

From: thomas.j.smith@census.gov [mailto:thomas.j.smith@census.gov]  
Sent: Monday, September 12, 2005 4:01 PM  
To: Harris-Kojetin, Brian A.  
Cc: hermann.habermann@census.gov; nancy.m.gordon@census.gov;  
thomas.l.mesenbourg.jr@census.gov; preston.j.waite@census.gov;  
alan.r.tupek@census.gov; ruth.ann.killion@census.gov  
Subject: Comments on Revisions to Stat Policy 1 & 2

Brian,

Thank you for the opportunity to comment on OMB's proposed revisions to Statistical Policy Directive No. 1, "Standards for Statistical Surveys," and Statistical Policy Directive No. 2, "Publication of Statistics."

Here are the Census Bureau's comments on the proposed revisions. These comments are also attached as a MSWord file.

1. The new standards and guidelines are much more extensive/lengthy than the Directives they are replacing. While more specific, it is unclear how the detail is intended to be used by Federal agencies. In particular, what are the expectations placed upon agencies to follow the proposed standards and guidelines? The document lists 20 standards, each of which is accompanied by several guidelines that are identified as "best practices useful in fulfilling the goals of the standard." The document includes a Glossary, but the terms standard and guideline are not defined. For example, is a standard a "goal?" The bottom line here is that the material presented is extensive and useful as far as providing guidance, but it is vague about the extent to which agencies must adhere to the 20 standards put forward.

2. The report does make several statements that seem to indicate that these standards are expected to be loosely adhered to and subject to agency judgment. As examples, the Introduction of the report states that the application of standards "requires judgment that balances such factors as the uses of the resulting information and the efficient allocation of resources" (p. 7), that "For each statistical survey in existence when these standards are issued and for each new survey, the sponsoring agency should evaluate compliance with applicable standards" (p. 8), and "The provision of standards and guidelines cannot substitute for agency judgment about the most appropriate expenditure of funds" (p. 8). If so, extra burden placed upon agencies in complying with these standards would seem to be minimal.

3. Occasionally, acronyms or special terms were used that made it difficult to interpret the standard or guideline. For example, in guideline 2.2.1 the statement "Provides any additional information to potential respondents that the agency is required to supply (e.g., see 5 CFR 1320.8(b)(3))" is not understandable.

4. Glossary comments. The definitions used for some terms differ from those that a demographic glossary would use for the same terms, such as coverage, coverage error, or estimates. Also, SIPP is usually identified as a longitudinal survey, but it does not seem to fit the definition put forward in the glossary. Finally, some of the definitions use other terms that are not elsewhere defined, but are not common use terms, such as convenience sampling, judgement sampling, quota sampling, and snowball

sampling (see probabilistic methods definition).

5. On Page 8, item k, cost estimates, could a website be given for the guide to estimate reporting costs?

6. Page 8, questionnaire section-- should the privacy notice be mentioned as something that is needed on the questionnaire or advance letters?

7. In the pretesting section, page 7, should there be mention of cognitive testing, focus groups, behavioral coding similar to what we have in the demographic pretesting standard? Also, in the questionnaire and instructions section on page 8, reference should be made to the pretesting of questionnaires, referring back to that section.

8. Standard 1.3 under Section 1.3 should read: "Nonresponse bias analyses should be conducted," rather than "Nonresponse bias analyses must be conducted."

9. Please explain the term "substitution" in Guideline 1.3.1.

10. Each of the nonsampling error thresholds in Guidelines 1.3.2, 1.3.4, and 1.3.5 seem unrealistically high. The introductory statements indicate that OMB acknowledges that some proposed standards may be unattainable due to individual circumstances, but setting an expectation unrealistically high does not seem to be productive. We recommend changing the response rate threshold for surveys that develop frames for other surveys from 95% to 80% or 85%. The other nonsampling error thresholds should also be lowered to more reasonable numbers.

11. In Guidelines 2.3.3 or 2.3.4 there should be a statement indicating that if an agency is using telephone interviewing, the agency's name should appear on Caller ID. Given the flood of marketing calls, potential respondents might be more likely to answer if their caller ID indicated an agency of the Federal government instead of a telemarketer.

12. Guideline 3.2.2. The formula provided looks different from the response rate definition just released in a Census Bureau Standard. If you break it down, it's the same as the Alternative Response Rate (sum of all responses) / (by sum of all eligible units) + (sum of all units with unknown eligibility) \* (e) which is in the standard. This does not pose any problem.

13. Guideline 3.2.8: I would like to see the guideline add something here about studying variation within the respondent set as an acceptable nonresponse bias study technique -- this includes things like comparing response rates on subgroups; using prior wave data; and analyzing estimates by level of effort. These techniques are cheap and readily available unlike good sample frame variables or externally matched datasets.

14. Guideline 3.2.8: It would be helpful if the guideline added something here about studying variation within the respondent set as an acceptable nonresponse bias study technique -- this includes things like comparing response rates on subgroups; using prior wave data; and analyzing estimates by level of effort. These techniques are cheap and readily available unlike good sample frame variables or externally matched datasets.

15. The material is heavily oriented toward the design, collection, and

processing of survey data, and provides a more cursory treatment of estimates, analysis, review, and reports. While footnote 2 notes that the section 7 heading was changed to "Data Dissemination" from the original Federal Register notice category, "Dissemination of data by published reports, electronic files, and other media requested by users," the new document does not seem to deal explicitly with the publication of statistics. For example, where is the equivalent section in the new standards to the section in Directive No. 1 with the heading, "Preparation and Publication of Final Report"?

16. The dissemination section refers to "information products," which may also be intended to include publications. Guideline 7.1.1 states, in part, that major information products should "Ensure equivalent, timely access to all data users." I would just point out that the Census Bureau has adopted a fairly common practice to release some data products on the Internet only and that publications are frequently issued on the Internet in ".pdf" format weeks before printed copies are available. One could argue that such dissemination does not provide equivalent access to all data users.

17. Not sure as to whether or not the Census Bureau currently follows some of the guidelines presented. For example, in a few places, the document refers to quantifying nonsampling error, as in Guideline 7.3.3, which states, "develop and implement methods for bounding or estimating the nonsampling error." Further, Guideline 7.3.4 states "produce a periodic evaluation report, such as a methodology report, that itemizes all sources of identified error." Do such reports exist from our current surveys?

18. In the release of data, should there be anything said about not releasing data because of low response and other errors?

19. It is unclear how these guidelines are to be interpreted in relation to the dissemination of estimates and reports prepared by U.S. statistical agencies based on statistical surveys conducted in other countries.

(See attached file: stat policy comments.wpd)

Thomas Smith  
Management Analyst  
Administrative & Management Systems Division  
U.S. Census Bureau  
Room 3110, FB3  
301-763-1181

From: alan.r.tupek@census.gov [mailto:alan.r.tupek@census.gov]  
Sent: Thursday, September 15, 2005 5:18 PM  
To: mary.m.savoy@census.gov  
Cc: Harris-Kojetin, Brian A.; Cohen\_Steve@bls.gov;  
jay.casselberry@eia.doe.gov; Madans, Jennifer H.; kevin.cecco@irs.gov;  
lowanda.r.rivers@census.gov; marilyn.seastrom@ed.gov;  
mary.m.savoy@census.gov; patrick.e.flanagan@census.gov;  
william\_arends@nass.usda.gov  
Subject: More Census Bureau comments

I gave the Census Bureau's Economic Directorate an extension on comments.  
Here they are --

(See attached file: OMB Directive 1-2 ECON R&M-9-14-05.doc)

A1

### OMB Directive 1 and 2 Comments from ECON R&M Areas

1. General comment - Through p15 guidelines use the word 'should'. After that, the guidelines are phrased in such a way to make it sound as though they are required. Consistency may be desired.
2. p8 - The last paragraph gives folks an out -- if strict application of a standard is not practical or feasible the agency should consider alternative methods to achieve the standard's purpose. The agency is to document reasons why it could not meet and what actions it has or will take to address any resulting issues. This paragraph allows for the flexibility needed under tight budget and staffing conditions.
3. p13 - Key Terms - Add coverage, measure of size, universe, expected yield by stratum, and estimated efficiency of sample design.
4. p15 - For Economic Surveys, whether to conduct nonresponse bias analyses should be based on quantity response rates or total quantity response rates. So, standard 1.3 should be modified to reflect this. A suggested rewording follows:

**Standard 1.3:** Agencies must design the survey to achieve the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data collection costs, to ensure that survey results are representative of the target population so that they can be used with confidence to inform decisions. Nonresponse bias analyses must be conducted when response rates suggest the potential for bias. Response rate definitions appropriate for the type of data being collected should be used to make this determination.

5. p16 - A suggested revision to Guideline 1.3.1 follows:

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

6. p16 - A suggested revision to Guideline 1.3.2 follows:

**Guideline 1.3.2:** Design data collections that will be used for sample frames for other surveys (e.g., the Decennial Census, and the Common Core of Data collection by the National Center for Education Statistics) to meet a target response rate or provide a justification for a lower anticipated rate. For demographic data the target unit response rate should be at least 95 percent. For economic data the target quantity or total quantity rate should be at least 80 percent.

7. p16 - We suggest combining and modifying Guidelines 1.3.4 and 1.3.5 as follows:

**Guideline 1.3.4:** Plan for a nonresponse bias analysis for demographic data collections if the expected unit response rate is below 80 percent, or if the expected item response rate is below 70 percent for any items used in a report. For economic data collections, plan for a nonresponse bias analysis if the quantity or total quantity response rate is below 65 percent.

8. p17 - A suggested revision to Standard 1.4 follows:

**Standard 1.4:** Agencies must pretest all components of a survey either through direct testing or through prior successful fielding or experience to ensure that measurement error is controlled and that the components function as intended when implemented in the full-scale survey.

9. p17 - Key Terms - The list is lengthy, but not exhaustive. This will lead some readers to think that this standard covers only the topics listed in the Key Terms. From the “soft” qualitative side, a number of topics are not covered, and we do not pretend to make the list complete with our recommended additions. Most noticeably, “usability testing” is not included in the Key Terms list, but is included in Guideline 1.4.1. Other methodologies that we suggest adding to the Key Terms list include:

- Respondent debriefings: Response analysis surveys (RAS), which are included in the Key Terms list, are “glorified” respondent debriefings in that the term RAS usually implies a formal, systematic implementation of respondent debriefings.
- Record-keeping studies
- Exploratory or feasibility studies - Is this what might be meant by the reference to “item feasibility” in Guideline 1.4.2.)
- Anthropological or ethnographic studies

Some of the items listed in the Key Terms do not seem appropriate to the notion of “pretesting” – e.g., response rates. Also, response analysis surveys (RAS) are more likely used for data quality evaluation, rather than pretesting, while respondent debriefings might be dove-tailed with a pilot prior to production data collection – the latter is pretesting; the former is not. However, in ongoing periodic surveys, a RAS might be

conducted after an iteration to inform decisions about future redesign of the form or other data collection and processing procedures.

10. p17 - Guideline 1.4.1 - Psychologists might consider this use of the term “cognitive testing” to be inappropriate. A lot of the pretesting we do for establishment surveys is not, strictly speaking, cognitive. Our investigations into the availability of requested data in business records are not cognitive. Our investigations into identifying appropriate respondent(s) are not cognitive. Our response task analyses are not cognitive. Yet they are integral aspects of pretesting for establishment surveys.  
The standard should not name specific pretesting methodologies. It should refer more generally to the selection and use of appropriate pretesting methodologies to meet the goals of a study. Examples may be provided, but it should be clear that the listed examples are not exhaustive. Also, OMB needs to be careful and recognize that the terminology for some of these methods is not consistently used or defined in the survey research field. (We learned this very quickly in our “Team USA” interagency research paper for QDET.)

11. pp17, 18 - Guidelines 1.4.2 - The description of “field test” seems to imply the need for a full “dress rehearsal” to test all systems together. We hope that is not the intent, because doing full “dress rehearsals” for every survey, every time there’s a change, would be burdensome for business respondents, not to mention cost prohibitive for agencies. Besides, full dress rehearsals are not necessarily needed, depending on the research issue. Again, from the standpoint of questionnaire or electronic instrument design, full-scale “field tests” or “pilots” may not be necessary. The design or the scope can be very targeted to the needs of the research question, problem or issue.

A number of the activities listed in Guideline 1.4.2 could sometimes be tested in simulations, rather than requiring new, live field-collected data. Past data could be used to test sample selection procedures, electronic data collection capabilities, edits, estimation, files, tabulations, etc. These may well be more effective than a field test, because we know what the answer should be when we run the tests, and can then evaluate the differences or failures.

Our main point is, again, that the methodology selected for doing the testing should fit the needs of the test. Some things can be tested without going to the “field.” And adequate testing may not require every survey step being tested in lockstep with the others. Also, the guideline needs to clearly differentiate between “field testing” (also called pilot testing) and a “full dress rehearsal.”

12. p22 - A suggested revision to Guideline 2.3.2.3 follows:

**Guideline 2.3.2.3** - The questionnaire is pretested to identify problems with interpretability and ease in navigation or has been shown to be successful in previous implementations.

13. p23 - Guideline 2.3.5.3 - We assume that what is meant here by selecting a "random subsample of nonrespondents" is selecting a probabilistic subsample that is representative of the nonrespondents, not simply an SRS. We may want the selection to be similar to the initial full selection, or to over-sample some components of nonrespondents to ensure that we get at least some responses from these components.
14. p26 - We suggest avoiding the use of 'standard formulas' to avoid confusion. Formulas considered standard for OMB reporting may not be the appropriate formulas for determining the potential nonresponse bias for Economic data. We propose that Standard 3.2 be reworded as follows:

**Standard 3.2:** Agencies must appropriately measure, adjust for, report, and nonresponse to assess the impact on data quality and to inform users. Response rates must measure the proportion of the eligible sample or the proportion of key characteristics that is represented by the responding units in each study, as an indicator of potential nonresponse bias.

15. p27 - We suggest dropping Guideline 3.2.1. First, it says that the response rates can be weighted and unweighted. Then it says that they are based on an either/or. The second part of the either/or says that it applies only to establishment surveys which means that household surveys must satisfy the first part, i.e., that response must be based on the probability of selection which makes them weighted, which contradicts that they can be unweighted, unless unweighted only applies to establishment surveys. Also probability of selection and 'proportion representative of the total industry' are not parallel structures that can be in an either/or. Certainly, we calculate the proportions representative of the total industry using probabilities of selection.
16. pp27-29 - Guidelines 3.2.2, 3.2.3, 3.2.4, 3.2.5, and 3.2.6 are for demographic surveys. For guideline 3.2.3, we suggest dropping the statement in parenthesis that says 'It is the product, not the sum.' For economic surveys the following response rates apply:

Unweighted Response Rate -The rate of responding units to the sum of eligible units and units of unknown eligibility.

$$[R/(E+U)] * 100$$

Quantity Response Rate - The rate of total weighted quantity for responding units to the total estimated quantity for all units eligible for tabulation.

$$\left[ \sum_{i=1}^R w_i t_i / T \right] * 100$$

Total Quantity Response Rate -The rate of total weighted quantity of data from responding units and from sources determined to be of equivalent quality as data provided by respondents to the total estimated quantity for all units eligible for tabulation.

$$\left[ \sum_{i=1}^{R+A} w_i t_i / T \right] * 100$$

where E is the number of units eligible for data collection. This is the number of units for which an attempt has been made to collect data and it is known that the unit belongs to the target population. Eligible units include units that provide sufficient information to be considered a response as well as units that do not provide sufficient information to be considered a response.

U is the number of units for which eligibility for data collection could not be determined. This occurs if there is an attempt to collect data from a unit, and this attempt is not successful and there is no information available about whether or not the unit is a member of the target population. Units whose forms are not deliverable as addressed have unknown eligibility.

A is the number of units belonging to the target population for which it was decided to not collect survey data, but instead to obtain administrative data<sup>1</sup> from sources determined to be of equivalent quality as data provided by respondents or to impute data from data based on a validated model. The decision to not collect survey data must have been made for survey efficiency only and for reasons other than that a unit had been a refusal in the past.

R is the number of eligible units for which an attempt was made to collect data, the unit belongs to the target population, and the unit provided sufficient data to be classified as a response. In a multi-mode survey or census, responses could be obtained by mail, internet, telephone, fax, or touch-tone data entry/voice recognition.

$w_i$  is the sampling weight for the  $i$ th unit.

$t_i$  is the quantity of a key variable for the  $i$ th unit.

T is the estimated (weighted) total of the variable t over the entire population represented by the frame. T is based on actual data (and administrative data for some surveys) and on imputed data or nonresponse adjustment.

<sup>1</sup> Here, administrative data means data that are collected for other than statistical purposes, such as data needed to manage programs in a non-statistical agency.

17. p29 - Guideline 3.2.7 - Some area sample surveys like those conducted in the international surveys select more than one supplemental unit in an area to be used as substitutes for nonresponse units. So 'matched pairs' is not the only way to do this.

18. pp 29, 30 - We suggest combining and modifying Guidelines 3.2.8 and 3.2.9 as follows:

**Guideline 3.2.8:** For demographic data collections with an overall unit response rate of less than 80 percent and for economic data collections with a quantity or total quantity response rate of less than 65 percent, conduct an analysis of nonresponse using rates as defined above, with an assessment of whether the data are missing completely at random. Where appropriate, for a multistage (or wave) survey, focus the analysis on each stage, with particular attention to the “problem” stages. Make comparisons between respondents and nonrespondents across subgroups using available sample frame variables. In the analysis consider a multivariate modeling of response using respondent and nonrespondent frame variables to determine if nonresponse bias exists. Comparison of the respondents to known characteristics of the population from an external source can provide an indication of possible bias, especially if the characteristics in question are related to the survey’s key variables. For demographic data collections, if the item response rate is less than 70 percent, conduct an item nonresponse analysis to determine if the data are missing at random at the item level for at least the items in question.

19. p30 - Guideline 3.2.11 - A suggested modification follows:

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

20. p34 - Guideline 3.4.3 - Change the last part of the first sentence to 'can have read only, write only, or both read and write access to that data set.'
21. p35 - Key terms - Include bias.
22. p35 - Guideline 3.5.1 - Item 2 should be included in item 1.
23. p37 - Key terms - Include calibration.
24. p37 - Guideline 4.1.1.1 - Chop the first sentence to 'Employ weights appropriate for the sample design to calculate population estimates.' Also clarify what is meant by weights since in ratio estimation we employ weights. Does 'employ weights' mean simple expansion estimator? If so, ratio estimator is an alternative.
25. p38 - Guideline 4.1.5 - Is the Census Bureau's new policy for making all data, except confidential, available (and proposed standard for limited releases) in conflict with this guideline?
26. p48 - Guideline 7.3.1.12 - Drop 'and how it was calculated.' This was addressed in item 9.

27. pp48, 49 - Guideline 7.3.1 - For Economic data collections this should include unweighted response rates, quantity response rates, and total quantity response rates instead of the items 14, 15, and 16 listed for demographic collections.

28. Glossary -

Bias -- The definition says 'under a specific design' so what does 'with the same constant error' add to the definition since we are talking about one design. Now if 'with the same constant error' is trying to refer to sample then it is misplaced.

Cross-sectional -- In 3.2.3, cross-sectional is used in a sense not covered in the glossary.

Domain -- In sections 1.2, 1.2.2 and glossary item 'minimum substantively significant effect', domain was not used in the sense in the glossary but in the sense of domain of estimation or population domain.

Add definitions for expected yield per stratum and estimated efficiency of sample design.



**Department of Energy**  
Washington, DC 20585

September 12, 2005

Brian Harris-Kotejin, PhD.  
Statistical and Science Policy Office  
Office of Information and Regulatory Affairs  
Office of Management and Budget  
NEOB Room 10201  
725 17<sup>th</sup> Street NW  
Washington, DC 20503

Dear Brian,

In response to the Office of Management and Budget's (OMB) July 14, 2005 *Federal Register* notice requesting comments on recommendations OMB received from the Federal Committee on Statistical Methodology (FCSM) Subcommittee on Standards for Statistical Surveys to update and revise OMB's Statistical Policy Directives 1 and 2, enclosed are comments of the Energy Information Administration (EIA). The *Proposed Standards and Guidelines for Statistical Surveys* are important updates to Directives 1 and 2. Along with agencies' Information Quality Guidelines, the *Proposed Standards and Guidelines* should help ensure that the results of statistical surveys sponsored by the Federal Government are as reliable and useful as possible. The coverage of all key aspects of planning, conducting, processing, and disseminating statistical surveys will be useful to the principal Federal statistical agencies as well as to other agencies that conduct surveys for statistical purposes.

If you have any questions or would like to discuss the comments, please contact Jay Casselberry of my staff at 202.586.8616 or [jay.casselberry@eia.doe.gov](mailto:jay.casselberry@eia.doe.gov).

Sincerely,

/s/

Guy F. Caruso  
Administrator  
Energy Information Administration  
U.S. Department of Energy

Enclosure

Comments of the Energy Information Administration (EIA) on the  
*Proposed Standards and Guidelines for Statistical Surveys*

**GENERAL COMMENTS**

**Checklist of Standards** - OMB should consider providing a checklist (possibly as an attachment or as a separate document) that has only the standards so a user can more quickly see the 20 standards as they interrelate. This would also facilitate communicating the standards more quickly than requiring staff to read the entire 59-page document.

**Relationship of Statistical Information Collected Across the U.S. Government** - A multi-agency initiative, chaired by the Department of Homeland Security and coordinated by OMB, is currently underway to define the Data Reference Model of the Federal Enterprise Architecture. Although still in draft phase, the concept of uniform metadata tagging in a government-wide schema should be referenced to ensure that survey data maintained over time and often shared between agencies are accessible and of maximum utility to agencies and the public. While outside the scope of the proposed standards and guidelines, a reference to this topic could alert agencies to consider this topic when designing, collecting, and disseminating statistical survey information.

**Measuring Nonresponse Bias** – The document does not address measuring nonresponse bias. While it devotes several pages to a discussion of response rates and the formulae, there is no discussion and no formulae given for measuring the possible bias resulting from nonresponse. It is suggested that the formula for measuring nonresponse bias and some of the discussion found in section 4.2.6 of OMB’s Statistical Policy Working Paper 31 be added to or referenced in this document.

**SPECIFIC COMMENTS**

**Introduction**

Page 9, paragraph beginning “The standards are presented in seven Sections. For each standard, a list of key words for relevant . . .” The wording should be changed to “a list of key terms for . . .” Also, some items listed as “key terms” in the standards (e.g., “survey system” in standards 1.1 and 1.4, and “variance” in 1.2) are not defined in the Glossary. This should be rechecked..

**Section 1**

Section 1.1, Survey Planning

Guideline 1.1.2 should mention priorities within the goals and objectives, which are often critical in balancing opposing goals such as accuracy of estimates vs. increased user requirements for more detailed estimates. When the goals and objectives are both defined and prioritized at the planning stage, the balance of goals and resolution of conflicts can be prevented or resolved more readily and accounted for throughout the survey design and survey process. This could be addressed with a sentence in guideline 1.1.2, item 1, such as “When the goals and objectives are both defined and prioritized at

the planning stage, the balance and resolution of conflicts can be addressed more readily and accounted for throughout the survey design and survey process.”

- Guideline 1.1.2, the eighth activity discusses the analysis plan that identifies analysis issues, objectives, key variables, minimum substantively significant effect sizes (MSSE), and proposed tests. A reference to section 5.1 would be useful and would parallel the section citations for other activities.
- Guideline 1.1.2 should be modified with regard to the preservation of data, documentation, and information products to indicate that a survey plan should include considerations of preservation to ensure the usefulness and reproducibility of survey information over time. This could be addressed by including an additional item in guideline 1.1.2, such as “A plan for the preservation of survey data, documentation, and information products.” Also, up-to-date, complete documentation of a survey is critical to long-term quality and use of survey information and the need for maintaining such documentation should be explicitly mentioned in a guideline.

### Section 1.2, Survey Design

- Guideline 1.2.2 – This guideline addresses the information needed to ensure the sample will yield the data required to meet the objectives. In addition to the areas specified, consider whether the following areas should also be mentioned: 1) use of panels and the effect of panels; 2) sample replacement/rotation/resampling plan (based on criteria such as sample deterioration, sampling error/CV, etc); 3) justification for the variable used for measure of size for Probability Proportional to Size (PPS) (parallel to criteria for stratifying or clustering); 4) nonresponse adjustment and/or imputation methodology (include within the “estimation and weighting plan”); 5) post-stratification plan if appropriate; and 6) sample implementation plan (screening, initiating, substitution or sample replacements, etc).
- Guideline 1.2.4 – This guideline addresses the additional information to be included with the survey instrument. It is recommended that the last sentence be revised to read: “A clear, logical and easy-to-follow flow of questions *from a respondent’s point-of-view* is a key element of a successful survey” to emphasize the focus on clarity to the survey respondents with respect to the instrument.

### Section 1.3, Survey Response Rates

- Guideline 1.3.4 requires that nonresponse bias analysis be *planned for* if the *expected* unit response rate is below 80%. It is recommended that guideline 3.2.8 be referenced (“also see section 3.2.8”) which addresses that a nonresponse analysis *be conducted* when *actual* overall unit response rate is less than 80% (and similarly for 1.3.5 citing 3.2.9 with respect to item response rate of less than 70%).
- Guideline 1.3.4. While it is recognized that the lower the response rate the greater the chance that nonresponse bias exists (i.e., the response values deviates from the whole population due to the differences between respondents and nonrespondents), there has been no determination of how large a response rate is needed to avoid nonresponse bias, and more importantly quantifying the relative size with respect to response bias, other non-sampling error and sampling error—total survey error. The importance of the (potential) systematic nature of nonresponse is downplayed by a guideline that focuses on a unit response rate of 80% and item response rate of 70%. The use of

those rates and not higher or lower rates as the cutoff for analysis should be justified or eliminated in order for the spirit of the guideline to be clear. An agency should focus on whether nonrespondents differ with respect to respondents. Also, no mention is made in the guidelines of subpopulations or other crosscuts of the data (as opposed to item nonresponse) provided by releasable estimates where more risk may exist for non-response bias.

## **Section 2**

### Section 2.1, Developing Sampling Frames

- Guideline 2.1.1, item 1 should also mention descriptions of frame maintenance. Suggested wording is “The manner in which the frame was constructed *and the maintenance procedures.*”
- Guideline 2.1.1 addresses the items required for a frame description; i.e., how constructed, exclusions, coverage issues, coverage mitigation. It is suggested that item 5 be expanded to include accuracy: “Other limitations of the frame including the timeliness *or accuracy* of the frame (*misclassification, eligibility, etc*)”
- Guideline 2.1.3 states coverage rates in excess of 95% overall and for each stratum as desirable, and if below 85% an evaluation of bias be conducted. This guideline only addresses under-coverage despite potential frame issues with over-coverage. Furthermore, it is often the case that there exists no alternative source to use to evaluate coverage, particularly with the limited data sharing available among many statistical agencies. Also, it is suggested that this guideline be expanded to include time dimensions to address deterioration/turn-over of the frame from the time of construction to use. While this comment is related to guideline 2.1.1 regarding periodic evaluation of coverage, it is not clear at what point in time/times the rate requirement needs of 2.1.3 to be met. It is not clear what to do with a dynamic frame, where births and deaths occur frequently. Finally, similar to the comment for response rates; it is also recommended that the specific cutoff of 85% be justified—why 85% not 80% for an evaluation? It appears the cutoff is arbitrarily selected.

## **Section 3**

### Section 3.1, Data Editing

- Guideline 3.1 focuses on checking and editing data to mitigate errors and lists eight specific items to check. It is recommended that this description be modified to also mention the use of exploratory data analysis (including graphical approaches) in assisting to determine potential outliers. The importance of these techniques, as demonstrated in Statistical Policy Working Papers 18 and 25, lies in their ability to more efficiently find outliers without predefined parameters (prespecified fixed ranges based on expert judgment) which constantly need review and updates, yet operate only according to the past. In addition, whenever possible, edit rules and edit parameters should be based on analysis of data with subject matter specialists input in order to produce more effective editing. While much has been written on this subject, little change has taken place and editing continues to be the most time- consuming, resource-intensive part of the survey process, yet performance metrics when available show poor performance.

- Guideline 3.1.3 addresses coding the data set to indicate any actions take during editing and the retention of unedited and edited data. It is recommended that this guideline mention the importance of this coding is to evaluate and improve the performance of the edits and the edit process.

### Section 3.2, Nonresponse Analysis and Response Rate Calculation

- Guideline 3.2.1, as currently written, is not specific on how one would calculate a weighted response rate. It is suggested that this guideline be rewritten as “Calculate response rates, based either on the proportion of units responding or, in the case of establishment surveys, on the estimated proportion that the responding establishments represent of the total industry.” Additionally, some formulae and discussion of weighted response rates should be added here based on section 4.2.3 of Statistical Policy Working Paper 31.
- Guideline 3.2.2 specifies how the response rate will be calculated for unit response rates. In particular, it specifies the measurement referenced in American Association for Public Opinion Research (AAPOR) as RR3, one of six possible approaches. The major differences between the approaches are in the treatment of partial responses and in the treatment of units with unknown eligibility. As a result, AAPOR recognizes a wide range in response rates that may vary depending on the survey, and therefore, recommends that multiple calculations be made and a range of rates be produced. RR3 that is used in this guideline does not specifically address partials but rather includes I (number of **completed** interviews) in the numerator and denominator, and refusals, other nonrespondent eligibles, noncontacted eligibles, non-interviewed unknown eligibility units. Therefore, it leaves partial responses up to the individual to interpret whether these responses qualify as “completed”. If not qualified, it appears that the partials are not included in any of the other categories, and therefore are left out of the calculation. To clarify this, “completed” could be defined in the glossary to include partials, or the formula could be modified to include partials in the denominator at least, and in the numerator according to the intentions of the guideline. In addition, this calculation uses the popular e(U) component where the survey estimates the proportion of units of unknown eligibility that are eligible (e) and multiplies that by the number of non-interviewed sample units of unknown eligibility (U) in order to include only eligible non-interviews in the calculation, yielding a higher response rate than if all unknowns were assumed eligible. The literature contains multiple approaches of estimating “e”, and the biases inherent in the approaches. While it is recognized that e(U) is intended to correct for over-estimating nonresponse, its impact can result in under-estimating nonresponse. Please consider whether the guidelines should recommend that e(U) be calculated using multiple approaches, to provide a range of response rates with clear explanation of each.
- Guideline 3.2.2, which is based on AAPOR’s RR3, uses the terminology “interview” that is associated with personal or phone interviews, as opposed to other collection methods such as mail surveys and internet data collections. It is recommended that a more generic term be used if possible.
- Guideline 3.2.2 uses a formula for unit response rate taken from OMB’s Statistical Policy Working Paper 31, p. 4-4. It is used by statistical agencies that conduct only **interview** surveys and as such it is too narrowly defined for those agencies

conducting non-interview establishment surveys. OMB should consider substituting or adding a response rate formula that is also applicable to establishment survey filings, e.g.,

Unit response rate = number of eligible sampled units responding/ number of eligible sampled units

where eligible units do not include establishments that are out-of-scope, out-of-business, or duplicates.

- Guideline 3.2.3 - For clarity, insert at the bottom of page 27 after “where”: “ $RRU_i$  is the unit level response for stage  $i$ , and  $K = . . .$ ”
- Guideline 3.2.2 - Rewrite the first sentence of the first paragraph on the top of page 28 to read “When the sample is drawn with probability proportional to size (PPS), then the interpretation of  $RRO^C$  can be improved by using size-weighted response rates for the  $K$  stages.” This substitutes “ $K$ ” for the undefined and seemingly out-of-context “ $k_1, \dots, k_{K-1}$ .”
- Guideline 3.2.4 – Based on the text, a reader may expect a more general formula; however the formula provided is only for wave 1. To make the formula more general change the “1’s” to “i’s” and “ $BRO^L$ ” to “ $BRO^{Li}$ .” If this is done, some of the wording (definitions) still needs further work because the content would be then inconsistent with the formula.

#### Section 4

No comments.

#### Section 5

##### Section 5.1 Analysis and Report Planning

- Guideline 5.1.1 addresses what should be included in the analysis plan, and includes the significance level to be used. It is recommended that type II errors also be mentioned because of their importance in many data uses/interpretations.
- Guideline 5.1.2 addresses the inclusion of standard elements of project management in the analysis plan, citing target completion dates and the resources. Please consider whether to emphasize the targets being achievement-based, and that a third critical element, risk planning, be mentioned.

#### Section 6

##### Section 6.1, Review of Information Products

Guideline 6.1.3 should clarify the reference to Section 508 compliance by citing the Act.

#### Section 7

##### Section 7.2 Data Protection and Disclosure Avoidance for Dissemination

Guideline 7.2.2 references FCSM Working Paper 22. It is recommended that a citation be provided for the forthcoming OMB Statistical Policy Directive on Release and Dissemination of Statistical Products Produced by the Federal Government.

##### Section 7.3 Survey Documentation

Guideline 7.3.3. The implementation of estimating bounds on nonsampling error seems more appropriate to section 4 on production of estimates. If the intent is the documentation of the methods or their implementation, it is suggested that this guideline be rewritten to reflect that and/or be combined with 7.3.4 that addresses evaluation reports for recurring surveys.

Guideline 7.3.5 makes reference to “archival policy,” but for all Federal agencies, archival policy is National Archives and Records Administration policy. Suggest in 7.3.5 and 7.4.6 changing the word “policy” to “program.”

## **Glossary**

Suggest the following terms for inclusion in the glossary and key terms as appropriate.

- Dissemination – Suggest using definition from OMB’s Information Quality Guidelines (“Dissemination” is defined to mean “agency initiated or sponsored distribution of information to the public.”)
- Collection of information – Use OMB’s regulations for the paperwork Reduction Act.
- Record - The word “record” is listed in the report ten times. (Page 2; Guideline 2.3.4; Guideline 3.1.1 (twice); Guideline 7.3.1; Guidelines 7.3.5 and 7.4.6 (the only references in the report to “archival policy” which is not defined); Standard 7.4 and Guideline 7.4.2 in the context of “record layout;” parts of the definitions of “imputation;” and “measurement error).

With respect to existing definitions, the following are suggestions:

- “Editing is a procedure that uses available information and some assumptions to derive substitute values for inconsistent values in a data file.” This definition is not compatible with the use of the word editing in the guidelines. The guidelines make use of the more common broader definition that includes both error detection and error correction. It is recommended that the definition be modified consistent with the United Nations Economic Commission for Europe (UNECE) definition: “Editing is the activity aimed at detecting and correcting errors.”
- “Imputation is a procedure that uses available information and some assumptions to derive substitute values for missing values in a data file”. This definition only addresses imputation for missing values and does not include the case of unusable data, i.e. data identified in the editing and corrected. Imputation for correction is addressed indirectly in section 3. Guideline 3.1.1 states, “Editing uses available information and some assumptions to derive substitute values for inconsistent values in a data file.” It is recommended that the definition be modified in the glossary consistent with the broader use of the term and consistent with the UNECE definition: “Imputation is the procedure for entering a value for a specific data item where the response is missing or unusable.”
- “Substitutions are done using matched pairs, in which the alternate member of the pair does not have an independent probability of selection”. This is not a comprehensive definition of “substitutions” but rather an example of one particular method of substitution. It is recommended that a broader perspective be taken such as that of OECD: “In sampling inquiries it is sometimes difficult to make contact with, or obtain information from, a particular member of the sample. In such cases it is sometimes the practice to substitute a more conveniently examined member of the

population in order to maintain the size of the sample. Any such substitution should, however, be carried out upon a strictly controlled plan in order to avoid bias. Using this perspective, it is recommended that the definition be modified to: "Substitution is the process of maintaining and adding to the sample in an unbiased manner in order to ensure it continues to be representative of the population". Also, it is of particular concern that the proposed definition when considered with respect to certain sample selection approaches may eliminate valid methods such as sequential sampling. Furthermore, it may also conflict with the use of frame information used in the sample design to determine probabilities of selection.

From: A E Powell [mailto:PowellAE@GAO.GOV]  
Sent: Tuesday, November 22, 2005 11:06 AM  
To: Harris-Kojetin, Brian A.; Wallman, Katherine K.; Schechter, Susan  
Cc: Maya Chakko; Susan Ragland  
Subject: SAN\_FRANCISCO-#104280-v2-  
PROPOSED\_CHANGE\_TO\_OMB'S\_PROPOSED\_STANDARDS\_FOR\_STATISTICAL\_SURVEYS.DOC

Brian -

Here is a copy of our recommended change to the draft standard. We appreciate your consideration.

A. Elizabeth Powell  
Senior Analyst, Strategic Issues  
U.S. General Accounting Office  
441 G St., NW, Rm. 2440C  
Washington, DC 20548  
Telephone - 202-512-6268  
Fax - 202-512-6880  
Email - powellae@gao.gov

### Potential Change to OMB's *Proposed Standards and Guidelines for Statistical Surveys*

In reviewing OMB's *Proposed Standards and Guidelines for Statistical Surveys*, we are concerned that standard 1.1 (survey planning) only indirectly addresses the importance of assessing potential duplication and overlap. The **standard should require** that the written plan show that each survey must be assessed to ensure that it does not contain "unnecessary duplication." However, the current language does not include any direct mention of the importance of limiting overlap and eliminating unnecessary duplication.

We wanted to raise this as an issue for your consideration as OMB is moving to finalize this guidance. Although it is past the open comment period, we hope that you agree that this point merits attention both for new surveys and to ensure a complete reexamination of sources that may have become available since existing surveys were initiated or revised.

Our suggestion is to revise the standard to include after "related and previous surveys"

*"and steps taken to prevent unnecessary duplication with other available sources of information;" ...*

We would also suggest a slight revision to the related Guideline 1.1.2. In 2, revise "to ensure that part or all of the data are not available from an existing source," to "*to ensure that part or all of the survey would not unnecessarily duplicate available data from an existing source,*"

-----Original Message-----

From: Susan Ashtianie [mailto:susan.ashtianie@nara.gov]  
Sent: Tuesday, September 13, 2005 4:04 PM  
To: Harris-Kojetin, Brian A.  
Cc: Cheryl StadelBevans; Debra Leahy  
Subject: RE: FR Doc. 05-13837

Dear Dr. Harris-Kojetin:

Thank you for the opportunity to provide comments on the Proposed Standards and Guidelines for Statistical Surveys that the Office of Management and Budget (OMB) received from the Federal Committee on Statistical Methodology. The National Archives and Records Administration (NARA) has two comments.

NARA recommends that Guideline 7.4.6 be changed to read: "Agencies should also arrange to archive data with the National Archives and Records Administration and other data archives, as appropriate, so that data are available for historical research in future years." Guideline 7.4.6 currently reads, "All Microdata products and document should be retained by an agency according to its records disposition and archival policy." The use of "its" as a modifier to "archival policy" may imply that each agency has its own archival policy when, in fact, the Federal Government has a single archival policy embedded in the Federal Records Act. Our recommendation follows the advice of the prestigious National Research Council of the National Academies as stated in its recently published third edition of *Principle and Practices for a Federal Statistical Agency* (Washington, DC: The National Academies Press, 2005), which states that the practices of a Federal statistical agency should include wide dissemination of its data.

In addition, NARA recommends that the guidelines should also include a requirement that any survey have a data management plan that identifies the records, specifies formats, and provides for the authorized disposition of the records.

Sincerely,

Susan Ashtianie  
Acting Director  
Policy and Communications Staff  
National Archives and Records Administration  
(301) 837-1490

Robert P. Parker  
Consultant on Federal Statistics  
6010 Woodacres Drive  
Bethesda, Md. 20816  
<[parkerrobertp@aol.com](mailto:parkerrobertp@aol.com)>

September 12, 2005

Mr. Brian A. Harris-Kojetin  
Office of Information and Regulatory Affairs  
Office of Management and Budget  
725 17<sup>th</sup> Street, NW  
New Executive Office Building, Room 10201  
Washington, DC 20503

Dear Mr. Harris-Kojetin;

This letter provides comments on the July 14, 2005, *Federal Register* notice “Proposed Revisions to OMB Statistical Policy Directive No. 1, Standards for Statistical Surveys, and OMB Statistical Policy Directive No. 2, Publication of Statistics.”

I have reviewed the proposed replacement directive from the perspective of a regular user of Federal statistics. In this capacity, information about the relevance and reliability of the statistics covered by the new directive are critical to decisions about whether a survey’s results can be used for a specific purpose. Although I found the coverage of proposed directive to be rather comprehensive, I recommend that the directive be changed to expand the minimum amount of detail required to be included in the survey documentation to be made available to users. In addition, the directive should be changed to require that survey documentation be made available on the agency Website.

The list of detail required to be included in the survey documentation in Guideline 7.3.1 should be expanded. First, the documentation should include information on comparisons with independent sources, the impact of item nonresponse and item imputations for all items, and additional information on data limitations. Second, the documentation should include OMB Form 83-I and related certifications and supporting statements now provided to OMB under the Paperwork Reduction Act. The proposed directive requires only limited information on data limitations and in Guideline 7.3.2 requires that information provided by agencies to OMB under the Paperwork Reduction Act only be included in internal documentation.

The proposed additions to survey documentation will provide critical information that will significantly improve the ability of users to assess the reliability of specific survey results. (Additional information on my recommendations follows.) Changing Standard 7.3 to require that survey documentation be available on the Website will insure that it is

readily available to users. The expanded survey documentation also will help agencies satisfy the requirements to make public information this information under their Information Quality Act guidelines.

### **Part I. -- Additional or expanded survey documentation**

Comparisons with independent sources -- I recommend that the Directive require agencies to prepare comparisons with independent sources and to include the results in the survey documentation – Although the new Directive mentions such comparisons as something of interest to users, there is no requirement for an agency to undertake such studies and to report the results as part of the survey documentation provided to the public.

Item 2 of Guideline 1.1.2 includes the following as a part of the survey planning process: “A review of related studies, surveys, and reports of Federal and non-Federal sources to ensure that part or all of the data are not available from an existing source, or could not be more appropriately obtained by adding questions to existing Federal statistical surveys. ...”<sup>1</sup>

Standard 3.5 of the new directive states that “Agencies must evaluate the quality of the data and make the evaluation public (through technical notes and documentation included in reports of results or through a separate report) to allow users to interpret results of analyses, and to help designers of recurring surveys focus improvement efforts.” The guidelines (3.5.1) for this standard suggests that agencies “Include an evaluation component in the survey plan that evaluates survey procedures, results, and measurement error (see Section 1.1). Review past surveys similar to the one being planned to determine likely sources of error, appropriate evaluation methods, and problems that are likely to be encountered.” Specific guideline 3.5.1 item 7 states, “Post-collection analyses of the quality of final estimates; the data and estimates derived from the data should be compared to other independent collections of similar data, if available.” Despite this reference to comparison with independent sources, the guideline supporting Standard 7.3, Survey documentation, does not list comparison with independent sources as a required item.

Support for the importance of the publication of such comparisons is in *FCSM Statistical Policy Working Paper No. 31, Measuring and Reporting Sources of Errors in Surveys*, one of the documents cited in the new OMB Directive. This paper devotes a full section (Section 8.2.1) to these comparisons.<sup>2</sup> It reports that “Comparing estimates from a survey to values from independent data sources is a useful method of examining the overall effect of errors on the estimates from the survey, but it is difficult to quantify the benefits. In most cases, the comparisons give a broad overview of the cumulative effect of errors in the survey. In a few cases, comparisons may reveal areas that need to be investigated further and this may lead directly to improvements in the survey procedures or methods.”

---

<sup>1</sup> See appendix A for a more complete excerpt of the Directive.

<sup>2</sup> See appendix B for the full text of this section.

The section continues stating that, “One of the primary beneficiaries of comparisons to independent sources are data users, especially those who are familiar with statistics produced from the independent sources. The comparisons provide users with insights into how the statistics from the survey align to statistics from other sources and highlight potential differences that might otherwise cause confusion.”

It should be noted, that for some series, such as the number employed, the number of persons with health insurance, household incomes, and compensation of employees, explanations of differences with independent sources are already provided by the agencies. In other cases, such as data on occupations, which are collected on several different surveys, comparisons are not readily available so there are no explanations of differences.

Impact of item imputation -- I recommend that guidelines include the reporting of the item nonresponse rates and the impact of imputations on the published data as part of the publicly available documentation. – Although the proposed directive recognizes the importance to users of information on item nonresponse and imputations, the relevant guidelines indicate that the survey documentation to be made public need not provide comprehensive information on item nonresponse, imputation procedures, or the impact of imputations.

Standard 3.2 of the proposed directive recognizes the importance to users of information on item nonresponse and imputations. It states “Agencies must appropriately measure, adjust for, report, and analyze unit and item nonresponse to assess their effects on data quality and to inform users.” Standard 7.3 requires agencies to make survey documentation readily accessible to users. Guideline 7.3.1, which identifies the minimum survey documentation, states in item 16 that item response rates be provided only for “variables with rates below 70 percent” and suggests nothing about the impact of the resulting imputation of nonrespondents. Guideline 7.3.2, which identifies minimum standards for “internally archived” documentation, states in item 11 that the internal documentation should include imputation specifications.

Under the guidelines cited above, agencies may not include in their survey documentation comprehensive information on item nonresponse and most likely will include nothing on the impact of imputation. Information on the impact of imputations on specific items is important to users’ assessments of their reliability. Measures of impact are important because for many dollar-value items, such as individual or business incomes, capital expenditures, or contributions to pension plans, the value per respondent typically differs widely. Consequently, item nonresponse rates, which are regularly computed by statistical agencies, may not be indicative of the reliability of an item where the value of the imputation for a small number of nonrespondents might account for a sizeable amount of the published value.

To include this information in the survey documentation, I recommend that item 16 of Guideline 7.3.1 item 16 be changed to delete “below 70 percent” and a new item 17 be

added that states “Impact of imputations on each variables.” (It should be noted that some agencies already provide such information on their Websites.)

Data limitations -- I recommend that guidelines on data limitations be expanded to insure more survey-specific information and to include user concerns.

The standards/guidelines in the proposed directive on the documentation of data limitations are inadequate in providing users with information to determine reliability and relevance. For example, item 8 of Guideline 3.5.1 suggests the inclusion of evaluation reports and item 6 of Guideline 6.1.2 suggests that data reviewers “Ensure that data sources and technical documentation, including data limitations, are included or referenced.” However, Guideline 7.3.1, which covers the scope of survey documentation, includes only some of the sources of data limitations; it excludes key information such as evaluation reports, comments by users, and, as noted above, information on the results of comparisons with independent sources. In addition, item 13 of this guideline suggests the inclusion of “sources of nonsampling error associated with the survey (e.g., coverage, measurement)” but does not indicate that agencies should report on how these sources specifically impact the survey.

The importance to data users of comprehensive information on data limitations was a key point of the National Academy of Sciences report *Principles and Practices for a Federal Statistical Agency (Third edition)*. That report covered this topic on pages 29-30 and included the following:

“Openness about data limitations requires much more than providing estimates of sampling error. In addition to a discussion of aspects that statisticians recognize as nonsampling errors, such as coverage errors, nonresponse, measurement errors, and processing errors, a description of the concepts used and how they relate to the major uses of the data is desirable. Descriptions of the shortcomings of and problems with the data should be provided in sufficient detail to permit the user to take them into account in the analysis and interpretation of the data.”<sup>3</sup>

To include this information in the survey documentation to be provided to users, I recommend that a new item should be added to Guideline 7.3.1 requiring a separate section on “data limitations” that would cover the types of limitations noted above.

## **Part II. -- Required public access to comprehensive documentation**

I recommend that the information reported on OMB Form 83-I and all of its attachments, certifications, supporting statements and comments (including responses to the *Federal Register* notice already submitted to OMB for information collection approval) be included in the survey documentation. In addition, to insure that all of this information is easily accessible to the users of these data, that an electronic version of the form 83 package for each information survey or study be posted on the agency’s web site with linkages to both the agencies IQA guidelines and the survey or program area.

---

<sup>3</sup> See appendix C for the full text from this section.

Standard 7.3 of the proposed directive requires survey documentation and requires that it be readily accessible to users. However, this standard does not require that information provided for the implementation of the proposed directive or information provided as part of the information collection process be included. (Guideline 7.3.2 indicates that the latter information need be maintained only internally.) Consequently, the proposed directive needs to be strengthened by requiring the survey documentation not only include this material, which must be available in order for agencies to seek OMB approval for a information collection, but also the additional information described in part I of this letter.

Sincerely,

[Signed]

Robert P. Parker  
Consultant on Federal Statistics

## **Appendix A. Selected Standards and Guidelines: Proposed OMB Statistical Policy Directive No. 1**

**Standard 1.1:** Agencies initiating a new survey or major revision of an existing survey must develop a written plan that sets forth a justification, including goals and objectives, potential users, and the decisions the survey may inform; key survey estimates; the precision required of the estimates (e.g., the size of differences that need to be detected); the tabulations and analytic results that will inform decisions and other uses; related and previous surveys; when and how frequently users need the data; and the level of detail needed in tabulations, confidential microdata, and public-use data files.

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

...

2. A review of related studies, surveys, and reports of Federal and non-Federal sources to ensure that part or all of the data are not available from an existing source, or could not be more appropriately obtained by adding questions to existing Federal statistical surveys. The goal here is to spend Federal funds effectively and minimize respondent burden. If a new survey is needed, efforts to minimize the burden on individual respondents are important in the development of new items.

...

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

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

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

1. Potential sources of error, including

- .  Coverage error (including frame errors);
  - .  Nonresponse error; and
  - .  Measurement error, including sources from the instrument, interviewers, and collection process;
2. Data processing error (e.g., keying, coding, editing, and imputation error);
  3. How sampling and nonsampling error will be measured, including variance estimation and studies to isolate error components;
  4. How total mean square error will be assessed;
  5. Methods used to reduce nonsampling error in the collected data;
  6. Methods used to mitigate nonsampling error after collection;
  7. Post-collection analyses of the quality of final estimates; the data and estimates derived from the data should be compared to other independent collections of similar data, if available; and
  8. Make evaluation studies public to inform data users.

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

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

...

6. Ensure that data sources and technical documentation, including data limitations, are included or referenced.

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

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

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

...

3. Quality-related documentation such as conceptual limitations and nonsampling error;

...

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

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

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

2. Description of the sample design, including strata and sampling unit identifiers to be used for analysis; and

3. Final instrument(s) or a facsimile thereof for surveys conducted through a computer-assisted telephone interview (CATI) or computer-assisted personal interview (CAPI) or Web instrument that includes the following:

• All items in the instrument (e.g., questions, check items, and help screens);

- . • Items extracted from other data files to prefill the instrument (e.g., dependent data from a prior round of interviewing); and
- . • Items that are input to the post data collection processing steps (e.g., output of an automated instrument);

4. Definitions of all variables, including all modifications;

...

10. Description of all editing and imputation methods applied to the data and how to remove imputed values from the data;

...

13. Description of the sources of nonsampling error associated with the survey (e.g., coverage, measurement);

14. Unit response rates (weighted and unweighted);

15. Overall response rates (weighted and unweighted); and

16. Item response rates for variables with rates below 70 percent.

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

1. Survey planning and design decisions, including OMB Information Collection Request package;

...

11. Final imputation plan specifications and justifications;

...

## **Appendix B. – Comparisons to Independent Sources: FCSM Statistical Policy Working Paper 31, Pages 8-3 to 8-5)**

### **8.2.1 Comparisons to Independent Sources**

Comparing estimates from a survey to values from independent data sources is a useful method of examining the overall effect of errors on the estimates from the survey, but it is difficult to quantify the benefits. In most cases, the comparisons give a broad overview of the cumulative effect of errors in the survey. In a few cases, comparisons may reveal areas that need to be investigated further and this may lead directly to improvements in the survey procedures or methods.

One of the primary beneficiaries of comparisons to independent sources are data users, especially those who are familiar with statistics produced from the independent sources. The comparisons provide users with insights into how the statistics from the survey align to statistics from other sources and highlight potential differences that might otherwise cause confusion.

The analysis and reporting of comparisons to independent sources is an important ingredient in assessing total survey error. However, these comparisons are not always released to the public, as is reported by the Federal Committee on Statistical Methodology (1988). When the comparisons are released, data producers have done so in a variety of formats. Kim et al. (1996) and Nolin et al. (1997) use a working paper format to compare a variety of statistics from the 1995 and 1996 National Household Education Survey to data from multiple independent sources. Vaughan (1988 and 1993) provides detailed aggregate comparisons of income statistics from the 1984 CPS, Survey of Income and Program Participation (SIPP), and administrative program data in a conference proceedings paper and in a technical report. The Energy Information Administration (EIA) publishes data comparisons as feature articles in its monthly publications (U.S. Energy Information Administration 1999a). Other formats for the release of the comparisons include chapters in quality profiles and appendices in survey reports.

The independent sources of statistics may be either administrative records prepared for nonstatistical purposes or estimates from other surveys. Administrative data or data from program sources are often viewed as more accurate than survey estimates; thus, comparisons to these sources may be used to measure the total survey error. This supposition, however, is not always true because administrative records are frequently not edited or subjected to other assessments as discussed in the previous chapters. Similarly, when the independent source is another survey, care must be taken in evaluating the differences because both surveys may have different error structures.

A key aspect of the evaluation of total survey error by comparing the survey estimates to independent sources is the issue of comparability. If the statistic from the independent data source is error-free, then the difference between it and the survey statistic accurately estimates total survey error. In practice, error-free statistics from independent sources are

virtually nonexistent. Nonetheless, the comparisons are still useful when the independent data source has a low level of error compared to the error in the survey estimate. In this case, the difference between the survey statistic and the independent source may be a useful indicator of the direction and magnitude of the total survey error.

When comparisons are made to independent data sources, the error sources of the independent data must be taken into account along with the error sources of the survey. The most common factors that must be considered when comparisons are made include: the time period of the data collection, coverage errors, sampling errors, nonresponse errors, processing errors, measurement errors, and mode effects. Many of the effects of these factors are easy to understand and need little or no explanation. For example, comparisons between estimates of the number of persons in the United States who were not born in the United States from the 1993 CPS and the 1990 Decennial Census will differ due to changes between the two time periods alone. The way some of these factors might affect comparisons are less obvious and are discussed below.

Coverage is an important aspect of comparability. For example, in random digit dial (RDD) telephone surveys only telephone households are covered so comparisons between RDD survey estimates and independent data sources may be informative about the nature and size of coverage biases in the RDD survey. Nolin et al. (1997) compared estimates from a national RDD survey, the National Household Education Survey, and an independent source from a survey conducted in both telephone and nontelephone households, the CPS. In the review, Nolin et al. (1997) noted that about 6 percent of adults aged 16 years and older who were not enrolled in elementary or secondary school lived in nontelephone households while about 10 percent of the children under 11 years old lived in nontelephone households. The comparisons were used to examine whether the estimates from the RDD survey which were statistically adjusted to reduce coverage biases differ significantly from the CPS estimates. Since the CPS estimates did not have this particular coverage bias, the absence of significant differences was taken as evidence of low levels of coverage bias in the estimates from the RDD survey.

If comparisons with external sources are made regularly over time, a change in the difference between two series may signal a change in the industry that needs to be resolved (coverage error). The EIA regularly compares its petroleum supply data to any available related information (U.S. Energy Information Administration 1999b). In the 1980's, the EIA compared fuel data with data from administrative records at the U.S. Department of Transportation, and in the regular comparison a 4 percent bias had always existed. When the bias increased to 7 percent, research showed that the industry had changed and new types of companies were producing gasoline by methods not covered by the EIA's survey.

The need to consider errors in both the survey and the independent data source is clearly demonstrated when the independent source is subject to sampling errors. In this situation, differences between the survey estimates and the independent data source have variances that are the sum of the variance due to the two sources. For example, Coder and Scoon-Rogers (1996) give differences between March CPS and SIPP estimates of income. Both

surveys are affected by sampling error, so any differences in the estimates from these surveys are also affected by this error

Other chapters in this volume have discussed different sources of measurement error in surveys and these apply equally to the survey and the independent source. Each potential error source, the interview, the questionnaire, the data collection method, and the respondent, ought to be considered. For example, administrative program data may collect family income to determine eligibility for a federal assistance program, but the program may define the concept of family differently from the survey. This could lead to differences that are not due to error in the survey, but to differences in the definitions. Another example is that program data and surveys may count different units: the survey may count persons and the program data may count households.

In summary, comparing survey estimates to independent sources can provide valuable information about the nature of the total survey error in the estimates, but the benefits are limited because the independent sources are also subject to error. These comparisons are most valuable when the statistics from the independent source are highly accurate. When this is the case, the differences observed can be considered valid measures of the total survey error. When the independent source has significant error of its own, the differences may still reveal important features of the survey that data users would find useful even though the differences cannot be considered valid measures of total survey error.

**Appendix C. – Information on Data Limitations from National Academy of Sciences Report, *Principles and Practices for a Federal Statistical Agency (Third edition)*: Pages 29-30**

**Practice 4: Openness About Sources and Limitations of the Data Provided**

An important means to instill credibility and trust among data users and data providers is for an agency to operate in an open manner with regard to the sources and the limitations of its data. Openness requires that an agency provide a full description of its data with acknowledgment of any uncertainty and a description of the methods used and assumptions made. Agencies should provide to users reliable indications of the kinds and amounts of statistical error to which the data are subject (see Brackstone, 1999; Federal Committee on Statistical Methodology, 2001a; see also President’s Commission on Federal Statistics, 1971). Some statistical agencies have developed detailed quality profiles for some of their major series, such as those developed for the American Housing Survey (Chakrabarty, 1996), the Residential Energy Consumption Survey (Energy Information Administration, 1996), the Schools and Staffing Survey (Kalton et al., 2000), and the Survey of Income and Program Participation (U.S. Census Bureau, 1998). Earlier, the Federal Committee on Statistical Methodology (1978c) developed a quality profile for employment as measured in the Current Population Survey. These profiles have proved helpful to experienced users and agency personnel responsible for the design and operation of major surveys and data series (see National Research Council, 1993a).

Openness about data limitations requires much more than providing estimates of sampling error. In addition to a discussion of aspects that statisticians recognize as nonsampling errors, such as coverage errors, nonresponse, measurement errors, and processing errors, a description of the concepts used and how they relate to the major uses of the data is desirable. Descriptions of the shortcomings of and problems with the data should be provided in sufficient detail to permit the user to take them into account in the analysis and interpretation of the data. Openness means that a statistical agency should describe how decisions on methods and procedures were made for a data collection program. It is important to be open about research conducted on methods and data and other factors that were weighed in a decision.

...

In summary, agencies should make an effort to provide information on the quality, limitations, and appropriate use of their data that is as frank and complete as possible. Such information, which is sometimes termed “metadata,” should be made available in ways that are easy for users to access and understand, recognizing that users differ in their level of understanding of statistical data (see National Research Council, 1993a, 1997b). Agencies need to work to educate users that all data contain some uncertainty

and error, which does not mean that the data are wrong but that they must be used with care.

The Information Quality Act of 2000 (see Appendix B) stimulated all federal agencies to develop written guidelines for maintaining and documenting the quality of their information programs and activities. Using a framework developed collaboratively by the members of the Interagency Council on Statistical Policy, individual statistical agencies have developed guidelines for their own data collection programs, which are available on the Internet (see Appendix B).