

Quality Requirements for Releasing Data Products

Version 1.0

Issued: 22 Mar 2007

Census Bureau Standard

Authored by:

Larry Cahoon (DSSD)
Gia Donnalley (DSMD)
Edison Gore (DMD)
Donna Kostanich (DSSD)
Ruth Runyan (CSD)
Ruth Detlefsen (SSSD)
Sharon Stern (HHES)

USCENSUSBUREAU

Helping You Make Informed Decisions



Document Management & Control *

Version	Issue Date	Approval	Description
1.0	22 Mar 07	M&S Council and Associate Directors	Initial Release

* **The most current version of this document is maintained on the Census Bureau Intranet and may be accessed from the website for the Methodology & Standards Council.**

Census Bureau Standard: Quality Requirements for Releasing Data Products

Introduction

The Census Bureau Quality Standards describe required survey or statistical methodology procedures and data product requirements for all U.S. Census Bureau program areas. The Methodology and Standards Council sets these standards, with the concurrence of the Associate Directors of the relevant areas, for Census Bureau data products developed from surveys, censuses, and administrative records.

The Census Bureau has issued five standards governing the dissemination of data products.

1. Review and Approval of Census Bureau Documents and Presentations
2. Dissemination of Census and Survey Data Products
3. Minimal Information to Accompany Any Report of Survey or Census Data
4. Describing the Sources and Accuracy of Tabulations and Estimates
5. Quality Requirements for Releasing Data Products (this standard)

All five standards work together to ensure that the information the Census Bureau publishes is accurate and reliable. In combination, these standards provide the foundation of the quality requirements, describe how customers are informed about the quality of data products, and detail other requirements related to data dissemination.

The Census Bureau's reputation for producing quality data is well deserved. Historically, Census Bureau data products have been high quality, statistically sound, with no identifiable problems, and meet the Public Data Release Criteria described in this standard. The Census Bureau releases them to the public under normal procedures. In rare instances, a product's quality does not meet the Public Data Release Criteria because the data are determined to be statistically unreliable or have unacceptably high nonsampling error. These data may be released, but they must be accompanied by supplemental information describing the problems or concerns.

This standard defines the Public Data Release Criteria and describes how to handle the release of any data products based on surveys or censuses that do not meet the Census Bureau's quality standards.

The Census Bureau is the primary source of high quality statistical data for the Congress, the Executive Branch, and the public. To meet this responsibility, in addition to collecting and analyzing data, the Census Bureau identifies data quality issues, if they exist, anticipates

problems that users might perceive to exist in the data, and addresses issues raised by data users. The Census Bureau provides metadata that accurately describe the data and the concepts measured and that inform users of any known data-related issues.

The Census Bureau recognizes that there are practical limitations to its role of assessing data quality and identifying unanticipated data concerns. The amount, scope, and detail of data collected by the Census Bureau make it difficult to anticipate, identify, and describe all of the potential data problems that might exist in the products produced.

This standard identifies three categories of data. Data in the first category satisfy the Public Data Release Criteria described in this standard. The second category consists of data that do not satisfy these criteria but have published release dates. Data in the second category will always be released. The third category consists of any other data that do not satisfy the criteria; they will be made available only on request. When data from the second and the third categories are released, they must be accompanied by specific metadata, beyond what is required for those in the first category, describing the nature of the quality concerns. This standard does not require that all data that satisfy the Public Data Release Criteria must be released. Often users request custom tabulations that are not part of a normal data release. These custom tabulations may or may not satisfy the Public Data Release Criteria. If these custom tabulations do not satisfy the criteria, the rules for the third category of data apply.

Historically, the Census Bureau has consulted with sponsors and data users regarding official publications and data releases. This practice will continue. Because the public has funded the data collections, the Census Bureau does not intend to withhold data unless their release would violate confidentiality requirements.

Occasionally, concerns may arise about the interpretation and application of this standard. In any such instance, the program managers associated with the project must bring these issues to the attention of the Methodology and Standards Council. The Council, supplemented with six representatives, two each from the Economic, the Demographic, and the Decennial directorates, will address these issues in a timely manner. The manager of the affected program will provide sufficient information to clearly convey the concerns about data quality, interpretation, or application. The extended Council will provide a written statement of its consensus on how to release the problematic data in conformity with this standard and a written clarification of its interpretation and application of the standards. This procedure will facilitate uniformity across the Census Bureau in the interpretation and application of this standard by creating a library documenting the decisions of the Council on specific cases.

Scope

This standard applies to all public releases of data by the Census Bureau using the results of surveys or censuses. A separate standard will detail the criteria for releasing data products based on administrative records. The Census Bureau has a responsibility to inform sponsoring agencies of any limitations of the data and to encourage sponsors to adhere to the requirements of this

standard when they issue data collected by the Census Bureau. Other data users also should be encouraged to adhere to this standard when they issue reports using source data from the Census Bureau.

Standard

The requirements of this standard are set forth in three documents:

1. The standard itself (this document), describing the Quality Requirements for Releasing Data Products ;
2. Metadata That Must Accompany All Census Bureau Data Releases (Supporting Document A); and
3. Metadata That Must Accompany Census Bureau Data Products That Do Not Satisfy the Public Data Release Criteria (Supporting Document B).

Key statistics, items, and classification variables must be identified in order to apply this standard. *Key statistics* should be defined for each data product. *Key items* on questionnaires are important components of the key statistics. *Key classification variables* are main classification variables (e.g., geography, demographic attributes, economic attributes, and industry) of the units to be studied. Each program area must identify the key statistics, key items and key variables and state them in the survey's goals and objectives.

For example, in the Annual Social and Economic Supplement (ASEC) to the Current Population Survey, total income is the sum of several income items. Total wage and salary income would likely be identified as a key item, because it determines the majority of the income for a family. Interest and dividend income likely would not be identified as key items, because these types of income comprise only a small proportion of total income. Similarly, if a specific level of detail is key for a survey, then the standard applies to estimates of the key statistics at that level of detail. In surveys where only a few key estimates do not meet the Public Data Release Criteria, the survey as a whole may nevertheless satisfy the criteria. However, if a majority of the detailed key estimates fail to meet the criteria, the survey as a whole will fail.

In designing a survey, the program manager should ensure that sample sizes would be large enough so that key estimates will satisfy the standard. Occasionally changes may occur that require a reduction in sample sizes, for example budget cuts in the program. In such situations, a survey may no longer be able to support the key statistics for a major classification. For example, the level of geographic detail may need to be reduced or the degree of detail in industry classifications may need to be reduced from four digits to three digits. The program manager must be aware of and document the effects of such changes.

1. Public Data Release Criteria

Two types of error affect the accuracy and reliability of data products, nonsampling error and sampling error. Sections A and B describe the standards for public release and how to address any data quality issues arising from these types of error.

A. Issues Due to Nonsampling Error

It is important to check nonsampling quality indicators, not only for the total survey but also for the various subsets defined by the key variables, such as population characteristics, industries, group quarters, and geography. While indirect indicators can point to potential nonsampling errors, nonsampling error is difficult to quantify. The following points provide guidance on how to assess nonsampling error and the steps to take if data quality concerns arise.

1. The primary concepts being measured by the survey must be clearly defined and measurable and should have been clearly defined when the survey was designed. In addition, derived statistics based on models must clearly measure the concept that is targeted by the model. In cases where the survey operations show that the concept is not clearly defined or that measurement of the concept failed for some reason, the data may be determined to be of unacceptable quality, in which case their release must follow the requirements for category 2 or category 3 data.

Similarly, the requirements for category 2 or category 3 data apply if the key estimates are inconsistent with our base of knowledge about the characteristic being estimated, or if problems that are serious enough to raise concerns about the accuracy of those data occur in any of the following areas: sample design, sampling methods, questionnaire design, data collection, data processing, estimation procedures, or the underlying assumptions of a model.

If a key estimate produced by a program shows no evidence of the problems mentioned above, even though secondary estimates are affected by such problems, the data may be released as category 1, provided that documentation of the problems in the secondary estimates are made available to the public. For example, change in a characteristic often is the key estimate and no evidence exists that data quality problems affect the estimate of change. However, the secondary estimate (the level of the characteristic) may suffer from data quality problems. In such a case, the data may be released as category 1, provided that documentation of the problems in the estimate of level is made available to the public.

2. The Census Bureau Standard: *Minimal Information to Accompany Any Report of Survey or Census Data* and Supporting Document A of this standard both require that certain quality indicators be computed and made available to the public:

survey unit response rates;

coverage rates;

key item response rates; and

the combined effect of these factors.

Except as noted in requirement 5 below, these quality indicators must be computed using weighted data. For demographic data, weights based on the probability of selection without post-collection modifications (such as noninterview adjustments, ratio adjustments, or similar adjustments to the weights) are appropriate.

However, economic survey data tend to be skewed because, for most estimates, a large proportion of the estimate comes from relatively few sample units. For such data, response rates, coverage rates, and similar indicators are more appropriately computed using the weighted variable of interest, such as the Quantity Response Rate described in the Census Bureau Standard: *Response Rate Definitions*, rather than the unit response rate. Again, weights based on the probability of selection without post-collection modifications, such as noninterview adjustments, are appropriate.

3. Threshold goals for the general quality indicators for household and person-based surveys are set at 70 percent for the survey unit response rate, 80 percent for the coverage rate, 80 percent for the response rates of the key items, and 50 percent for the combined effect of the indicators. (A simple multiplication of the first three factors may be sufficient for computing the combined effect.) Threshold goals for the general quality indicators for surveys about businesses, institutions, and government organizations are set at 60 percent for the survey unit response rate, 70 percent for the coverage rate, 70 percent for the response rates of the key items, and 50 percent for the combined effect of the indicators.

If the survey unit response rate, coverage rate, key item response rates, or the combined effect of these factors are below the threshold goals for a particular survey, release of the data must follow the requirements for data in the second or third categories.

4. Generally, Census Bureau data products must be derived from a census or a probability-based sample. Data collections that involve self-selected respondents or that are not from probability samples usually will be in the second or third category.

The Census Bureau, however, conducts a few surveys that do not use probability samples, whose results may be released under the rules for category one. Generally, these 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 Public Data Release Criteria for response rates, coverage rates, and combined effects are not appropriate. Instead, the responding units must represent at least 70 percent of the target universe, based on the unit response rate or the key statistics for the survey (e.g., sales), as appropriate. If the data do not meet this criterion, the Census Bureau will release the data product under the rules for the second or third category. Data products in the second and third categories must be accompanied by the supplemental metadata described in Supporting Document B of this standard.

5. In general, unweighted estimates from surveys may not be released, except for the situation noted below.

Unweighted measures of quality, indicators of quality, or documentation of survey processes and costs may be released, in addition to the weighted measures that must be included in the survey documentation and metadata. Three examples are: when using weighted measures would disrupt a time series of quality measures; when comparing with surveys using unweighted quality measures; and in field reports, because weighted measures of quality are inappropriate for monitoring and managing field activities.

In cases where it is not possible for the Census Bureau to calculate weighted measures of quality, releasing unweighted measures is acceptable. The main example is reports to sponsors for whom the Census Bureau simply collects data, and so has no access to the sampling weights and performs no post-collection processing. Such reports may provide unweighted measures such as response rates. In these surveys, the sponsor processes and releases the data and the accompanying measures of quality.

B. Issues Due to Sampling Error

If the estimated coefficients of variation (CV) for key statistics are larger than 30 percent, the data product will be released under the requirements for category 2 or category 3 data. This standard does not apply to estimates of change of the key statistics, differences between them, or rare characteristics, as long as the key statistics meet the standard.

The application of this standard is not rigid and does not apply to every estimate in every table. The requirement is that the CV standard should be met by the majority of the key estimates. It should not be unduly influenced by geographic areas where the characteristic of interest is rare or small. The Census Bureau is committed to, and will continue to, publish data equivalent to the data we have historically provided.

2. Metadata That Must Accompany All Census Bureau Data Releases

All Census Bureau data products released to the public must include or make available metadata describing the accuracy and reliability of the data and describing the methods used to collect, process, and analyze the data. The Census Bureau Standard: *Minimal Information to Accompany Any Report of Survey or Census Data* and Supporting Document A of this standard, *Metadata That Must Accompany All Census Bureau Data Releases*, describe these required metadata.

3. Release Requirements for Products Not Meeting the Public Data Release Criteria

The Census Bureau has a well-deserved reputation for the production of quality data. However, in rare instances, the data will not meet the quality expectations of the Census Bureau. When data that have quality problems are released (categories 2 and 3 described above), supplemental information necessary to describe the problems and concerns must be included, as described in Supporting Document B. This section describes restrictions on the Census Bureau's use of such data, provides disclaimers to use when they are released, and notes the key requirements for the additional metadata to be provided.

When data problems affect only part of the data release, a statement describing the problem and the Census Bureau's plans to examine and resolve the problem must be included with the data release. The problematic subset of the data set may be released to the public, but it must be accompanied by the supplemental information describing the problems and concerns (see Supporting Document B).

Efforts to address the problem should not degrade other aggregates in the data set. Specifically, combining problem data with other data is not an appropriate solution if it lowers the quality of the aggregates so they do not satisfy the Public Data Release Criteria.

A. Restriction on the Use of Data Not Meeting the Public Data Release Criteria

Census Bureau employees, or other analysts receiving financial support from the Census Bureau, may not prepare professional papers that provide subject matter analysis of data that do not meet the criteria for public release. However, professional papers providing

information and analysis about the data quality problems may be released. Similarly, researchers with Special Sworn Status may not have access to confidential data that, in their public-use form, does not meet the criteria for public release, except to analyze the data quality problems, including developing potential solutions. The revised data also may be used for substantive analyses, provided that the data quality problems have been resolved to the researcher's satisfaction and the Census Bureau has determined that the researcher's solutions are appropriate.

B. Disclaimers

1. The following disclaimer should be included when data in the second category are released.

“Note that the data in this release do not meet all of the criteria outlined in the Census Bureau Standard: *Quality Requirements for Releasing Data Products* and they should be used with caution.”

Immediately following the disclaimer, include a summary of the concerns. The last line of the summary should include the statement, “See <location of description> for a detailed description of the specific quality issues and information about the impact on the data.”

2. The following disclaimer should be included when data in the third category are released.

“These data are being released on request, despite concerns about their quality. It is the Census Bureau's policy not to withhold data that are available, unless releasing such data would violate confidentiality requirements. Because these data do not meet the criteria outlined in the Census Bureau Standard: *Quality Requirements for Releasing Data Products*, the Census Bureau recommends caution when using them.”

Immediately following the disclaimer, include a summary of the concerns. The last line of the summary should include the statement, “See < location of description> for a detailed description of the specific quality issues and information about the impact on the data.”

C. Additional Metadata Requirements

The specific problem or data quality issue and information about the impact on the data must be described, including why the data product does not meet the Public Data Release Criteria. Supporting Document B, “Metadata That Must Accompany Census Bureau Data Products That Do Not Satisfy the Public Data Release Criteria,” describes the additional metadata that are required.

Responsibilities

Program areas that collect, analyze, or release data have the following responsibilities.

Implement and ensure compliance with this standard by:

including requirements and plans to satisfy this standard in the design, schedule, and budget of activities, programs and products;

developing and following procedures needed to implement this standard;

producing and maintaining documentation of the implementation, such as completed checklists, clearance forms, and so forth; and

when circumstances warrant it, following the process described in the Census Bureau Standard: *Waiver Procedure* to request a waiver from requirements of this standard.

Identify key statistics, items, and classification variables and compute quality measures to compare data against thresholds.

Review data releases and indicate data concerns or confirm there are no apparent data problems.

Raise relevant data concerns to the appropriate level (including the Extended Methodology and Standards Council).

Issue a report on all major unresolved quality issues identified with the data. The report should include the input from all areas involved in the production of the data.

The Methodology and Standards Council has the following responsibilities.

Provide guidance to program areas in implementing this standard.

Initiate evaluations, reviews, and updates to this standard, as necessary.

The Extended Methodology and Standards Council has the following responsibilities.

Provide timely feedback on issues sent to them for resolution.

Apply this standard in a consistent manner.

Document agreements on how to release data with quality problems and clarify the interpretation and application of this standard.

Maintain records of prior decisions.

Implementation

The standard is effective April 1, 2007.

Inquiries

Inquiries relating to the interpretation of this standard should be addressed to the Census Bureau's Methodology and Standards Council.

Supporting Documents

- A. Metadata That Must Accompany All Census Bureau Data Releases
- B. Additional Metadata That Must Accompany Census Bureau Data Products That Do Not Satisfy the Criteria for Public Data Release

Recommended by the Census Bureau Methodology and Standards Council:

Howard Hogan 2 / 21 / 07
Howard Hogan
Chair, Methodology and Standards Council
Date

David C. Whitford 2 / 16 / 07
David C. Whitford
Chief, Decennial Statistical Studies Division
Date

Rita Petroni 2 / 16 / 07
Rita Petroni
Chief, Office of Statistical Methods and
Research for Economic Programs
Date

Ruth Ann Killion 2 / 16 / 07
Ruth Ann Killion
Chief, Demographic Statistical Methods Division
Date

Tommy Wright 2 / 16 / 07
Tommy Wright
Chief, Statistical Research Division
Date

Concurrence:

<u><i>Nancy M. Gordon</i></u>	<u><i>3 / 22 / 07</i></u>
Nancy M. Gordon	Date
Associate Director for Strategic Planning and Innovation	

<u><i>Thomas L. Mesenbourg</i></u>	<u><i>3 / 22 / 07</i></u>
Thomas L. Mesenbourg	Date
Associate Director for Economic Programs	

<u><i>Teresa Angueira</i></u>	<u><i>3 / 22 / 07</i></u>
Teresa Angueira	Date
Acting Associate Director for Decennial Census	

<u><i>Howard Hogan</i></u>	<u><i>3 / 22 / 07</i></u>
Howard Hogan	Date
Associate Director for Demographic Programs	