refer both to the Supplementary Information to the final guidelines published in the **Federal Register** on September 28, 2001 (66 FR 49718), and also to the Supplementary Information published today. We stress that the three ''Underlying Principles'' that OMB followed in drafting the guidelines that we published on September 28, 2001 (66 FR 49719), are also applicable to the amended guidelines that we publish today.

In accordance with section 515, OMB has designed the guidelines to help agencies ensure and maximize the quality, utility, objectivity and integrity of the information that they disseminate (meaning to share with, or give access to, the public). It is crucial that information Federal agencies disseminate meets these guidelines. In this respect, the fact that the Internet enables agencies to communicate information quickly and easily to a wide audience not only offers great benefits to society, but also increases the potential harm that can result from the dissemination of information that does not meet basic information quality guidelines. Recognizing the wide variety of information Federal agencies disseminate and the wide variety of dissemination practices that agencies have, OMB developed the guidelines with several principles in mind.

First, OMB designed the guidelines to apply to a wide variety of government information dissemination activities that may range in importance and scope. OMB also designed the guidelines to be generic enough to fit all media, be they printed, electronic, or in other form. OMB sought to avoid the problems that would be inherent in developing detailed, prescriptive, ''one-size-fits-all'' government-wide guidelines that would artificially require different types of dissemination activities to be treated in the same manner. Through this flexibility, each agency will be able to incorporate the requirements of these OMB guidelines into the agency's own information resource management and administrative practices.

Second, OMB designed the guidelines so that agencies will meet basic information quality standards. Given the administrative mechanisms required by section 515 as well as the standards set forth in the Paperwork Reduction Act, it is clear that agencies should not disseminate substantive information that does not meet a basic level of quality. We recognize that some government information may need to meet higher or more specific information quality standards than those that would apply to other types of government information. The more important the information, the higher the quality standards to which it should be held, for example, in those situations involving ''influential scientific, financial, or statistical information'' (a phrase defined in these guidelines). The guidelines recognize, however, that

information quality comes at a cost. Accordingly, the agencies should weigh the costs (for example, including costs attributable to agency processing effort, respondent burden, maintenance of needed privacy, and assurances of suitable confidentiality) and the benefits of higher information quality in the development of information, and the level of quality to which the information disseminated will be held.

Third, OMB designed the guidelines so that agencies can apply them in a common-sense and workable manner. It is important that these guidelines do not impose unnecessary administrative burdens that would inhibit agencies from continuing to take advantage of the Internet and other technologies to disseminate information that can be of great benefit and value to the public. In this regard, OMB encourages agencies to incorporate the standards and procedures required by these guidelines into their existing information resources management and administrative practices rather than create new and potentially duplicative or contradictory processes. The primary example of this is that the guidelines recognize that, in accordance with OMB Circular A–130, agencies already have in place well-established information quality standards and administrative mechanisms that allow persons to seek and obtain correction of information that is maintained and disseminated by the agency. Under the OMB guidelines, agencies need only ensure that their own guidelines are consistent with these OMB guidelines, and then ensure that their administrative mechanisms satisfy the standards and procedural requirements in the new agency guidelines. Similarly, agencies may rely on their implementation of the Federal Government's computer security laws (formerly, the Computer Security Act, and now the computer security provisions of the Paperwork Reduction Act) to establish appropriate security safeguards for ensuring the ''integrity'' of the information that the agencies disseminate.

In addition, in response to concerns expressed by some of the agencies, we want to emphasize that OMB recognizes that Federal agencies provide a wide variety of data and information. Accordingly, OMB understands that the guidelines discussed below cannot be implemented in the same way by each agency. In some cases, for example, the data disseminated by an agency are not collected by that agency; rather, the information the agency must provide in a timely manner is compiled from a variety of sources that are constantly updated and revised and may be confidential. In such cases, while agencies' implementation of the guidelines may differ, the essence of the guidelines will apply. That is, these agencies must make their methods transparent by providing documentation, ensure quality by reviewing the underlying methods used in developing the data and consulting (as appropriate) with experts and users, and keep users informed about corrections and revisions.

Summary of OMB Guidelines

These guidelines apply to Federal agencies subject to the Paperwork Reduction Act (44 U.S.C. chapter 35). Agencies are directed to develop information resources management procedures for reviewing and substantiating (by documentation or other means selected by the agency) the quality (including the objectivity, utility, and integrity) of information before it is disseminated. In addition, agencies are to establish administrative mechanisms allowing affected persons to seek and obtain, where appropriate, correction of information disseminated by the agency that does not comply with the OMB or agency guidelines. Consistent with the underlying principles described above, these guidelines stress the importance of having agencies apply these standards and develop their administrative mechanisms so they can be implemented in a common sense and workable manner. Moreover, agencies must apply these standards flexibly, and in a manner appropriate to the nature and timeliness of the information to be disseminated, and incorporate them into existing agency information resources management and administrative practices.

Section 515 denotes four substantive terms regarding information disseminated by Federal agencies: quality, utility, objectivity, and integrity. It is not always clear how each substantive term relates—or how the four terms in aggregate relate—to the widely divergent types of information that agencies disseminate. The guidelines provide definitions that attempt to establish a clear meaning so that both the agency and the public can readily judge whether a particular type of information to be disseminated does or does not meet these attributes.

In the guidelines, OMB defines ''quality'' as the encompassing term, of which ''utility,'' ''objectivity,'' and ''integrity'' are the constituents. ''Utility'' refers to the usefulness of the information to the intended users. ''Objectivity'' focuses on whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner, and as a matter of substance, is accurate, reliable, and unbiased. ''Integrity'' refers to security—the protection of information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification. OMB modeled the definitions of ''information,'' ''government information,'' ''information dissemination product,'' and ''dissemination'' on the longstanding definitions of those terms in OMB Circular A–130, but tailored them to fit into the context of these guidelines.

In addition, Section 515 imposes two reporting requirements on the agencies. The first report, to be promulgated no later than October 1, 2002, must provide the agency's information quality guidelines that describe administrative mechanisms allowing affected persons to seek and obtain, where appropriate, correction of disseminated information that does not comply with the OMB and agency guidelines. The second report is an annual fiscal year report to OMB (to be first submitted on January 1, 2004) providing information (both quantitative and qualitative, where appropriate) on the number, nature, and resolution of complaints received by the agency regarding its perceived or confirmed failure to comply with these OMB and agency guidelines.

Public Comments and OMB Response

Applicability of Guidelines. Some comments raised concerns about the applicability of these guidelines, particularly in the context of scientific research conducted by Federally employed scientists or Federal grantees who publish and communicate their research findings in the same manner as their academic colleagues. OMB believes that information generated and disseminated in these contexts is not covered by these guidelines unless the agency represents the information as, or uses the information in support of, an official position of the agency.

As a general matter, these guidelines apply to ''information'' that is ''disseminated'' by agencies subject to the Paperwork Reduction Act (44 U.S.C. 3502(1)). See paragraphs II, V.5 and V.8. The definitions of ''information'' and ''dissemination'' establish the scope of the applicability of these guidelines. ''Information'' means ''any communication or representation of knowledge such as facts or data * * *'' This definition of information in paragraph V.5 does ''not include opinions, where the agency's presentation makes it clear that what is

being offered is someone's opinion rather than fact or the agency's views.''

''Dissemination'' is defined to mean ''agency initiated or sponsored distribution of information to the public.'' As used in paragraph V.8, ''agency INITIATED * * * distribution of information to the public'' refers to information that the agency disseminates, e.g., a risk assessment prepared by the agency to inform the agency's formulation of possible regulatory or other action. In addition, if an agency, as an institution, disseminates information prepared by an outside party in a manner that reasonably suggests that the agency agrees with the information, this appearance of having the information represent agency views makes agency dissemination of the information subject to these guidelines. By contrast, an agency does not ''initiate'' the dissemination of information when a Federally employed scientist or Federal grantee or contractor publishes and communicates his or her research findings in the same manner as his or her academic colleagues, even if the Federal agency retains ownership or other intellectual property rights because the Federal government paid for the research. To avoid confusion regarding whether the agency agrees with the information (and is therefore disseminating it through the employee or grantee), the researcher should include an appropriate disclaimer in the publication or speech to the effect that the ''views are mine, and do not necessarily reflect the view'' of the agency.

Similarly, as used in paragraph V.8., ''agency * * * SPONSORED distribution of information to the public'' refers to situations where an agency has directed a third-party to disseminate information, or where the agency has the authority to review and approve the information before release. Therefore, for example, if an agency through a procurement contract or a grant provides for a person to conduct research, and then the agency directs the person to disseminate the results (or the agency reviews and approves the results before they may be disseminated), then the agency has ''sponsored'' the dissemination of this information. By contrast, if the agency simply provides funding to support research, and it the researcher (not the agency) who decides whether to disseminate the results and—if the results are to be released—who determines the content and presentation of the dissemination, then the agency has not ''sponsored'' the dissemination even though it has funded the research and even if the Federal agency retains ownership or other intellectual property rights because the Federal government paid for the research. To avoid confusion regarding whether the agency is sponsoring the dissemination, the researcher should include an appropriate disclaimer in the publication or speech to the effect that the ''views are mine, and do not necessarily reflect the view'' of the agency. On the other hand, subsequent agency dissemination of such information requires that the information adhere to the agency's information quality guidelines. In sum, these guidelines govern an agency's dissemination of information, but generally do not govern a third-party's dissemination of information (the exception being where the agency is essentially using the third-party to disseminate information on the agency's behalf). Agencies, particularly those that fund scientific research, are encouraged to clarify the applicability of these guidelines to the various types of information they and their employees and grantees disseminate.

Paragraph V.8 also states that the definition of ''dissemination'' does not include ''* * * distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes.'' The exemption from the definition of ''dissemination'' for ''adjudicative processes'' is intended to exclude, from the scope of these guidelines, the findings and determinations that an agency makes in the course of adjudications involving specific parties. There are well-established procedural safeguards and rights to address the quality of adjudicatory decisions and to provide persons with an opportunity to contest decisions. These guidelines do not impose any additional requirements on agencies during adjudicative proceedings and do not provide parties to such adjudicative proceedings any additional rights of challenge or appeal.

*The Presumption Favoring Peer-Reviewed Information.*As a general matter, in the scientific and research context, we regard technical information that has been subjected to formal, independent, external peer review as presumptively objective. As the guidelines state in paragraph V.3.b.i: ''If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity.'' An example of a formal, independent, external peer review is the review process used by scientific journals.

Most comments approved of the prominent role that peer review plays in the OMB guidelines. Some comments contended that peer review was not accepted as a universal standard that incorporates an established, practiced, and sufficient level of objectivity. Other comments stated that the guidelines would be better clarified by making peer review one of several factors that an agency should consider in assessing the objectivity (and quality in general) of original research. In addition, several comments noted that peer review does not establish whether analytic results are capable of being substantially reproduced. In light of the comments, the final guidelines in new paragraph V.3.b.i qualify the presumption in favor of peer-reviewed information as follows: ''However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance.''

We believe that transparency is important for peer review, and these guidelines set minimum standards for the transparency of agency-sponsored peer review. As we state in new paragraph V.3.b.i: ''If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity. However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance. If agency-sponsored peer review is employed to help satisfy the objectivity standard, the review process employed shall meet the general criteria for competent and credible peer review recommended by OMB–OIRA to the President's Management Council (9/20/01) (http://www.whitehouse.gov/omb/inforeg/oira_review-process.html), namely, 'that (a) peer reviewers be selected primarily on the basis of necessary technical expertise, (b) peer reviewers be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers be expected to disclose to agencies their sources of personal and institutional funding (private or public sector), and (d) peer reviews be conducted in an open and rigorous manner.' ''

The importance of these general criteria for competent and credible peer review has been supported by a number of expert bodies. For example, ''the work of fully competent peer-review panels can be undermined by allegations of conflict of interest and bias. Therefore, the best interests of the Board are served by effective policies and procedures regarding potential conflicts of interest, impartiality, and panel balance.'' (*EPA's Science Advisory*

Board Panels: Improved Policies and Procedures Needed to Ensure Independence and Balance, GAO–01–536, General Accounting Office, Washington, DC, June 2001, page 19.) As another example, ''risk analyses should be peer-reviewed and accessible—both physically and intellectually—so that decision-makers at all levels will be able to respond critically to risk characterizations. The intensity of the peer reviews should be commensurate with the significance of the risk or its management implications.'' (*Setting Priorities, Getting Results: A New Direction for EPA,* Summary Report, National Academy of Public Administration, Washington, DC, April 1995, page 23.)

These criteria for peer reviewers are generally consistent with the practices now followed by the National Research Council of the National Academy of Sciences. In considering these criteria for peer reviewers, we note that there are many types of peer reviews and that agency guidelines concerning the use of peer review should tailor the rigor of peer review to the importance of the information involved. More generally, agencies should define their peer-review standards in appropriate ways, given the nature and importance of the information they disseminate.

Is Journal Peer Review Always Sufficient? Some comments argued that journal peer review should be adequate to demonstrate quality, even for influential information that can be expected to have major effects on public policy. OMB believes that this position overstates the effectiveness of journal peer review as a quality-control mechanism.

Although journal peer review is clearly valuable, there are cases where flawed science has been published in respected journals. For example, the NIH Office of Research Integrity recently reported the following case regarding environmental health research:

> ''Based on the report of an investigation conducted by [XX] University, dated July 16, 1999, and additional analysis conducted by ORI in its oversight review, the US Public Health Service found that Dr. [X] engaged in scientific misconduct. Dr. [X] committed scientific misconduct by intentionally falsifying the research results published in the journal SCIENCE and by providing falsified and fabricated materials to investigating officials at [XX] University in response to a request for original data to support the research results and conclusions report in the SCIENCE paper. In addition, PHS finds that there is no original data or other corroborating evidence to support the research results and conclusions reported in the SCIENCE paper as a whole.'' (66 FR 52137, October 12, 2001).

Although such cases of falsification are presumably rare, there is a significant scholarly literature documenting quality problems with articles published in peer-reviewed research. ''In a [peer-reviewed] meta-analysis that surprised many—and some doubt—researchers found little evidence that peer review actually improves the quality of research papers.'' (*See, e.g., Science,* Vol. 293, page 2187 (September 21, 2001.)) In part for this reason, many agencies have already adopted peer review and science advisory practices that go beyond journal peer review. *See, e.g.,* Sheila Jasanoff, *The Fifth Branch: Science Advisers as Policy Makers,* Cambridge, MA, Harvard University Press, 1990; Mark R. Powell, *Science at EPA: Information in the Regulatory Process.* Resources for the Future, Washington, DC., 1999, pages 138–139; 151–153; *Implementation of the Environmental Protection Agency's Peer Review Program: An SAB Evaluation of Three Reviews,* EPA–SAB–RSAC–01–009, A Review of the Research Strategies Advisory Committee (RSAC) of the EPA Science Advisory Board (SAB), Washington, DC., September 26, 2001. For information likely to have an important public policy or private sector impact, OMB believes that additional quality checks beyond peer review are appropriate.

Definition of ''Influential''. OMB guidelines apply stricter quality standards to the dissemination of information that is considered ''influential.'' Comments noted that the breadth of the definition of ''influential'' in interim final paragraph V.9 requires much speculation on the part of agencies.

We believe that this criticism has merit and have therefore narrowed the definition. In this narrower definition, ''influential'', when used in the phrase ''influential scientific, financial, or statistical information'', is amended to mean that ''the 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 intent of the new phrase ''clear and substantial'' is to reduce the need for speculation on the part of agencies. We added the present tense—''or does have''—to this narrower definition because on occasion, an information dissemination may occur simultaneously with a particular policy change. In response to a public comment, we added an explicit reference to ''financial'' information as consistent with our original intent.

Given the differences in the many Federal agencies covered by these guidelines, and the differences in the nature of the information they disseminate, we also believe it will be helpful if agencies elaborate on this definition of ''influential'' in the context of their missions and duties, with due consideration of the nature of the information they disseminate. As we state in amended paragraph V.9, ''Each agency is authorized to define 'influential' in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.''

Reproducibility. As we state in new paragraph V.3.b.ii: ''If an agency is responsible for disseminating influential scientific, financial, or statistical information, agency guidelines shall include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties.'' OMB believes that a reproducibility standard is practical and appropriate for information that is considered ''influential'', as defined in paragraph V.9—that ''will have or does have a clear and substantial impact on important public policies or important private sector decisions.'' The reproducibility standard applicable to influential scientific, financial, or statistical information is intended to ensure that information disseminated by agencies is sufficiently transparent in terms of data and methods of analysis that it would be feasible for a replication to be conducted. The fact that the use of original and supporting data and analytic results have been deemed ''defensible'' by peer-review procedures does not necessarily imply that the results are transparent and replicable.

Reproducibility of Original and Supporting Data. Several of the comments objected to the exclusion of original and supporting data from the reproducibility requirements. Comments instead suggested that OMB should apply the reproducibility standard to original data, and that OMB should provide flexibility to the agencies in determining what constitutes ''original and supporting'' data. OMB agrees and asks that agencies consider, in developing their own guidelines, which categories of original and supporting data should be subject to the reproducibility standard and which should not. To help in resolving this issue, we also ask agencies to consult directly with relevant scientific and technical communities on the feasibility of having the selected categories of original and supporting data subject to the reproducibility standard. Agencies are encouraged to address ethical, feasibility, and confidentiality issues

with care. As we state in new paragraph V.3.b.ii.A, ''Agencies may identify, in consultation with the relevant scientific and technical communities, those particular types of data that can practicably be subjected to a reproducibility requirement, given ethical, feasibility, or confidentiality constraints.'' Further, as we state in our expanded definition of ''reproducibility'' in paragraph V.10, ''If agencies apply the reproducibility test to specific types of original or supporting data, the associated guidelines shall provide relevant definitions of reproducibility (*e.g.,* standards for replication of laboratory data).'' OMB urges caution in the treatment of original and supporting data because it may often be impractical or even impermissible or unethical to apply the reproducibility standard to such data. For example, it may not be ethical to repeat a ''negative'' (ineffective) clinical (therapeutic) experiment and it may not be feasible to replicate the radiation exposures studied after the Chernobyl accident. When agencies submit their draft agency guidelines for OMB review, agencies should include a description of the extent to which the reproducibility standard is applicable and reflect consultations with relevant scientific and technical communities that were used in developing guidelines related to applicability of the reproducibility standard to original and supporting data.

It is also important to emphasize that the reproducibility standard does not apply to all original and supporting data disseminated by agencies. As we state in new paragraph V.3.b.ii.A, ''With regard to original and supporting data related [to influential scientific, financial, or statistical information], agency guidelines shall not require that all disseminated data be subjected to a reproducibility requirement.'' In addition, we encourage agencies to address how greater transparency can be achieved regarding original and supporting data. As we also state in new paragraph V.3.b.ii.A, ''It is understood that reproducibility of data is an indication of transparency about research design and methods and thus a replication exercise (*i.e.,* a new experiment, test, or sample) shall not be required prior to each dissemination.'' Agency guidelines need to achieve a high degree of transparency about data even when reproducibility is not required.

Reproducibility of Analytic Results. Many public comments were critical of the reproducibility standard and expressed concern that agencies would be required to reproduce each analytical result before it is disseminated. While several comments commended OMB for establishing an appropriate balance in the ''capable of being substantially reproduced'' standard, others considered this standard to be inherently subjective. There were also comments that suggested the standard would cause more burden for agencies.

It is not OMB's intent that each agency must reproduce each analytic result before it is disseminated. The purpose of the reproducibility standard is to cultivate a consistent agency commitment to transparency about how analytic results are generated: the specific data used, the various assumptions employed, the specific analytic methods applied, and the statistical procedures employed. If sufficient transparency is achieved on each of these matters, then an analytic result should meet the ''capable of being substantially reproduced'' standard.

While there is much variation in types of analytic results, OMB believes that reproducibility is a practical standard to apply to most types of analytic results. As we state in new paragraph V.3.b.ii.B, ''With regard to analytic results related [to influential scientific, financial, or statistical information], agency guidelines shall generally require sufficient transparency about data and methods that an independent reanalysis could be undertaken by a qualified member of the public. These transparency standards apply to agency analysis of data from a single study as well as to analyses that combine information from multiple studies.'' We elaborate upon this principle in our expanded definition of ''reproducibility'' in paragraph V.10: ''With respect to analytic results, 'capable of being substantially reproduced' means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error.''

Even in a situation where the original and supporting data are protected by confidentiality concerns, or the analytic computer models or other research methods may be kept confidential to protect intellectual property, it may still be feasible to have the analytic results subject to the reproducibility standard. For example, a qualified party, operating under the same confidentiality protections as the original analysts, may be asked to use the same data, computer model or statistical methods to replicate the analytic results reported in the original study. *See, e.g.,* ''Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality,'' A Special Report of the Health Effects Institute's Particle Epidemiology Reanalysis Project, Cambridge, MA, 2000.

The primary benefit of public transparency is not necessarily that errors in analytic results will be detected, although error correction is clearly valuable. The more important benefit of transparency is that the public will be able to assess how much an agency's analytic result hinges on the specific analytic choices made by the agency. Concreteness about analytic choices allows, for example, the implications of alternative technical choices to be readily assessed. This type of sensitivity analysis is widely regarded as an essential feature of high-quality analysis, yet sensitivity analysis cannot be undertaken by outside parties unless a high degree of transparency is achieved. The OMB guidelines do not compel such sensitivity analysis as a necessary dimension of quality, but the transparency achieved by reproducibility will allow the public to undertake sensitivity studies of interest.

We acknowledge that confidentiality concerns will sometimes preclude public access as an approach to reproducibility. In response to public comment, we have clarified that such concerns do include interests in ''intellectual property.'' To ensure that the OMB guidelines have sufficient flexibility with regard to analytic transparency, OMB has, in new paragraph V.3.b.ii.B.i, provided agencies an alternative approach for classes or types of analytic results that cannot practically be subject to the reproducibility standard. ''[In those situations involving influential scientific, financial, or statistical information * * *] making the data and methods publicly available will assist in determining whether analytic results are reproducible. However, the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections.'' Specifically, in cases where reproducibility will not occur due to other compelling interests, we expect agencies (1) to perform robustness checks appropriate to the importance of the information involved, e.g., determining whether a specific statistic is sensitive to the choice of analytic method, and, accompanying the information disseminated, to document their efforts to assure the needed robustness in information quality, and (2) address in their guidelines the

degree to which they anticipate the opportunity for reproducibility to be limited by the confidentiality of underlying data. As we state in new paragraph V.3.b.ii.B.*ii,* ''In situations where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken. Agency guidelines shall, however, in all cases, require a disclosure of the specific data sources that have been used and the specific quantitative methods and assumptions that have been employed.''

Given the differences in the many Federal agencies covered by these guidelines, and the differences in robustness checks and the level of detail for documentation thereof that might be appropriate for different agencies, we also believe it will be helpful if agencies elaborate on these matters in the context of their missions and duties, with due consideration of the nature of the information they disseminate. As we state in new paragraph V.3.b.ii.B.ii, ''Each agency is authorized to define the type of robustness checks, and the level of detail for documentation thereof, in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.''

We leave the determination of the appropriate degree of rigor to the discretion of agencies and the relevant scientific and technical communities that work with the agencies. We do, however, establish a general standard for the appropriate degree of rigor in our expanded definition of ''reproducibility'' in paragraph V.10: '' 'Reproducibility' means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased).'' OMB will review each agency's treatment of this issue when reviewing the agency guidelines as a whole.

Comments also expressed concerns regarding interim final paragraph V.3.B.iii, ''making the data and models publicly available will assist in determining whether analytic results are capable of being substantially reproduced,'' and whether it could be interpreted to constitute public dissemination of these materials, rendering moot the reproducibility test. (For the equivalent provision, see new paragraph V.3.b.ii.B.*i.*) The OMB guidelines do not require agencies to reproduce each disseminated analytic result by independent reanalysis. Thus, public dissemination of data and models *per se* does not mean that the analytic result has been reproduced. It means only that the result should be CAPABLE of being reproduced. The transparency associated with this capability of reproduction is what the OMB guidelines are designed to achieve.

We also want to build on a general observation that we made in our final guidelines published in September 2001. In those guidelines we stated: ''... in those situations involving influential scientific[, financial,] or statistical information, the substantial reproducibility standard is added as a quality standard above and beyond some peer review quality standards'' (66 FR 49722 (September 28, 2001)). A hypothetical example may serve to illustrate this point. Assume that two Federal agencies initiated or sponsored the dissemination of five scientific studies after October 1, 2002 (see paragraph III.4) that were, before dissemination, subjected to formal, independent, external peer review, i.e., that met the presumptive standard for ''objectivity'' under paragraph V.3.b.i. Further assume, at the time of dissemination, that neither agency reasonably expected that the dissemination of any of these studies would have ''a clear and substantial impact'' on important public policies, *i.e.,* that these studies were not considered ''influential'' under paragraph V.9, and thus not subject to the reproducibility standards in paragraphs V.3.b.ii.A or B. Then assume, two years later, in 2005, that one of the agencies decides to issue an important and far-reaching regulation based clearly and substantially on the agency's evaluation of the analytic results set forth in these five studies and that such agency reliance on these five studies as published in the agency's notice of proposed rulemaking would constitute dissemination of these five studies. These guidelines would require the rulemaking agency, prior to publishing the notice of proposed rulemaking, to evaluate these five studies to determine if the analytic results stated therein would meet the ''capable of being substantially reproduced'' standards in paragraph V.3.b.ii.B and, if necessary, related standards governing original and supporting data in paragraph V.3.b.ii.A. If the agency were to decide that any of the five studies would not meet the reproducibility standard, the agency may still rely on them but only if they satisfy the transparency standard and—as applicable—the disclosure of robustness checks required by these guidelines. Otherwise, the agency should not disseminate any of the studies that did not meet the applicable standards in the guidelines at the time it publishes the notice of proposed rulemaking.

Some comments suggested that OMB consider replacing the reproducibility standard with a standard concerning ''confirmation'' of results for influential scientific and statistical information. Although we encourage agencies to consider ''confirmation'' as a relevant standard—at least in some cases—for assessing the objectivity of original and supporting data, we believe that ''confirmation'' is too stringent a standard to apply to analytic results. Often the regulatory impact analysis prepared by an agency for a major rule, for example, will be the only formal analysis of an important subject. It would be unlikely that the results of the regulatory impact analysis had already been confirmed by other analyses. The ''capable of being substantially reproduced'' standard is less stringent than a ''confirmation'' standard because it simply requires that an agency's analysis be sufficiently transparent that another qualified party could replicate it through reanalysis.

Health, Safety, and Environmental Information. We note, in the scientific context, that in 1996 the Congress, for health decisions under the Safe Drinking Water Act, adopted a basic standard of quality for the use of science in agency decisionmaking. Under 42 U.S.C. 300g–1(b)(3)(A), an agency is directed, ''to the degree that an Agency action is based on science,'' to use ''(i) the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices; and (ii) data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data).''

We further note that in the 1996 amendments to the Safe Drinking Water Act, Congress adopted a basic quality standard for the dissemination of public information about risks of adverse health effects. Under 42 U.S.C. 300g–1(b)(3)(B), the agency is directed, ''to ensure that the presentation of information [risk] effects is comprehensive, informative, and understandable.'' The agency is further directed, ''in a document made available to the public in support of a regulation [to] specify, to the extent practicable—(i) each population addressed by any estimate [of applicable risk effects]; (ii) the expected risk or central estimate of

risk for the specific populations [affected]; (iii) each appropriate upper-bound or lower-bound estimate of risk; (iv) each significant uncertainty identified in the process of the assessment of [risk] effects and the studies that would assist in resolving the uncertainty; and (v) peer-reviewed studies known to the [agency] that support, are directly relevant to, or fail to support any estimate of [risk] effects and the methodology used to reconcile inconsistencies in the scientific data.''

As suggested in several comments, we have included these congressional standards directly in new paragraph V.3.b.ii.C, and made them applicable to the information disseminated by all the agencies subject to these guidelines: ''With regard to analysis of risks to human health, safety and the environment maintained or disseminated by the agencies, agencies shall either adopt or adapt the quality principles applied by Congress to risk information used and disseminated pursuant to the Safe Drinking Water Act Amendments of 1996 (42 U.S.C. 300g–1(b)(3)(A) & (B)).'' The word ''adapt'' is intended to provide agencies flexibility in applying these principles to various types of risk assessment.

Comments also argued that the continued flow of vital information from agencies responsible for disseminating health and medical information to medical providers, patients, and the public may be disrupted due to these peer review and reproducibility standards. OMB responded by adding to new paragraph V.3.b.ii.C: ''Agencies responsible for dissemination of vital health and medical information shall interpret the reproducibility and peer-review standards in a manner appropriate to assuring the timely flow of vital information from agencies to medical providers, patients, health agencies, and the public. Information quality standards may be waived temporarily by agencies under urgent situations (*e.g.,* imminent threats to public health or homeland security) in accordance with the latitude specified in agency-specific guidelines.''

Administrative Correction Mechanisms. In addition to commenting on the substantive standards in these guidelines, many of the comments noted that the OMB guidelines on the administrative correction of information do not specify a time period in which the agency investigation and response must be made. OMB has added the following new paragraph III.3.i to direct agencies to specify appropriate time periods in which the investigation and response need to be made. ''Agencies shall specify appropriate time periods for agency decisions on whether and how to correct the information, and agencies shall notify the affected persons of the corrections made.''

Several comments stated that the OMB guidelines needed to direct agencies to consider incorporating an administrative appeal process into their administrative mechanisms for the correction of information. OMB agreed, and added the following new paragraph III.3.ii: ''If the person who requested the correction does not agree with the agency's decision (including the corrective action, if any), the person may file for reconsideration within the agency. The agency shall establish an administrative appeal process to review the agency's initial decision, and specify appropriate time limits in which to resolve such requests for reconsideration.'' Recognizing that many agencies already have a process in place to respond to public concerns, it is not necessarily OMB's intent to require these agencies to establish a new or different process. Rather, our intent is to ensure that agency guidelines specify an objective administrative appeal process that, upon furthercomplaint by the affected person, reviews an agency's decision to disagree with the correction request. An objective process will ensure that the office that originally disseminates the information does not have responsibility for both the initial response and resolution of a disagreement. In addition, the agency guidelines should specify that if the agency believes other agencies may have an interest in the resolution of any administrative appeal, the agency should consult with those other agencies about their possible interest.

Overall, OMB does not envision administrative mechanisms that would burden agencies with frivolous claims. Instead, the correction process should serve to address the genuine and valid needs of the agency and its constituents without disrupting agency processes. Agencies, in making their determination of whether or not to correct information, may reject claims made in bad faith or without justification, and are required to undertake only the degree of correction that they conclude is appropriate for the nature and timeliness of the information involved, and explain such practices in their annual fiscal year reports to OMB.

OMB's issuance of these final guidelines is the beginning of an evolutionary process that will include draft agency guidelines, public comment, final agency guidelines, development of experience with OMB and agency guidelines, and continued refinement of both OMB and agency guidelines. Just as OMB requested public comment before issuing these final guidelines, OMB will refine these guidelines as experience develops and further public comment is obtained.

Dated: December 21, 2001.

John D. Graham,

Administrator, Office of Information and Regulatory Affairs.

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies

I. OMB Responsibilities

Section 515 of the Treasury and General Government Appropriations Act for FY2001 (Public Law 106–554) directs the Office of Management and Budget to issue government-wide guidelines that provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by Federal agencies.

II. Agency Responsibilities

Section 515 directs agencies subject to the Paperwork Reduction Act (44 U.S.C. 3502(1)) to—

1. Issue their own information quality guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by the agency no later than one year after the date of issuance of the OMB guidelines;

2. Establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency that does not comply with these OMB guidelines; and

3. Report to the Director of OMB the number and nature of complaints received by the agency regarding agency compliance with these OMB guidelines concerning the quality, objectivity, utility, and integrity of information and how such complaints were resolved.

III. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies

1. Overall, agencies shall adopt a basic standard of quality (including objectivity, utility, and integrity) as a performance goal and should take appropriate steps to incorporate information quality criteria into agency information dissemination practices. Quality is to be ensured and established at levels appropriate to the nature and timeliness of the information to be disseminated. Agencies shall adopt

specific standards of quality that are appropriate for the various categories of information they disseminate.

2. As a matter of good and effective agency information resources management, agencies shall develop a process for reviewing the quality (including the objectivity, utility, and integrity) of information before it is disseminated. Agencies shall treat information quality as integral to every step of an agency's development of information, including creation, collection, maintenance, and dissemination. This process shall enable the agency to substantiate the quality of the information it has disseminated through documentation or other means appropriate to the information.

3. To facilitate public review, agencies shall establish administrative mechanisms allowing affected persons to seek and obtain, where appropriate, timely correction of information maintained and disseminated by the agency that does not comply with OMB or agency guidelines. These administrative mechanisms shall be flexible, appropriate to the nature and timeliness of the disseminated information, and incorporated into agency information resources management and administrative practices.

i. Agencies shall specify appropriate time periods for agency decisions on whether and how to correct the information, and agencies shall notify the affected persons of the corrections made.

ii. If the person who requested the correction does not agree with the agency's decision (including the corrective action, if any), the person may file for reconsideration within the agency. The agency shall establish an administrative appeal process to review the agency's initial decision, and specify appropriate time limits in which to resolve such requests for reconsideration.

4. The agency's pre-dissemination review, under paragraph III.2, shall apply to information that the agency first disseminates on or after October 1, 2002. The agency's administrative mechanisms, under paragraph III.3., shall apply to information that the agency disseminates on or after October 1, 2002, regardless of when the agency first disseminated the information.

IV. Agency Reporting Requirements

1. Agencies must designate the Chief Information Officer or another official to be responsible for agency compliance with these guidelines.

2. The agency shall respond to complaints in a manner appropriate to the nature and extent of the complaint. Examples of appropriate responses include personal contacts via letter or telephone, form letters, press releases or mass mailings that correct a widely disseminated error or address a frequently raised complaint.

3. Each agency must prepare a draft report, no later than April 1, 2002, providing the agency's information quality guidelines and explaining how such guidelines will ensure and maximize the quality, objectivity, utility, and integrity of information, including statistical information, disseminated by the agency. This report must also detail the administrative mechanisms developed by that agency to allow affected persons to seek and obtain appropriate correction of information maintained and disseminated by the agency that does not comply with the OMB or the agency guidelines.

4. The agency must publish a notice of availability of this draft report in the **Federal Register**, and post this report on the agency's website, to provide an opportunity for public comment.

5. Upon consideration of public comment and after appropriate revision, the agency must submit this draft report to OMB for review regarding consistency with these OMB guidelines no later than July 1, 2002. Upon completion of that OMB review and completion of this report, agencies must publish notice of the availability of this report in its final form in the **Federal Register**, and post this report on the agency's web site no later than October 1, 2002.

6. On an annual fiscal-year basis, each agency must submit a report to the Director of OMB providing information (both quantitative and qualitative, where appropriate) on the number and nature of complaints received by the agency regarding agency compliance with these OMB guidelines and how such complaints were resolved. Agencies must submit these reports no later than January 1 of each following year, with the first report due January 1, 2004.

V. Definitions

1. ''Quality'' is an encompassing term comprising utility, objectivity, and integrity. Therefore, the guidelines sometimes refer to these four statutory terms, collectively, as ''quality.''

2. ''Utility'' refers to the usefulness of the information to its intended users, including the public. In assessing the usefulness of information that the agency disseminates to the public, the agency needs to consider the uses of the information not only from the perspective of the agency but also from the perspective of the public. As a result, when transparency of information is relevant for assessing the information's usefulness from the public's perspective, the agency must take care to ensure that transparency has been addressed in its review of the information.

3. ''Objectivity'' involves two distinct elements, presentation and substance.

a. ''Objectivity'' includes whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner. This involves whether the information is presented within a proper context. Sometimes, in disseminating certain types of information to the public, other information must also be disseminated in order to ensure an accurate, clear, complete, and unbiased presentation. Also, the agency needs to identify the sources of the disseminated information (to the extent possible, consistent with confidentiality protections) and, in a scientific, financial, or statistical context, the supporting data and models, so that the public can assess for itself whether there may be some reason to question the objectivity of the sources. Where appropriate, data should have full, accurate, transparent documentation, and error sources affecting data quality should be identified and disclosed to users.

b. In addition, ''objectivity'' involves a focus on ensuring accurate, reliable, and unbiased information. In a scientific, financial, or statistical context, the original and supporting data shall be generated, and the analytic results shall be developed, using sound statistical and research methods.

i. If data and analytic results have been subjected to formal, independent, external peer review, the information may generally be presumed to be of acceptable objectivity. However, this presumption is rebuttable based on a persuasive showing by the petitioner in a particular instance. If agency-sponsored peer review is employed to help satisfy the objectivity standard, the review process employed shall meet the general criteria for competent and credible peer review recommended by OMB–OIRA to the President's Management Council (9/20/01) (http://www.whitehouse.gov/omb/inforeg/oira_review-process.html), namely, ''that (a) peer reviewers be selected primarily on the basis of necessary technical expertise, (b) peer reviewers be expected to disclose to agencies prior technical/policy positions they may have taken on the issues at hand, (c) peer reviewers be expected to disclose to agencies their sources of personal and

institutional funding (private or public sector), and (d) peer reviews be conducted in an open and rigorous manner.''

ii. If an agency is responsible for disseminating influential scientific, financial, or statistical information, agency guidelines shall include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties.

A. With regard to original and supporting data related thereto, agency guidelines shall not require that all disseminated data be subjected to a reproducibility requirement. Agencies may identify, in consultation with the relevant scientific and technical communities, those particular types of data that can practicable be subjected to a reproducibility requirement, given ethical, feasibility, or confidentiality constraints. It is understood that reproducibility of data is an indication of transparency about research design and methods and thus a replication exercise (i.e., a new experiment, test, or sample) shall not be required prior to each dissemination.

B. With regard to analytic results related thereto, agency guidelines shall generally require sufficient transparency about data and methods that an independent reanalysis could be undertaken by a qualified member of the public. These transparency standards apply to agency analysis of data from a single study as well as to analyses that combine information from multiple studies.

i. Making the data and methods publicly available will assist in determining whether analytic results are reproducible. However, the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections.

ii. In situations where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken. Agency guidelines shall, however, in all cases, require a disclosure of the specific data sources that have been used and the specific quantitative methods and assumptions that have been employed. Each agency is authorized to define the type of robustness checks, and the level of detail for documentation thereof, in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.

C. With regard to analysis of risks to human health, safety and the environment maintained or disseminated by the agencies, agencies shall either adopt or adapt the quality principles applied by Congress to risk information used and disseminated pursuant to the Safe Drinking Water Act Amendments of 1996 (42 U.S.C. 300g–1(b)(3)(A) & (B)). Agencies responsible for dissemination of vital health and medical information shall interpret the reproducibility and peer-review standards in a manner appropriate to assuring the timely flow of vital information from agencies to medical providers, patients, health agencies, and the public. Information quality standards may be waived temporarily by agencies under urgent situations (*e.g.,* imminent threats to public health or homeland security) in accordance with the latitude specified in agency-specific guidelines.

4. ''Integrity'' refers to the security of information—protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.

5. ''Information'' means any communication or representation of knowledge such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms. This definition includes information that an agency disseminates from a web page, but does not include the provision of hyperlinks to information that others disseminate. This definition does not include opinions, where the agency's presentation makes it clear that what is being offered is someone's opinion rather than fact or the agency's views.

6. ''Government information'' means information created, collected, processed, disseminated, or disposed of by or for the Federal Government.

7. ''Information dissemination product'' means any books, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic, an agency disseminates to the public. This definition includes any electronic document, CD–ROM, or web page.

8. ''Dissemination'' means agency initiated or sponsored distribution of information to the public (see 5 CFR 1320.3(d) (definition of ''Conduct or Sponsor'')). Dissemination does not include distribution limited to government employees or agency contractors or grantees; intra- or inter-agency use or sharing of government information; and responses to requests for agency records under the Freedom of Information Act, the Privacy Act, the Federal Advisory Committee Act or other similar law. This definition also does not include distribution limited to correspondence with individuals or persons, press releases, archival records, public filings, subpoenas or adjudicative processes.

9. ''Influential'', when used in the phrase ''influential scientific, financial, or statistical information'', means that the 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. Each agency is authorized to define ''influential'' in ways appropriate for it given the nature and multiplicity of issues for which the agency is responsible.

10. ''Reproducibility'' means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased). If agencies apply the reproducibility test to specific types of original or supporting data, the associated guidelines shall provide relevant definitions of reproducibility (*e.g.,* standards for replication of laboratory data). With respect to analytic results, ''capable of being substantially reproduced'' means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error.

[FR Doc. 02–59 Filed 1–2–02; 1:36 pm]
BILLING CODE 3110–01–M

Editorial Note: Due to numerous errors, this document is being reprinted in its entirety. It was originally printed in the **Federal Register** on Thursday, January 3, 2002 at 67 FR 369–378 and was corrected on Tuesday, February 5, 2002 at 67 FR 5365.

[FR Doc. R2–59 Filed 2–21–02; 8:45 am]
BILLING CODE 1505–01–D