United States Government Accountability Office

GAO

Report to Congressional Requesters

November 2004

DATA QUALITY

Census Bureau Needs to Accelerate Efforts to Develop and Implement Data Quality Review Standards

GAO-05-86

November 2004

DATA QUALITY

Census Bureau Needs to Accelerate Efforts to Develop and Implement Data Quality Review Standards

GAO Accountability • Integrity • Reliability

Highlights

Highlights of GAO-05-86, a report to congressional requesters

Why GAO Did This Study

Data from the decennial census are used to apportion and redistrict seats in the House of Representatives, distribute billions of dollars of federal funds, and guide the planning and investment decisions of the public and private sectors. Given the importance of these data, it is essential that they meet high quality standards before they are distributed to the public. After questions arose about the quality of certain data from the 2000 Census, the requesters asked GAO to review U.S. Census Bureau (Bureau) standards on the quality of data disseminated to the public.

What GAO Recommends

GAO recommends that the Bureau

- accelerate its effort to establish a comprehensive set of data quality review standards by developing and making public a detailed plan, including interim milestones for developing such standards and procedures, and
- include the implementation of data quality review standards in the Bureau's plans for the 2010 Census, and test new draft guidelines on data quality review using the annual American Community Survey or other surveys.

Commerce agreed with our second recommendation but not the first. However, because the Bureau has yet to approve and make public data quality review standards, we continue to believe that it needs to accelerate its effort.

www.gao.gov/cgi-bin/getrpt?GAO-05-86.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Patricia A. Dalton at (202) 512-6806 or daltonp@gao.gov.

What GAO Found

The Bureau did not have detailed agencywide standards for the review of data from the 2000 Census to determine if the data were of sufficient quality for public dissemination. Instead, analysts and managers in different parts of the Bureau primarily used their own judgment and unwritten, program-specific guidance to decide when and whether data should be released and what supporting information should accompany the data. The lack of sufficient data quality review standards led to a variety of problems, including missed opportunities for correcting data before release, inconsistent decisions on disseminating data with similar quality issues, and inadequate communication to users about the reasons for dissemination decisions. As a result, some users of data from the 2000 Census lost confidence in the quality of the data and in the Bureau's review procedures.

In the 4 years since the 2000 Census, the Bureau has publicly issued general information quality guidelines, including eight performance principles, and one new standard that allows individuals to request correction of certain errors in data disseminated by the Bureau. Both of these documents resulted from the enactment of the Information Quality Act in 2000 and the subsequent guidelines issued by the Office of Management and Budget in 2002. However, except for the one standard, the Bureau did not provide any specific guidelines or procedures on the implementation of the general guidelines. The Bureau also began work on other standards, including one on minimal information that must be provided with data and another on discussion of errors in data released to the public. Neither has been issued in final form. In response to our earlier recommendations, the Bureau created an interdirectorate working group charged with developing and publicly issuing Bureau-wide standards for quality in data releases. The working group has taken some steps, but the Bureau has not provided information on the scope or the time frame for its efforts to develop these standards.

The standards that the Bureau has under development and the activities of the working group are encouraging. However, it will be important for the Bureau to proceed with greater urgency to ensure that fully tested standards are in place for the 2010 Census. Until spring 2004, no additional resources were provided to support the working group, and over a year after it began, it has not issued any new standards or said when it will be ready to do so.

A comprehensive, Bureau-wide data quality framework, with interrelated standards, and specific implementing procedures could help ensure consistent decisions about the quality of the data from the next decennial census and conditions under which the data will be disseminated. Moreover, the benefits the Bureau can achieve by developing and effectively implementing comprehensive data quality standards would not be limited to the decennial census. Because they would apply to all data disseminated by the Bureau, it will be important for any new standards to be developed promptly, implemented across the Bureau, and released to the public.

Contents

Letter 1

Results in Brief 2
Background 5
Scope and Methodology 7
Professional Judgment Drove Data Dissemination Decisions 9
The Bureau Has Made Limited Progress in Publicly Issuing New Standards on the Quality of Data Disseminated to the Public since the 2000 Census 14
Greater Commitment to New Standards for Public Dissemination of Data Could Help Bureau Avoid Problems in Disseminating 2010 Census and Other Data 19
Conclusions 21
Recommendations for Executive Action 22
Agency Comments and Our Evaluation 23

Appendix

Appendix I: Comments from the Department of Commerce 25

GAO
Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, D.C. 20548

November 17, 2004

The Honorable Wm. Lacy Clay
Ranking Minority Member
Subcommittee on Technology,
Information Policy, Intergovernmental Relations
and the Census
Committee on Government Reform
House of Representatives

The Honorable Danny K. Davis
The Honorable Carolyn B. Maloney
The Honorable Charles A. Gonzalez
House of Representatives

As one of the nation's principal statistical agencies, the U.S. Census Bureau (Bureau) collects and disseminates data that are used to apportion and redistrict seats in the House of Representatives, distribute billions of dollars of federal funds, and guide the planning and investment decisions of the public and private sectors. Given the importance of Bureau data to our economy and system of governance, census information, like other federal statistics, must be of high quality before it is released to the public. Specifically, the data must be accurate, timely, accessible, relevant, and objective. Failure to meet this threshold could impair decision making and erode public confidence in the information and the Bureau's credibility.

Producing high-quality data is a continuing challenge, in part because the methods used to collect and process census data are complex and subject to some degree of error. Consequently, the Bureau must decide if and when the quality of each set of data is high enough for it to be released and what caveats, if any, are needed to inform users of any shortcomings that could affect whether and how the data are used. The development and use of comprehensive data quality review standards—if they are well documented, transparent, clearly defined, and consistently applied—help statistical agencies make such decisions and communicate the results of the decisions to the public.

After the reliability of certain publicly released data from the 2000 Census was called into question, concerns were raised about the adequacy of the Bureau's data quality review standards. Chief among these concerns was that the Bureau did not routinely and consistently include an adequate

discussion of limitations to the data it disseminates or provide information on how it reaches its dissemination decisions.

At your request, we reviewed the Bureau's data quality review standards.[1] Specifically, as discussed with your offices, we (1) examined the review standards that the Bureau had in place to guide decisions to disseminate 2000 Census data, (2) determined if the Bureau has subsequently developed additional review standards to guide decisions about data quality, and (3) assessed whether any such standards are likely to address for the 2010 Census the data quality review concerns raised after the release of certain data from the 2000 Census.

To meet these objectives, we interviewed Bureau officials, reviewed relevant documents prepared both before and after the enactment of the Information Quality Act of 2000, and examined the guidelines other statistical agencies and organizations have developed governing the public dissemination of data. We did our audit work in Washington, D.C., and at the Bureau's headquarters in Suitland, Maryland, from August 2003 through October 2004, in accordance with generally accepted government auditing standards.

Results in Brief

The Bureau did not have detailed agencywide standards for reviewing data from the 2000 Census to determine if the data were of sufficient quality for public dissemination. Instead, analysts and managers within the different parts of the Bureau primarily used their own judgment and unwritten, program-specific practices to decide when and whether data should be released and what supporting information about data limitations, if any, should accompany them. This led to (1) the dissemination of data with uncorrected and undisclosed quality problems, (2) inconsistent decisions on disseminating data with similar quality problems, and (3) inadequate communication to users about the reasons for dissemination decisions. As a result, some users of data from the 2000 Census lost confidence in the quality of the data and in the Bureau's quality review procedures.

[1] The terms "standards" and "guidelines" are often used without clear definition and sometimes interchangeably. The Bureau defines standards as methodological procedures that are *required* for all Bureau program areas and guidelines as procedures that are *recommended* for all Bureau program areas. We follow that distinction in discussing Bureau guidance. However, the term "guidelines" is also used, particularly in reference to governmentwide requirements, to refer to a broad set of related standards, guidelines, or a combination of these, and we follow that usage where appropriate.

In the 4 years since the 2000 Census, the Bureau has publicly issued information quality guidelines that contain general quality goals and principles and one new standard that allows individuals to request correction of errors in data disseminated by the Bureau. Both of these initiatives came as a result of the enactment of the Information Quality Act in 2000[2] and the subsequent guidelines issued by the Office of Management and Budget (OMB) in 2002.[3] However, except for the one standard, the Bureau did not provide specific guidelines or procedures on the implementation of the general principles articulated in the information quality guidelines.

Since the 2000 Census, the Bureau has also initiated work on several other standards and guidelines on the quality of data released to the public. Some have been approved for internal use but have not yet been made publicly available. For example, one such standard specifies minimal information that must accompany any report of Bureau data. Additionally, the Bureau has identified several other initiatives on data quality review standards, which are in earlier stages of development. For example, the Bureau is working on a Bureau-wide standard for discussion and presentation of errors in data disseminated to the public that will be based on an existing working paper on the subject. Bureau officials said that the Bureau plans to make completed standards publicly available on its Internet site by the end of 2004.

In response to the recommendations contained in our 2003 reports on census counts of Hispanic subgroups[4] and the homeless,[5] the Bureau established an interdirectorate working group charged with developing Bureau-wide standards for quality in data releases. According to Bureau officials, the working group has taken some steps to address the tasks laid

[2] Consolidated Appropriations, 2001, Pub. L. No. 106-554 (2000) (enacting H.R. 5658, §515) referred to by the Office of Management and Budget as the Information Quality Act.

[3] Office of Management and Budget, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Federal Government; Republication*, 67 Fed. Reg. 8452 (Feb. 22, 2002).

[4] GAO, *Decennial Census: Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement*, GAO-03-228 (Washington, D.C.: Jan. 17, 2003).

[5] GAO, *Decennial Census: Methods for Collecting and Reporting Data on the Homeless and Others without Conventional Housing Need Refinement*, GAO-03-227 (Washington, D.C.: Jan. 17, 2003).

out in its charter. However, the Bureau has not provided information on the scope or the time frame for developing these standards.

The standards that the Bureau has under development and activities of the working group are steps in the right direction. However, the Bureau needs to accelerate its efforts to develop and implement quality standards for data it disseminates. Until spring 2004, no additional resources were provided to support the work of the group, and over a year after it began, the group has not issued any new standards or guidelines, nor indicated when it will be ready to do so. Although Bureau officials said that 2010 Census dissemination decisions would adhere to new Bureau dissemination guidelines, the actions the Bureau has taken to date are not enough to ensure that it will avoid in 2010 the types of problems encountered in disseminating data from the 2000 Census. Also, because the Bureau is distributing data from the American Community Survey (ACS),[6] development of needed standards should not wait until 2010.

The development and implementation of a comprehensive, Bureau-wide data quality framework, with interrelated standards, and specific procedures will help ensure (1) the consistency of decisions about the quality of data from the next decennial census, the ACS, and other surveys and (2) the conditions under which the data will be disseminated. Thus, the benefits the Bureau can achieve by implementing comprehensive data quality review standards will not be limited to the decennial census. Because the standards will apply to all of the data publicly disseminated by the Bureau, the standards should be developed promptly and implemented across the Bureau.

Therefore, we recommend that the Secretary of Commerce direct the Director of the U.S. Census Bureau to (1) accelerate the Bureau's effort to establish comprehensive data quality standards and (2) include the implementation of data quality review standards in the Bureau's plans for the 2010 Census.

[6] The ACS is designed to replace the long-form census questionnaire and provide annual data for areas with populations of 65,000 or more and multiyear averages for smaller geographic areas using population and housing counts from the Intercensal Population Estimates. See GAO, *The American Community Survey: Accuracy and Timeliness Issues*, GAO-02-956R (Washington, D.C.: Sept. 30, 2002), and *ACS: Key Unresolved Issues*, GAO-05-82 (Washington, D.C.: Oct. 8, 2004).

The Secretary of Commerce provided written comments on a draft of this report (see app. I). Commerce agreed with our recommendation that the Bureau include the implementation of data quality review standards in its plans for the 2010 Census, and said that the quality review standards will be used for the 2010 Census and for all applicable Bureau programs, including the ACS. However, Commerce did not agree with our recommendation that the Bureau accelerate its effort to establish comprehensive data quality standards. Commerce maintained that the Bureau has already completed much of the work of establishing comprehensive data quality standards and will continue to develop new standards where needed. While these are important steps, most of these standards are not available to the public, and the Bureau still lacks well-documented, transparent, clearly defined quality review guidelines and standards. Thus, we stand by our recommendation and urge the Bureau to accelerate its pace in completing the development of these standards and effectively implementing them.

Background

The Bureau is best known for counting the nation's population every 10 years. In the future, the Bureau intends to collect much of the data that have traditionally been collected during the decennial census from the long-form questionnaire with the annual ACS. Beyond the decennial census, the Bureau also conducts numerous other surveys and censuses that measure changing individual and household demographics and the economic condition of the nation. Lawmakers and agency officials at the federal, state, and local levels rely on these data when they make decisions in a wide range of policy areas. Private-sector decision makers also use census data to guide their business plans.

Because of the critical and varied uses of census information, it is important that the Bureau's published data meet minimum quality standards. In addition, when the data are made public, it is equally important for the Bureau to disclose what has been done to ensure the quality of the data and identify any limitations so that potential consumers can decide whether the data are appropriate for a particular use.

Some degree of error in the census (and in virtually any survey) is inevitable because of limitations in enumeration, processing, and dissemination methods and errors in responses and imputation of data for nonresponses. Given the size and diversity of the U.S. population, the effort to count the entire population and provide detailed demographic characteristics every 10 years is one of the most complex of all government

operations. The Bureau devotes significant resources to minimizing error and improving the quality of the decennial census.

Data quality standards and standardized quality control procedures can provide a consistent basis for making data dissemination decisions and informing the public about the quality of the data made available to it. In 2000, Congress passed what is now known as the Information Quality Act. This legislation directed OMB to issue governmentwide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." The legislation also required each agency to issue its own implementing guidelines that include administrative mechanisms allowing affected persons to correct information maintained and disseminated by the agency.

The OMB guidelines,[7] issued in final form in February 2002, define quality as encompassing utility, objectivity, and integrity. The guidelines require agencies to issue their own implementing guidelines by October 1, 2002. Additionally, they mandate that agencies adopt a standard of quality as a performance goal and act to incorporate data quality criteria into their data dissemination practices. The guidelines also require agencies to develop processes for reviewing the quality of data before they are disseminated.

Although OMB had some general guidance for survey processes prior to the enactment of the Information Quality Act, other than requirements for the evaluation of selected monthly and quarterly economic indicators,[8] there were no governmentwide requirements relating to the quality of data disseminated by the federal agencies. Some statistical agencies within the United States developed their own extensive guidelines and standards that apply to data disseminated to the public. In July 2001, OMB identified the statistical agencies within the Departments of Education and Energy, the National Center for Education Statistics (NCES) and the Energy Information Administration, as good examples of agencies that have

[7] 67 Fed. Reg. 8452.

[8] Office of Management and Budget, *Statistical Policy Directive Number 3: Compilation, Release and Evaluation of Principal Federal Economic Indicators* (Washington, D.C.: July 1985).

developed specific guidelines to implement their broad principles and diverse professional standards.[9]

Statistical agencies in other countries have also developed good examples of comprehensive guidelines for ensuring the quality of data disseminated to the public. Since 1985, Statistics Canada, the central statistical agency of the Canadian government, has published quality guidelines for its statistical activities. Subsequently, it added guidelines on quality assurance processes and management context and developed a policy and standards on informing users about data quality. More recently, the European Union recognized the importance of comprehensive, well-documented guidelines and standards to support its task of developing high-quality, comparable statistics from member countries. All members of the European Statistical System (ESS)[10] have signed a quality declaration and approved 22 recommendations for quality for future work within the system.[11]

Scope and Methodology

To address our first question on the standards that the Bureau had in place to guide its data dissemination decisions, we interviewed census officials, reviewed relevant agency documents, talked to data users, and reviewed various complaints about the quality of 2000 Census data. We built on our prior reports about the quality of data from the 2000 Census on Hispanic subgroups[12] and the homeless[13] and the Bureau's decision-making processes for its decisions on whether to release those data. We also reviewed other GAO reports addressing aspects of the Bureau's procedures

[9] Office of Management and Budget, *Statistical Policy Working Paper 31: Measuring and Reporting Sources of Error in Surveys* (Washington, D.C.: June 2001).

[10] ESS includes Eurostat, the statistical directorate of the European Union; national statistical institutions of member countries; and a variety of academic and other statistical institutes.

[11] European Statistical System, *Quality Declaration of the European Statistical System* (Brussels: September 2001).

[12] GAO-03-228.

[13] GAO-03-227.

for assessing the quality of disseminated data.[14] From these reports, we identified examples of several types of problems the Bureau encountered with 2000 Census data, which might have been alleviated if the Bureau had implemented data quality standards and procedures. Our examples of data quality problems are not comprehensive, but illustrative.

To determine whether the Bureau has since developed Bureau-wide data quality standards, and, if so, whether they would likely address for the 2010 Census the data quality problems raised after the 2000 Census, we interviewed census officials responsible for developing agencywide standards, examined documents related to the development of new standards on data quality review, and reviewed the agency's Internet site for information on data quality review standards available to the public. We also reviewed OMB guidelines on the quality of data disseminated by federal agencies as well as the action taken by the Department of Commerce and the Bureau in response to the guidelines. We attended meetings of the Secretary of Commerce's Decennial Census Advisory Committee, the National Academy of Science Panel on Research on Future Census Methods, and the Washington Statistical Society's conference on Quality Assurance in the Government, all of which examined issues related to the quality of the data disseminated by the Bureau. We also discussed information quality standards and guidelines with officials in Eurostat, the statistical directorate of the European Union.

Additionally, we considered how the Bureau's actions in developing dissemination guidelines could improve the quality of data disseminated after the 2010 Census and for other Bureau data collection programs, such as the ACS that among other things, is intended to replace the long-form census questionnaire. To benchmark the Bureau's progress in developing data quality review standards with that of other statistical agencies, we also reviewed documents from entities that have developed standards for the quality of data disseminated to the public, including NCES; Statistics Canada, the central statistical agency of Canada; and ESS. However, we did not evaluate the implementation or effectiveness of these guidelines and standards or their specific applicability to the Bureau.

[14] See GAO, *2000 Census: Refinements to Full Count Review Program Could Improve Future Data Quality*, GAO-02-562 (Washington, D.C.: July 3, 2002) and *2000 Census: Coverage Measurement Programs' Results, Costs, and Lessons Learned*, GAO-03-287 (Washington, D.C.: Jan. 29, 2003).

Our work addressed only standards and guidelines on data quality review. Although OMB's information quality guidelines and the Bureau's guidelines and performance principals cover all the key steps in data collection, analysis, and dissemination, we did not look at the Bureau's guidelines or standards for ensuring quality during the planning and data collection stages. Instead, as requested, we looked at Bureau guidance on steps taken after data collection, that is, guidance related to processing data, assessing their quality, and making them available to the public. We looked for documents spelling out standards, guidelines, procedures, and other criteria to guide decisions about identifying and correcting errors, determining if and when to release data, and revising data after release.

Our audit work was conducted in Washington, D.C., and at the Bureau's headquarters in Suitland, Maryland, from August 2003 through October 2004. Our work was done in accordance with generally accepted government auditing standards.

We requested comments on a draft of this report from the Secretary of Commerce. On September 27, 2004, the Secretary provided written comments on the draft. The comments are reprinted in appendix I.

Professional Judgment Drove Data Dissemination Decisions

The Bureau had no agencywide standards or guidelines in place to guide decisions about disseminating data from the 2000 Census. Instead of agencywide, written guidance, professionals within the different parts of the Bureau primarily used their judgment and program-specific practices to decide when and whether data should be released and what supporting information, if any, should accompany them. This led to instances when (1) data were released with uncorrected and undisclosed quality problems, (2) inconsistent decisions were made on whether to release data sets with similar quality problems, and (3) the reasons for certain data dissemination decisions were inadequately communicated.

The Bureau Lacked Agencywide, Written Standards and Guidelines on the Quality of Census Data Disseminated to the Public

At the time the Bureau was making decisions about disseminating data from the 2000 Census, it did not have written, agencywide guidelines or standards to help inform its decisions on whether the data were of sufficient quality to be released. Although Bureau officials emphasized that the Bureau has a long tradition of high standards and procedures that yield quality data, they acknowledged that these practices were primarily part of the agency's institutional knowledge. According to one official, key

individuals in each program area, relying primarily on professional judgment, determined whether the quality of the data was acceptable for release to the public. The official explained that the program areas develop their own guidance and procedures for ensuring data quality. Sometimes their guidance and procedures were written, but more often they were not. Further, the Bureau had no central inventory or repository of the guidance and practices of the different divisions.

Lack of Data Quality Review Guidelines Led to Inadequate Analysis of Potential Errors and Release of Data without Adequate Disclosure

As noted earlier, decennial census data are used to apportion and redistrict Congress. As release of data for each of these purposes is required by statute, they are known collectively as "public law" data. The Bureau had a number of quality assurance programs and procedures for assessing the accuracy of, and correcting errors in, public law and other data prior to their release. However, the lack of standard procedures and guidelines for dealing with quality problems contributed to lost opportunities to correct errors in the count of the population identified before the data were disseminated.

One such quality assurance program we reviewed was known as Demographic Full Count Review, in which analysts were to identify, investigate, and document suspected data discrepancies or "issues" in order to clear census data files and products for subsequent processing or public release.[15] The Bureau contracted out some of the analysts' work because it lacked sufficient staff to conduct the Full Count Review on its own. Bureau reviewers were to determine whether and how to correct the data by weighing quality improvements against time and budget constraints. Analysts identified 4,809 possible discrepancies, such as instances when the location, population count, demographic characteristics, or a combination of these for housing units and group living facilities differed from what analysts expected. According to Bureau officials, only 5 of the 4,809 issues were investigated and corrected prior to the release of the public law data. All five involved group living facilities the Bureau calls "group quarters" for which the Bureau had the correct population counts, but placed the living facilities in the wrong places. The Bureau did not investigate most of the remaining issues prior to the release of the data in large part because they were insufficiently documented and the Bureau lacked the time and people to further investigate these issues.

[15] GAO, *2000 Census: Refinements to Full Count Review Program Could Improve Future Data Quality*, GAO-02-562 (Washington, D.C.: July 3, 2002).

Subsequently, according to Bureau officials, the remaining issues that contained sufficient documentation were investigated as a part of the Count Question Resolution program, which ended in September 2003.

As we noted in our July 2002 report, the fact that public law data were released with over 4,800 unresolved data issues of unknown validity, magnitude, and impact is cause for concern. To the extent these unresolved discrepancies were in fact true errors in the population count or geography, they could have affected the drawing of congressional districts as well as other purposes for which census data are used.

The existence of data quality review guidelines could have helped the Bureau in this situation. For example, we found that the Bureau's lack of clearly defined requirements for documenting data issues resulted in a significant number of cases with inadequate documentation that the Bureau could not use to resolve the issues. Additionally, the Bureau had no mechanism for setting priorities for resolving these potential data errors. A sufficient set of guidelines could have helped the Bureau to ensure that the documentation of potential errors was adequate for decision making and to maximize the use of scarce resources in addressing the various data issues, giving top priority to investigating discrepancies likely to have the most adverse affect on the data.

The quality of certain data from the census long-form questionnaire have been called into question as well. In its 2004 comprehensive review of the 2000 Census, a panel of the National Research Council of the National Academy of Sciences assessed the quality of the long-form data using various benchmarks, and found that the overall quality of the information was less than that of the short-form questionnaire and had deteriorated since the 1990 Census.[16] For example, at least 32 percent of the respondents failed to provide information on their property taxes, and 30 percent did not respond to all or some of the questions relating to income (compared with 12 and 13 percent, respectively, in 1990). Additionally, the panel noted that the Bureau did not measure and report the impact of some of the steps it took to address problems with missing data and recommended that the Bureau develop such measures and inform users about the need for caution in analyzing and interpreting these data.

[16] National Research Council, *The 2000 Census: Counting Under Adversity* (Washington, D.C.: 2004).

Even more significant quality problems plagued the data for residents of group quarters. The panel found these data to be poor in comparison with the data for household residents, and also in comparison with data for group quarters from the 1990 Census. In 2000, missing data rates for some items were over 25 percent—one item was over 50 percent—for all residents of group quarters, and as high as 75 percent for prison inmates. Given the prevalence of missing data from residents of group quarters, the panel questioned whether the Bureau should have published these data at all for some or all types of group quarters.

The Bureau Made Conflicting Dissemination Decisions on Data with Similar Quality Problems

Our earlier work on Hispanic subgroups and the homeless showed that the Bureau's approach to data quality review led to inconsistent decision making. Faced with similar quality problems in data from the 2000 Census, Bureau officials made different decisions about disseminating data and did not explain the reasons for their decisions.

For example, in an effort to improve the count of Hispanics and simplify the questionnaire, the Bureau redesigned its 2000 Census question on Hispanic origin and dropped a list of examples of Hispanic subgroups included in the 1990 Census. In May 2001, the Bureau released data on Hispanics and Hispanic subgroups as part of its first release summarizing the results of the 2000 Census. The Bureau also published *The Hispanic Population*, a 2000 Census brief that provided an overview of the size and distribution of the Hispanic population in 2000 and highlighted changes in the population since the 1990 Census. For the first time, the Bureau released data on Hispanic subgroups as a part of its release of the Full Count Review data even though it had not fully tested the impact of questionnaire changes on the subgroup data and provided little discussion of the potential limitations of the data.

Shortly after the Hispanic and Hispanic subgroup data from the 2000 Census were released to the public, questions were raised about the counts for specific Hispanic subgroups. For example, the reported count of Dominican Hispanics was significantly lower than counts reported in other Bureau surveys. Representatives of affected Hispanic subgroups asked for an investigation and explanation of why the Bureau reported data that these subgroups considered to be of questionable quality. We found that a key factor behind the Bureau's release of apparently less-than-accurate Hispanic subgroup data appeared to be a lack of adequate guidelines

governing decisions on quality considerations that should be addressed before making data publicly available.[17]

In contrast, the Bureau, citing quality problems, decided not to separately report certain information on people without conventional housing, including those commonly referred to as "homeless." Enumerating this segment of the population has been an ongoing challenge for the Bureau. To help locate and count these people in 2000, the Bureau partnered with organizations providing services to the homeless and with local governments, some of which put considerable resources into the effort. When the Bureau decided not to separately report the number of people in transitional and emergency shelters as originally planned because of data quality problems, some of the organizations and local governments, which had expected to use the data for directing services to the homeless, questioned the Bureau's process for making that decision. Additionally, we found that the decision about when and whether to release data on people in emergency and transitional shelters changed several times. Decisions about the release of data with identified quality problems were not well documented and communicated with some Bureau partners and other stakeholders.[18]

As a result, outside parties interested in both the Hispanic and homeless data from the 2000 Census questioned the quality of the data, the procedures the Bureau used to determine what data to release, and the value of their own participation in helping the Bureau prepare for the 2000 Census. Because the Bureau's reasons for data release decisions were not obvious, and it had no guidelines or standards that spelled out criteria for decisions, the Bureau left itself open to questions about the objectivity of its decisions and risked loss of public confidence.

In our reports on Hispanic and homeless Census 2000 data, we recommended that the Bureau (1) develop agencywide guidelines for its decisions on the level of quality needed to release data to the public, how to characterize any limitations in the data, and when it is acceptable not to release data and (2) ensure that these guidelines are documented, transparent, clearly defined, and consistently applied. We also recommended that the Bureau ensure that its plans for releasing data are

[17] GAO-03-228.

[18] GAO-03-227.

clearly and consistently communicated to the public. The Bureau agreed with each of our recommendations and asked its Methodology and Standards Council[19] to review existing statistical and quality guidelines, bring them together in one place, and develop data quality standards. We discuss the Bureau's actions later in this report.

The Bureau Has Made Limited Progress in Publicly Issuing New Standards on the Quality of Data Disseminated to the Public since the 2000 Census

Since the first results of the 2000 Census were released, the Bureau has publicly issued a set of information quality guidelines and one new standard on the quality of data disseminated to the public. As required by the Information Quality Act and the OMB guidelines, the Department of Commerce[20] and the Bureau published *Information Quality Guidelines*, but the guidelines contain only general quality goals and principles and do not provide any specific guidelines or procedures on the implementation of the general principles. Also as required by the Information Quality Act and the OMB guidelines, the Bureau published a standard that described a procedure allowing individuals to seek correction of certain errors in data disseminated by the Bureau. Additionally, the Bureau has begun developing several other standards on the quality of data disseminated to the public, but none have been publicly released in final form.

In March 2003, in response to our recommendations, the Bureau established an interdirectorate working group charged with the broad mandate of developing Bureau-wide standards for quality in data releases. The working group has taken some steps to address the tasks laid out in its charter. However, the Bureau has not provided information on the scope or the time frame for developing these standards.

[19] The Bureau's Methodology and Standards Council sets standards for the Bureau's surveys and censuses. It is chaired by the Associate Director for Methodology and Standards and includes division chiefs from across the Bureau.

[20] Department of Commerce, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Disseminated Information*, 67 Fed. Reg. 62685 (Oct. 8, 2002). The Department of Commerce took a distributed approach, requiring its operating units (including the Bureau) to document and make available to the public their own information quality standards.

The Bureau Has Taken Steps to Expand Its Guidance on Data Quality Review

Recognizing the paucity of Bureau-wide written standards on the quality of data disseminated to the public, the Bureau established a Quality Program in 1999 to develop consistent processes for producing quality products across the Bureau. The Bureau's Associate Director for Methodology and Standards with input from chiefs in a number of divisions compiled an inventory of data quality review documents used in different divisions[21] and developed a Bureau-wide quality framework. The resulting quality framework was adopted to serve as a vehicle through which "the demographic, economic, and decennial areas can share and support common principles, standards, and guidelines." This framework provides the organization for documents in the intranet portal known as the Quality Management Repository (QMR). Additionally, the Bureau's description of the quality framework spells out the process for developing, reviewing, and approving quality framework documents. The document describing the quality framework and most of the documents in the QMR are internal documents not available to the public through the agency's Internet site. However, Bureau officials indicated that they intend to make some of the standards available through the Internet later in calendar year 2004.

The Bureau has publicly issued two data quality review documents and made them available through the Internet. In October 2002, in response to the requirements of the OMB guidelines, the Bureau published a set of information quality guidelines in eight performance areas, including the establishment of review procedures. The Bureau's guidelines identify broad quality goals and principles, but do not provide specific guidance to ensure consistent decisions. For example, the guideline on predissemination review of data says that "all documents released by the Census Bureau undergo extensive review that encompasses the content, statistical and survey methodology, and policy implications of the document," and that this review "ensures that the data and text of the document meet Census Bureau standards for quality" or the Bureau reserves the right to withhold the data from the public. However, the guideline does not indicate what the Bureau "standards for quality" are, how the Bureau will know if the data meet the standards, or who within the Bureau is responsible for the review.

[21] Some Bureau programs have their own data quality guidance and report extensively on the quality and limitations of the data. For example, for the 2003 Annual Social and Economic Supplement of the Current Population Survey, the Bureau provides information on the limitations, including a recommendation on using cells with a small number of respondents. See U.S. Census Bureau, *Current Population Survey, 2003 Public Use File Technical Documentation* (Washington, D.C.: 2003), G-7.

The second document issued and made available on the Bureau's Web site is *Census Bureau Standard: Correcting Information That Does Not Comply with Census Bureau Section 515 Information Quality Guidelines* in March 2002. This standard was also issued in response to the specific requirements of the Information Quality Act and the OMB guidelines that agencies provide procedures for correcting certain errors identified in data they disseminated and post these guidelines on their Web sites. The standard established procedures that allow individuals to request a correction of information they believe is erroneous and the Bureau to review the evidence and determine whether a correction is warranted.

The Bureau has also approved several additional Bureau-wide data quality review documents for implementation and internal distribution through the QMR on its intranet. On March 18, 2003, the Bureau issued *Census Bureau Standard: Minimal Information to Accompany Any Report of Census Bureau Data* for a 6-month trial period. The standard identifies 13 specific items that the Bureau should report for every survey or census and specifies who is responsible for ensuring adherence to the standard. An accompanying memorandum from the Associate Director for Methodology and Standards to program associate directors said that implementation issues would be documented during the trial period and appropriate changes made prior to the final release of the standard. Even though the trial period is over, the Bureau has not made such changes or publicly issued the standard in final form. However, the standard is still in effect on a trial basis, according to one Bureau official.

The Bureau also released its *Census Bureau Guideline: Quality Profiles* on March 9, 2004, through the QMR. The document outlines a standardized quality profile, recommended for all recurring surveys and certain other programs, which is intended to present a consistent set of information on the quality of each program. As a guideline rather than a standard, this guidance is recommended rather than mandatory.

In addition, the Bureau has also initiated work on several proposals for additional standards. For example, a standard for discussion and presentation of errors in data disseminated to the public is under development. This standard is based on a technical paper that was issued in 1974 and revised in 1987. The Bureau said it would be issued in the near future, but has not provided a specific date.

Bureau Working Group Has Begun Developing Additional Standards on Data Quality Review, but None Have Been Issued

In response to our recommendations from reports on both homeless and Hispanic subgroup data from the 2000 Census, the Bureau established an interdirectorate working group on March 3, 2003, with the broad mandate to develop Bureau-wide standards for quality in data releases. However, the working group has not yet issued any draft or final standards or developed a time frame for doing so.

The working group is composed primarily of assistant division chiefs from the program areas—decennial, demographic, and economic. An assistant division chief from the Demographic Statistical Methods Division chairs the group.

According to the working group's charter, its mission is to

- "Document current Census Bureau data review procedures,
- "Benchmark Census Bureau review procedures with that of other agencies,
- "Document Census Bureau situations where review of data indicates data does not meet "quality requirements" and the outcome of those situations,
- "Propose standards for quality in Census Bureau data products,
- "Benchmark quality requirements for data release with other agencies,
- "Develop Census Bureau Standard: Quality in Census Bureau Data Releases."

Bureau officials told us that the working group has reviewed the published detailed guidelines from NCES and the Canadian national statistical office. Benchmarking discussions have taken place with the Bureau of Labor Statistics and the National Center for Health Statistics. Additionally, the working group met with an official from the New Zealand national statistical office to discuss its standards. The group is also planning meetings with additional federal agencies. These organizations have published detailed guidance on how broad principles on data quality are to be put into practice, notably the organizational responsibilities and internal control mechanisms for applying them.

For example, Statistics Canada, the central statistical agency of the Canadian government, has developed an extensive and detailed set of quality guidelines that covers the quality of data disseminated to the public and the quality control processes that are supposed to be applied to ensure the quality of the data.[22] In March 2000, Statistics Canada published its *Policy on Informing Users of Data Quality and Methodology*, which specifies the organization's responsibilities to inform users about the concepts and methodology for collecting, processing, and analyzing its data; the accuracy of these data; and any other features that affect their quality or fitness for use. By detailing mandatory documentation standards, guidelines for additional documentation, and examples of mandatory standardized summary documentation, the policy enhances the likelihood of consistent decision making throughout the organization. Additionally, making this information public ensures that any data user can determine what has been done to ensure the quality of the data and Statistics Canada's reasons for its decisions about release.

NCES has developed detailed standards designed to implement its broader policies on dissemination of statistical data. An NCES standard includes a section entitled "Establishment of Review Procedures," which includes a table showing the required reviews for each type of product and an illustration of the key steps in the review and adjudication process. As with the Statistics Canada policy, the NCES standard provides information on the quality assessments and reviews that data must undergo before being released to the public.[23]

According to the Bureau's Associate Director for Methodology and Standards, the working group is making progress in conducting the work laid out in its charter. She said that the working group has reviewed different practices in divisions across the Bureau and benchmarked these practices against appropriate organizations. It has moved on to the task of identifying quality problems that have resulted from data quality review practices in different parts of the Bureau and assessing what could have been done differently. However, the Bureau did not provide any time frame for the working group's activities, information on how the Bureau intends

[22] Statistics Canada, *Statistics Canada Quality Guidelines*, 3rd ed. (Ottawa: October 1998), and *Statistics Canada's Quality Assurance Framework* (Ottawa: 2002).

[23] As noted earlier in this report, we did not evaluate the implementation or effectiveness of these guidelines and standards or their specific applicability to the Bureau.

to use the benchmarking exercises, or the intended scope and content of the Bureau-wide standard on quality in Bureau data releases.

The working group's charter indicates that its schedule should reflect an expeditious effort to complete its tasks. The Associate Director for Methodology and Standards, to whom the working group reports, emphasized that setting standards is a long-term process and pointed out that the Bureau has never issued a standard in less than a year. She noted that participation in the working group is added to the other responsibilities of its members and that initially the working group had no dedicated staff.[24] Additionally, she said that the working group does not have a time frame for completing these activities.

Greater Commitment to New Standards for Public Dissemination of Data Could Help Bureau Avoid Problems in Disseminating 2010 Census and Other Data

The standards that the Bureau has under development and activities of the working group are steps in the right direction. However, the Bureau has provided limited indication that developing and implementing standards on the quality of data it disseminates is a priority. It has no official plans for such an initiative, and these issues are not included in the Bureau's plan for the 2010 Census. Until spring 2004, no additional resources were provided to support the working group, and a year and a half after it began, the group has not developed any new standards or guidelines or indicated when it will be ready to do so. Although Bureau officials said that 2010 Census dissemination decisions would adhere to Bureau dissemination guidelines, the actions the Bureau has taken to date are not enough to ensure that it will avoid in 2010 the types of problems encountered in disseminating data from the 2000 Census. A publicly issued, comprehensive, Bureau-wide data quality framework, with interrelated standards, and specific procedures (as evident in NCES, ESS, and Statistics Canada) could help ensure consistency of decisions about the quality of data from the next decennial census and the conditions under which the data will be disseminated. The benefits the Bureau can achieve by implementing data quality review standards should not be limited to the decennial census. Because the standards could apply to all of the data publicly distributed by the Bureau, the sooner they are developed and implemented across the Bureau, the sooner the Bureau will begin to reap their benefits.

[24] Following our inquiries about staff support, the Bureau established a Quality Program Staff of three individuals in the spring of 2004 to support all of the working groups chartered by the Methodology and Standards Council.

Developing and Implementing Bureau-Wide Data Quality Review Standards Are Not Part of Official Bureau Plans

As noted above, the Bureau has not provided specific plans for further developing Bureau-wide data quality review standards or for implementing the broad data quality principles and guidelines outlined in its response to the OMB guidelines. It has not spelled out what needs to be done, how long it will take, what resources will be required, or how performance will be measured.

The Bureau's evolving plans for the 2010 Census devote little attention to data quality review issues. As it has for past decennial censuses, the Bureau focuses its plans for the 2010 Census on ensuring the quality of information collected during the data collection phase, rather than on how it will address potential quality problems that might be identified before the data are released. Bureau officials told us that whatever standards are developed will be applied to disseminating data from the 2010 Census. However, they said that the next decennial census is still a number of years away, and disseminating data from the 2010 Census is still farther in the future.

Data Quality Review Standards Could Also Aid Other Data Programs before 2010

The 2010 Census is to differ significantly from is predecessor. The 2010 Census, if implemented as planned, will ask the entire population to provide only basic information on the short form necessary for congressional apportionment. It will no longer collect more extensive information on a longer questionnaire from a sample of the population. Instead, the Bureau has developed the ACS that among other things, is intended to replace the long-form census questionnaire. The detailed data on social and economic conditions that were previously collected as a part of the decennial census will in the future be collected annually in the ACS. In fact, the ACS is a key component of the Bureau's plan for a reengineered 2010 Census. The ACS data are being collected and released annually for larger geographic areas, and data quality review standards could help improve the quality of these data immediately.[25]

The Bureau has developed several measures of quality for the information included in the ACS and began reporting these measures on its Web site in December 2003. These reported measures are important steps in the right direction for the Bureau, but these program-specific measures have not been adopted as Bureau-wide standards for similar collections. A Bureau

[25] See GAO-02-956R and GAO-05-82.

official said that these measures meet the requirements for minimum information on data quality of the Bureau's standard, which is being piloted. The measures developed for the ACS program are being reviewed for possible implementation in other household surveys.

Conclusions

Fully documented, transparent, clearly defined, and consistently applied standards on the quality of data disseminated to the public can help ensure that the Bureau makes consistent decisions about how it addresses data quality problems. Additionally, such standards can help the public understand the Bureau's reasons for its dissemination decisions, and can help protect the Bureau from allegations that it was inappropriately releasing or suppressing data. Because the cooperation and trust of the public is essential to a successful census, the Bureau must work to avert any loss of public confidence in the quality of data and in the integrity and objectivity of the Bureau.

Taken together, the quality problems that affected certain data from the 2000 Census underscore the importance of comprehensive data quality review guidelines for ensuring the Bureau makes more uniform decisions on data quality review and informs the public of limitations that could affect whether and how the data are employed.

The Bureau still has a long way to go in developing standards for the release of data to the public that will help avoid in the 2010 Census (and the ACS) the types of problems experienced in 2000. Additionally, since the standards would apply to all Bureau data collections, delay in their development and implementation means the Bureau is missing an opportunity for improving the quality of the other data it collects and disseminates. To avoid the problems it had with the dissemination of 2000 Census data the Bureau should place greater emphasis on developing and implementing data quality standards.

Although the Bureau has established a program for addressing standards development, we identified the following causes of concern.

- In the absence of more detailed information about the activities and schedule of the working group, it is difficult to assess the Bureau's progress in developing these standards. Over a year and a half after establishing the working group, the Bureau has publicly issued no new standards and has not publicly released plans that provide information on its schedule and agenda for developing the standards. Also the

Bureau has not publicly sought comments on the working group's initiatives through its advisory committees.

- Plans for the 2010 Census do not address procedures for dealing with data quality problems that are identified during the data quality review phase.

- The Bureau has not publicly announced any comprehensive plans for developing and implementing written, Bureau-wide quality standards and quality control processes.

A number of statistical agencies in the United States and elsewhere have developed comprehensive data quality review standards and quality control procedures that could serve as models for the Bureau. A Bureau-wide set of quality standards on data disseminated to the public covering both the quality of the data and quality control procedures would apply not only to the decennial census, but also to all other data collected by the Bureau and released to the public. Such standards could help the Bureau avoid some of the problems it experienced in disseminating data from the 2000 Census. Much of the data that were previously collected during the decennial census are now being collected under the ACS. Because these data are collected and released annually, the ACS, or other annual household surveys, could serve as a test for proposed standards.

Recommendations for Executive Action

To ensure that the 2010 Census, the ACS, and other Census data products will provide public data users with more complete, accurate, and useful information, we recommend that the Secretary of Commerce direct the Director of the U.S. Census Bureau to take the following two actions:

1. Accelerate the Bureau's effort to establish comprehensive data quality standards by developing and making public a detailed plan, including interim milestones, for developing such standards and procedures.

2. Include the implementation of the data quality review standards in the Bureau's plans for the 2010 Census, and test new draft guidelines on data quality review using the annual ACS test program and other surveys.

Agency Comments and Our Evaluation

The Secretary of Commerce provided us with written comments on a draft of this report on September 27, 2004, which are reprinted in appendix I. Commerce agreed with one of our two recommendations–namely, to establish data quality review standards as part of its plans for the 2010 Census, and as indicated in the Secretary's letter, the Bureau is taking steps to implement it. However, Commerce disagreed with our first recommendation that the Bureau accelerate its effort to establish comprehensive data quality standards. Commerce also identified some specific issues and suggested changes to provide additional context and clarification and in some cases technical corrections. We made these changes and corrections to the text as appropriate, but believe our first recommendation still applies.

Commerce took exception to our characterization of the amount of work that the Bureau has completed in developing comprehensive data quality review standards and in developing a specific standard for decisions on data release. However, the activities and documents Commerce cited to demonstrate the Bureau's progress were mentioned in our draft report. For example, Commerce noted that the Bureau had developed a quality framework for Bureau documents, inventoried quality guidance used in specific program areas, and created an in-house repository of such documents. Commerce also pointed to the quality principles the Bureau developed and included as a part of its *Information Quality Guidelines* issued in response to OMB requirements. Our draft report credited the Bureau with all of these activities, although not always at the same level of detail as Commerce described in its comments.

Moreover, while these are important steps, most of this work is not available to the public. As we observed in our draft report, the only documents that have been made public on the agency's Internet site are the documents required by the Information Quality Act and the related OMB guidelines: (1) the Bureau's *Information Quality Guidelines* and (2) the standard allowing individuals to seek correction of certain errors in data disseminated by the Bureau.

Indeed, our primary concern is not with how much work has been done but whether the Bureau has well-documented, transparent, clearly defined quality review guidelines and standards, and whether the pace of its efforts is sufficient. As yet, the Bureau has not produced such guidelines nor has it documented plans for completing this work. Bureau officials said they will

make existing standards available to the public on the agency's Internet site by the end of 2004, but have not indicated which standards will be included.

Therefore, we reaffirm our recommendation that the Bureau should accelerate its efforts to establish such data quality review standards by making public a detailed plan, including interim milestones, for developing such standards and procedures. Such a plan can assist the Bureau in prioritizing its work and addressing the resource constraints that will inevitably be present. If, as Commerce maintained, much of the work has already been completed, implementing the recommendation should not be unduly burdensome or time consuming. While we commend the Bureau for agreeing with our recommendation to implement data review guidelines and standards for the 2010 Census and the ACS, we believe it needs to accelerate its efforts to complete, make public, and fully implement these data review standards. The more time that elapses, the greater the risk of releasing data with quality problems.

As agreed with your offices, unless you release its contents earlier, we plan no further distribution of this report until 30 days from its issue date. At that time, we will send copies of this report to the Chairman of the House Committee on Government Reform, the Secretary of Commerce, and the Director of the U.S. Census Bureau. Copies will be made available to others on request. This report will also be available at no charge on GAO's home page at http://www.gao.gov.

Please contact me on (202) 512-6806 or by e-mail at daltonp@gao.gov if you have any questions. Other key contributors to this report were Robert Goldenkoff, Elizabeth Powell, Robert Parker, Michael Volpe, and Andrea Levine.

Patricia A. Dalton
Director
Strategic Issues

Comments from the Department of Commerce

THE SECRETARY OF COMMERCE
Washington, D.C. 20230

September 27, 2004

Ms. Patricia A. Dalton
Director, Strategic Issues
U.S. Government Accountability Office
Washington, DC 20548

Dear Ms. Dalton:

The U.S. Department of Commerce appreciates the opportunity to comment on the U.S. Government Accountability Office draft report entitled *Data Quality: Census Bureau Needs to Accelerate Efforts to Develop and Implement Data Quality Review Standards*. The Department's comments on this report are enclosed.

Sincerely,

Donald L. Evans

Enclosure

Comments from the U.S. Department of Commerce,
Regarding the U.S. Government Accountability Office
Draft Report Entitled *Data Quality: Census Bureau Needs to Accelerate Efforts to Develop and Implement Data Quality Review Standards*

The U.S. Department of Commerce thanks the Government Accountability Office (GAO) for the opportunity to review the draft report, *Data Quality: Census Bureau Needs to Accelerate Efforts to Develop and Implement Data Quality Review Standards* (GAO-04-469). This report discusses an important issue of concern to the Census Bureau—ensuring the quality of Census Bureau data releases.

The focus of the GAO report alternates between discussing what the Census Bureau is doing to establish comprehensive data quality standards for its products, develop data quality review standards, and develop guidelines for decisions on the level of quality needed to release data to the public. This report considerably expands the scope of two previous GAO reports on the quality of the Census Bureau data products that were issued in January 2003. These reports are *Methods for Collecting and Reporting Data on Homeless and Others Without Conventional Housing Need Refinement* and *Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement*. In both of these reports, the GAO recommendation in this area was directed at the more limited area of data quality review. The reports recommended that:

> "The Secretary of Commerce should direct the Bureau to . . . (2) develop guidelines for decisions on the level of quality needed to release data to the public, how to characterize any limitations, and when it is acceptable to suppress data; . . . "

The GAO approach in these three reports points to the multifaceted aspect of producing and releasing high-quality data to the public. Developing comprehensive standards on overall data quality is an important aspect of responding to this issue.

The Census Bureau has made substantial progress in developing comprehensive standards and developing a specific standard for decisions on data release. The significant amount of work that has been completed in this area is not reflected in the GAO report. Both of these efforts have the same objectives: to gain consensus within the Census Bureau on a specific standard, to document the standard, to implement the standard consistently across Census Bureau products, and to inform Census Bureau data users of the quality procedures that are incorporated into Census Bureau products.

The Census Bureau established a Quality Program in 1999 to develop consistent processes for producing quality products across the Census Bureau. This program encompassed building quality into Census Bureau processes, developing training on quality procedures for survey operations, documenting current practices that lead to quality products, developing best practices for survey procedures, and promoting communication on quality procedures. An initial effort produced a 30-year inventory of quality guidance issued by Census Bureau directorates.

1

The Census Bureau developed a Quality Framework for Census Bureau documents—Principles, Standards, Guidelines—modeled after the Statistics Canada Quality Guidelines. A portal was created as an in-house repository for the documents to provide accessibility to staff within the Census Bureau. Working groups were chartered to develop documents on an as-needed basis. In the spring of 2002, the Census Bureau developed a set of nine Quality Principles as a part of its Section 515 Information Quality Guidelines. These Quality Principles reference the Census Bureau Quality Framework and are consistent with those established by the federal statistical agencies in their *Federal Register* Notice (67 FR 38267-38470 – June 4, 2002), stating that the issuing agencies all had existing principles ensuring the quality of information disseminated to the general public. The Census Bureau's Section 515 Information Quality Guidelines provides a comprehensive statement of its approach to producing quality data. To date, the Census Bureau has had no information quality complaints.

These same nine principles are part of the Census Bureau Quality Framework. Beginning in 2001, documents were issued into the Census Bureau Quality Framework. Now there are nine Principles, seven Standards, eight Guidelines, and over one hundred Current Practices. The documents in the Quality Framework represent either updated documentation of a standard or guideline that was previously issued by some Census Bureau organization so that it relates to the entire Census Bureau (e.g. *Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses; Census Bureau Guideline: Quality Profiles*), or a standard or guideline for a new area of concern (e.g. *Census Bureau Standard: Minimal Information to Accompany any Report of Census Bureau Data*; *Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials*). The current totality of Census Bureau principles, standards, and guidelines addresses many, but not all, procedures that might be put in place to ensure quality in its products and processes. Components of these quality procedures relate to data review. Implementation of these Quality Framework documents will prevent many of the specific quality issues noted concerning Census 2000 data.

The Census Bureau directed its Methodology and Standards Council to address agency-wide issues of quality. Each Quality Framework standard has been developed by working groups chartered by the Methodology and Standards Council and composed of individuals within the organization who have expertise in the specific topic. All the affected operating units in the Census Bureau review the standard and comment on it, helping to assure both an appropriate and comprehensive standard and acceptance by the operating units. Once the Census Bureau Methodology and Standards Council approves the standard, the program associate directors are asked to concur. The program associate directors then take responsibility for implementing and enforcing the standard. The Quality Program Staff assists the program areas in implementing the standard through training and later evaluates the effectiveness of the implementation of the standard. The Census Bureau plans to make its existing standards available on its Internet site by the end of 2004.

A working group was chartered by the Methodology and Standards Council in March 2003 to address concerns raised in the two previous GAO reports relating to Census 2000 data releases. The charter directed the group to develop an agency-wide documented standard for quality in Census Bureau data releases as recommended in the earlier reports and is not limited exclusively to decennial data. The group has not been asked to develop comprehensive quality standards and guidelines as these are being developed within the Quality Framework described above. Rather, it is addressing the issues of the quality needed to release data, how limitations in data quality should be described, and when it is advisable to suppress data.

The group has made substantial progress thus far. It has drafted a report documenting current Census Bureau data review procedures; it has examined 12 situations where there was a concern with the quality of data planned for release and prepared a draft paper documenting those situations; it has benchmarked with two external agencies—Bureau of Labor Statistics and the National Center for Health Statistics—on data review procedures incidental to making data-release decisions. This is all preliminary and necessary work to both fulfill the charter of the working group and to propose standards to the Census Bureau Methodology and Standards Council for implementation in the Census Bureau. The working group is meeting weekly and making progress on the tasks in its charter. The ultimate task of establishing standards for data release decisions is extremely difficult and complex. The group, at this time, is actively developing specific standards for Census Bureau approval. This work will be completed for incorporation in the 2010 census.

The Census Bureau does not agree with Recommendation No. 1 to accelerate the effort to establish comprehensive data quality standards. Much of the work required has already been completed in the development of the existing principles, standards, and guidelines. The Census Bureau Methodology and Standards Council has now addressed all the issues identified in the 30-year inventory that it prepared in 2000, by either issuing an updated document in the Quality Framework or determining that the original document was no longer relevant. The Council will continue to identify new areas where a standard or guideline is needed and charter working groups to address those concerns, prioritizing the needs that come to their attention. There will always be resource constraints for developing a standard in a specific area, as there are limits to the number and availability of individuals with given expertise.

We do, however, agree with Recommendation No. 2 to establish data quality review standards as part of our plans for the 2010 census. The quality standards developed within the Quality Framework discussed above will be used as the quality review standards for the 2010 census; the results of the standards working group on decisions on quality for data release will become a part of that Quality Framework. To that end, the working group chartered will complete its efforts to develop a standard. Once that standard has been approved and issued, the Census Bureau Program Quality Staff will assist the program directorates in implementing the standard. The objective is that the standard will be applicable to all Census Bureau programs, including the 2010 census and the American Community Survey (ACS).

Thank you for your support of our efforts to ensure quality in our Census Bureau data products. If you have additional questions or concerns, please contact Alan Tupek, Acting Chair of the Methodology and Standards Council, at 301/763-4287.

Specific Issues

The Executive Summary, paragraph 2, and page 4, paragraph 1, states that the Census Bureau did not provide any specific guidelines or procedures on the implementation of its Section 515 Information Quality Guidelines. This is not correct. The Census Bureau developed a specific standard to address Section 515 information quality complaints—*Census Bureau Standard: Correcting Information that does not Comply with Census Bureau Section 515 Information Quality Guidelines* (issued 5/16/02).

Page 2, paragraph 2, The Census Bureau now has a standard that requires a discussion of limitations to the data it disseminates—*Census Bureau Standard: Minimal Information to Accompany any Report of Census Bureau Data.*

Page 5, paragraph 2 and page 27, paragraph 2. The report states that the Census Bureau has not provided additional resources to support the working group on data release. This is not correct. The Census Bureau established a Quality Program Staff (currently three individuals) in the spring of 2004. This staff supports all of the working groups that the Methodology and Standards Council charters to develop principles, standards, and guidelines; assists with the implementation of those documents; and works with program staff to incorporate quality into Census Bureau products and processes.

Page 15, paragraph 1, "The Bureau did not investigate most of the remaining issues in large part because they were insufficiently documented and the Bureau lacked the time and people to further investigate these issues."

Comment: The Census Bureau investigated the remaining issues containing sufficient documentation as part of the Count Question Resolution program, which was implemented between June 30, 2001 and September 30, 2003.

Page 16, paragraph 2, "In 2000, missing data rates reached as high as 50 percent for all group quarters residents, and as high as 75 percent for prison inmates."

Comment: The allocation rates for sample characteristics presented in the National Academy of Sciences report of the total group quarters population range from 18 percent (marital status) to 50 percent (wages last year).

Five of the 17 sample characteristics have allocation rates below 25 percent; 11 rates are between 25 percent and 49.9 percent, and 1 rate is over 50 percent.

The highest rates are found in the correctional category, where the sample characteristics have allocation rates ranging from 31 percent (marital status) to 75 percent (occupation last year).

Four of the 17 allocation rates are under 25 percent, two rates are between 25 percent and 49.9 percent, 10 rates are between 50 percent and 74.9 percent, and there is one allocation rate above 75 percent (occupation last year).

Page 18, paragraph 2, "Additionally, we found that the decision about when and whether to release data on people in emergency and transitional shelters changed several times. Decisions about the release of data with identified quality problems appeared to be judgment calls made by individuals in parts of the Bureau different from those involved with Bureau partners and other stakeholders."

Comment: The Census Bureau worked closely with the National Coalition for the Homeless and other advocacy groups in developing plans and procedures for counting and producing tabulations for the population experiencing homelessness.

The Census Bureau considered several options for tabulating data in Summary File 1 (SF-1) from the Service-Based Enumeration (SBE) operation. Originally, the Census Bureau planned to separately publish the total number of people enumerated at emergency and transitional shelters.

People tabulated at the other service locations, such as soup kitchens, regularly scheduled mobile food vans, and targeted non-sheltered outdoor locations, would be included in the "Other noninstitutional group quarters" category.

After examining the 1998 Dress Rehearsal results, as well as the preliminary Census 2000 data. The Census Bureau decided in January 2001 to include all people tabulated at the service locations during the SBE operation in one category, "Other noninstitutional group quarters," as part of its Summary File 1 (SF-1) release. This change was based on the Census Bureau's increasing concerns that the census tabulations of people enumerated at emergency and transitional shelters, without the provision of appropriate qualifiers and other limitations, would be misinterpreted. As a result, the Census Bureau decided to issue a special report on the results of the enumeration of emergency and transitional shelters with the appropriate caveats."

Page 21, The inventory and quality framework described was compiled by the Associate Director for Methodology and Standards, with input from the chiefs of the Computer-Assisted Survey Research Office, Decennial Statistics Studies Division, Demographic Statistical Methods Division (DMSD), Economic Statistical Methods and Programming Division, Planning, Research, and Evaluation Division, and Statistical Research Division.

Page 24, middle of page. The composition of the group stated in the text is incorrect. The group is composed of assistant division chiefs from the program areas—decennial, demographic, and economic. It is chaired by an assistant division chief in DSMD.

Page 25, The information concerning the working group on data release standards is incorrect. A correct statement follows:

> The working group has reviewed the published detailed guidelines from the National Center for Education Statistics and from Statistics Canada. Benchmarking discussions have taken place with the Bureau of Labor Statistics and the National Center for Health Statistics. Additionally, the working group met with an official from Statistics New Zealand to discuss its standards. The group is planning meetings with additional federal agencies.

Page 29, The measures developed for the ACS program are being reviewed for possible implementation in other household surveys.

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Susan Becker, Acting Manager, BeckerS@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

United States
Government Accountability Office
Washington, D.C. 20548-0001

Official Business
Penalty for Private Use $300

Address Service Requested

Presorted Standard
Postage & Fees Paid
GAO
Permit No. GI00