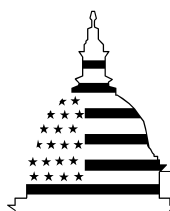


June 2005

DATA QUALITY

Improvements to Count Correction Efforts Could Produce More Accurate Census Data



G A O

Accountability ★ Integrity ★ Reliability



Highlights of [GAO-05-463](#), a report to congressional requesters

Why GAO Did This Study

The U.S. Census Bureau (Bureau) conducted the Count Question Resolution (CQR) program to correct errors in the count of housing units as well as dormitories and other group living facilities known as group quarters. GAO was asked to assess whether CQR was consistently implemented across the country, paying particular attention to whether the Bureau identified census errors that could have been caused by more systemic problems. GAO also evaluated how well the Bureau transitioned to CQR from an earlier quality assurance program called Full Count Review.

What GAO Recommends

GAO recommends that the Secretary of Commerce direct the Bureau to take such actions as consolidating CQR and Full Count Review into a single effort that systematically reviews and corrects any errors prior to the release of data for apportionment and redistricting; prioritizing the review of errors based on the magnitude of the problem; and ensuring the accuracy and accessibility of the revised data on its Web site. The Department of Commerce noted our report made several useful recommendations, but stated our approach was infeasible because of timing and other constraints. We believe our recommendations still apply because they could help the Bureau overcome these constraints and deliver better quality data.

www.gao.gov/cgi-bin/getrpt?GAO-05-463.

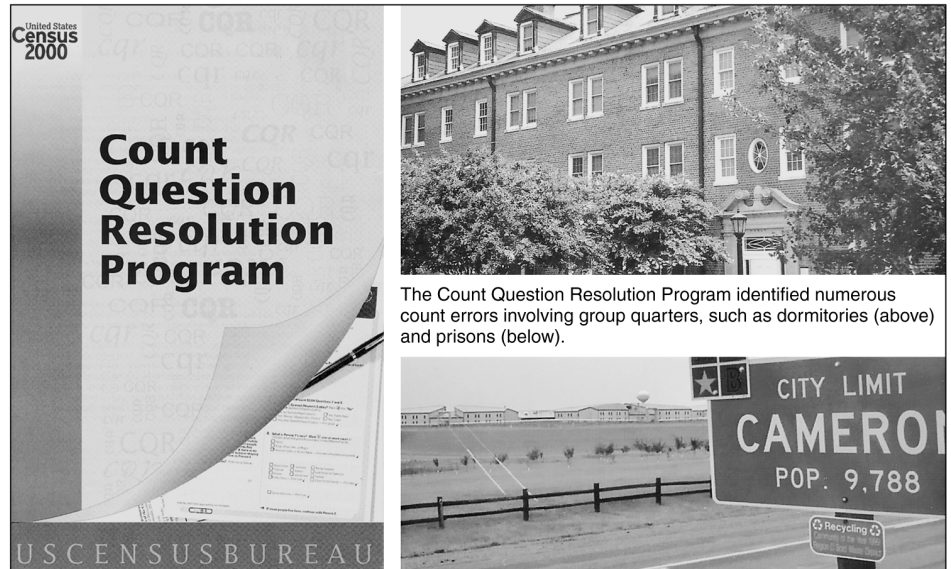
To view the full product, including the scope and methodology, click on the link above. For more information, contact Orice Williams at (202) 512-6806 or williamso@gao.gov.

DATA QUALITY

Improvements to Count Correction Efforts Could Produce More Accurate Census Data

What GAO Found

The CQR program, which ran from June 30, 2001, to September 30, 2003, played an important role in improving the quality of data from the 2000 Census in that it corrected numbers affecting 47 states and over 1,180 governmental units. Although this is a small percentage of the nation's more than 39,000 government entities, the count revisions impacted private homes, prisons, and other dwellings and, in some cases, were significant. For example, when the Bureau deleted duplicate data on students at the University of North Carolina at Chapel Hill and made other corrections, that state's head count dropped by 2,828 people. Similarly, CQR found that more than 1,600 people in Morehead, Kentucky, were counted in the wrong location.



Source: U.S. Census Bureau (left), and GAO (upper right and lower right).

GAO identified several shortcomings with the CQR program, including inconsistent implementation by the Bureau's regional offices and the posting of inaccurate data to the Bureau's Web-based errata report. Moreover, while CQR found the counting of group quarters to be particularly problematic, the Bureau did not perform an active, nationwide review of these known trouble spots, and thus missed an opportunity to potentially improve the accuracy of the data for these dwellings. Further, because CQR had more stringent documentation requirements compared to a preceding program called Full Count Review, CQR rejected hundreds of unresolved full count issues, missing another opportunity to improve the data. As its plans proceed for the 2010 Census, it will be important for the Bureau to address the operational issues GAO identified. Moreover, because the data for apportionment and redistricting were later found to be flawed for some jurisdictions, it will be important for the Bureau to develop a count correction program that is designed to systematically review and correct these essential figures.

Contents

Letter	1
Results in Brief	2
Background	4
Scope and Methodology	6
CQR Program Corrected Numerous Data Errors, but More Consistent Implementation and Other Improvements Are Needed	8
Better Strategic Planning and Other Actions Could Improve Future Count Correction Efforts	21
Conclusions	26
Recommendations for Executive Action	27
Agency Comments and Our Evaluation	28

Appendixes	
Appendix I: Change in State Populations As a Result of Count Question Resolution Program	31
Appendix II: Human Error and Other Factors Contributed to University of North Carolina Counting Errors	33
Appendix III: Comments from the Department of Commerce	35
Appendix IV: GAO Contact and Staff Acknowledgments	40

Figures	
Figure 1: Time Line Showing Relationship of CQR Program to Key Census 2000 Milestones	5
Figure 2: Map of Census Bureau’s 12 Regions	7
Figure 3: CQR Revisions Affected Numerous Governmental Units in Most States	10
Figure 4: Students in 26 UNC Dormitories Were Counted Twice in the Census	12
Figure 5: Prisoners in Cameron, Missouri, Were Mistakenly Omitted From the Town’s Population Count	13
Figure 6: CQR Table in the Bureau’s 2000 Notes and Errata Report Showing Faulty Links to Data	18
Figure 7: Initial Census Data on Bureau Web Site Do Not Inform Users That Some Numbers Have Been Revised	20

Contents

This is a work of the U.S. government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



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

June 20, 2005

The Honorable Wm. Lacy Clay
Ranking Minority Member
Subcommittee on Federalism and the Census
Committee on Government Reform
House of Representatives

The Honorable Carolyn B. Maloney
House of Representatives

Complete and accurate data from the decennial census are central to our democratic system of government. As required by the Constitution, census results are used to apportion seats in the House of Representatives. Census data are also used to redraw congressional districts, allocate billions of dollars in federal assistance to state and local governments, and for many other public and private sector purposes. Failure to deliver quality data could skew the equitable distribution of political power in our society, impair public and private decision making, and erode public confidence in the U.S. Census Bureau (Bureau).

To ensure it delivers accurate data, the Bureau employs a number of quality assurance programs throughout the course of the census. One such effort during the 2000 Census was the Count Question Resolution (CQR) program, which enabled state, local, and tribal governments to formally challenge the counts of housing units and “group quarters” (dormitories, prisons, and other group living facilities), and their associated populations. Bureau personnel could initiate a review of the counts as well.

Although the Bureau did not design CQR with the intention of incorporating any of the corrections that resulted from it into Census 2000 data products—including the numbers used for congressional apportionment and redistricting (figures commonly referred to as “public law data”)—governmental entities could use the updated information when applying for federal aid that uses census data as part of an allocation formula, as well as for other purposes. Because the count corrections could have political and financial implications for states and localities, it was important for the Bureau to carry out CQR consistent with its protocols. CQR began on June 30, 2001, and no new submissions were accepted after September 30, 2003.

This letter responds to your request to review the conduct of the CQR program. As agreed with your offices, we reviewed the results of the CQR program and assessed whether the program was consistently implemented across the country. In doing this, we paid particular attention to the extent to which the Bureau reviewed the census data for errors that could have been caused by broader, more systemic problems. We also evaluated how well the Bureau transitioned to CQR from an earlier quality assurance program called Full Count Review.

To meet these objectives, we reviewed relevant program documents and examined case files and conducted on-site inspections at four of the Bureau's regional offices where some of the largest CQR corrections took place. We also interviewed officials and staff responsible for administering the CQR program at the Bureau's headquarters and 12 regional offices. We did our audit work between February 2004 and March 2005 in accordance with generally accepted government auditing standards.

Results in Brief

The CQR program corrected data affecting over 1,180 of the nation's more than 39,000 governmental units including states, counties, and cities. Although the national and state-level revisions were relatively small, in some cases the corrections at the local level were substantial. For example, CQR increased Morehead, Kentucky's, population total by more than 1,600 people because the Bureau mistakenly attributed local university students, who lived in dormitories located within the city, to the population count of an unincorporated section of the county in which Morehead is located. Likewise, the Bureau added almost 1,500 persons to the population count of Cameron, Missouri, when CQR found that a prison's population was erroneously omitted.

That said, we also found critical aspects of the CQR program in need of improvement. For example, CQR was not consistently implemented by the Bureau's regional offices. Only the Bureau's Los Angeles Regional Office appeared to do any comprehensive, systematic research to identify possible count errors beyond those that were submitted by governmental units. Had it not been for Los Angeles' self-initiated review, several data errors—including instances where college dormitories were counted in the wrong geographic location—would have remained uncorrected because they were not identified by the affected jurisdiction.

In contrast, the Bureau's Charlotte Regional Office found that almost 2,700 students were counted twice at the University of North Carolina,

the discovery of which came about largely because two key census employees in Charlotte were alumni of the school and curious to see whether dormitories there were enumerated correctly. One factor behind the disparate execution of the CQR program seems to have been vague and sometimes inconsistent guidance and training that left staff in the regional offices with different understandings of whether they could conduct self-initiated research.

In addition, although the Bureau maintained an errata report on its Web site that listed the CQR revisions to the census data, our partial review of that information found several discrepancies between the updated figures, and what the numbers should have been. For example, the revised number of housing units for Sioux Falls, South Dakota, was almost 47,000 units too low. Likewise, the errata data on the total housing unit count for Burlington County, New Jersey, mistakenly excluded about 145,000 units. Moreover, embedded links on the Web site that were supposed to take users to revisions at lower levels of geography did not always work and produced error messages instead.

The CQR program was also poorly integrated with its predecessor program, Full Count Review. Although the Bureau planned to fold unresolved full count issues into CQR, the latter program had more rigorous documentation requirements. Consequently, hundreds of unresolved Full Count Review cases lacked CQR's necessary documentation, were rejected from CQR, and received no further review.

Overall, the CQR program was an important quality assurance tool, but the Bureau needs to address the operational issues we identified. Further, given the growing challenges to counting the nation's population, census errors are inevitable, and as the Bureau makes plans for the 2010 Census, it will be important for it to have a mechanism specifically designed to methodically review and correct errors in the public law data and subsequent data releases to the greatest extent possible. The lessons the Bureau has learned from CQR should provide valuable experience in developing such a program.

With that in mind, we recommend that the Secretary of Commerce direct the Bureau to improve its count correction efforts by taking such actions as: (1) consolidating Full Count Review and CQR into a single program that systematically reviews and corrects any errors prior to the release of the public law data; (2) expediting count correction efforts, in part, by using enumerators to help investigate data discrepancies while conducting their

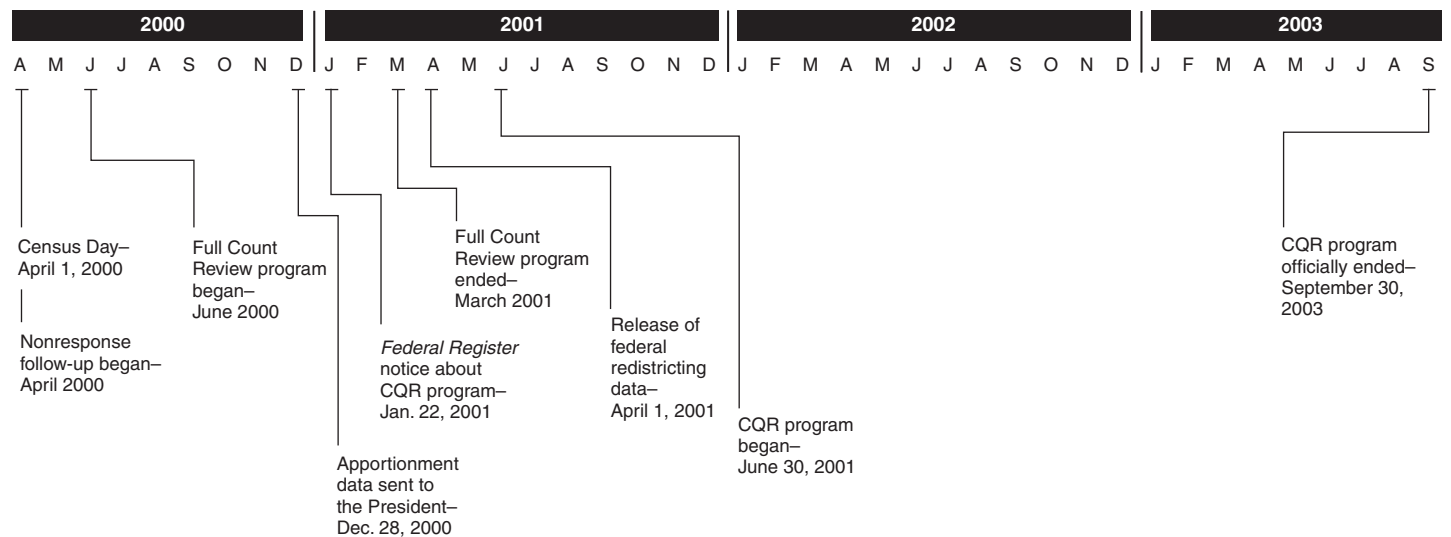
field work; (3) prioritizing the investigation of data challenges based on the magnitude of the suspected error; (4) ensuring the accuracy and accessibility of the revised data on its Web site; and (5) improving training and guidance provided to regional offices to help ensure count correction activities are consistently implemented.

The Acting Deputy Secretary of Commerce provided written comments on a draft of this report (see app. III). Commerce acknowledged that “the report provides a good overview of program results and makes several useful observations and recommendations,” and agreed with our finding that the process for conducting internal reviews was not consistently implemented. Nevertheless, Commerce took exception to our recommendations calling for the Bureau to design a count correction effort capable of identifying and correcting errors in the apportionment and redistricting data before that critical information is released. Commerce noted that such an approach was infeasible largely because of time and logistical constraints. Our report recognizes these challenges; further, the steps we recommend could help the Bureau overcome these very challenges and deliver more accurate public law data.

Background

The Bureau launched the CQR program on June 30, 2001, as the last in a series of quality assurance initiatives aimed at improving the accuracy of 2000 Census data (see fig. 1). Specifically, the CQR program provided a mechanism for state, local, and tribal governments to have the Bureau correct errors in certain types of census data. The Bureau referred to these challenges as “external cases.” Bureau personnel could also initiate reviews of suspected count errors, independent of these challenges, for further review. These were known as “internal cases.” Many of the internal cases were unresolved issues inherited from Full Count Review. Indeed, when the Full Count Review program began, the Bureau planned to fold unresolved issues from that program into CQR. The Bureau accepted no new submissions after the program officially ended on September 30, 2003, although it continued to review challenges submitted before the deadline and completed the final revisions in the summer of 2004.

Figure 1: Time Line Showing Relationship of CQR Program to Key Census 2000 Milestones



Source: GAO analysis of U.S. Census Bureau data.

Three types of corrections were permissible under the CQR program: (1) boundary corrections, where a jurisdictional boundary of a functioning governmental unit was in the wrong location; (2) geocoding corrections, where the Bureau placed the living quarters and their associated population in the wrong location; and (3) coverage corrections, where the Bureau properly enumerated specific living quarters and their corresponding population during the census but incorrectly added or deleted the information during data processing. Bureau officials were to research cases using existing Bureau data gathered during the 2000 Census; they could not conduct any new fieldwork to resolve count questions. The Bureau required governmental entities to accompany their challenges with specific documentation before it would investigate their claims.

Importantly, under the design of CQR, if a governmental unit had evidence that the Bureau missed housing units or group quarters that existed on Census Day 2000 (April 1), but the Bureau's records indicated that all of the Bureau's boundary information, geocoding, and processing were properly implemented, the Bureau would not change the data. Rather, the Bureau was to address this as part of the planning process for the 2010 Census.

If the CQR program corrected the population or housing unit counts of a particular entity, the Bureau was to issue revised, official figures for that

jurisdiction. The governmental unit could then use the updated numbers for future programs requiring 2000 Census data. CQR corrections were also used to modify annual post-censal estimates beginning December 2002 and were publicized on the Bureau's Census 2000 and American FactFinder Web sites (www.census.gov and www.factfinder.census.gov, respectively), as part of the 2000 Census notes and errata. However, CQR was not designed or publicized as a mechanism to correct the census results for purposes of apportionment and redistricting.

In compliance with legal requirements, the Bureau produced apportionment data by December 31, 2000, and redistricting data by April 1, 2001¹ (this information is known collectively as public law data). Although the law does not require that states use census data to redraw the boundaries of congressional districts, most states have always done so. Nothing would preclude the states from using the corrected data for redistricting. The general perception of the impartiality of the Bureau and the great cost and administrative effort required to take a census have been strong arguments in favor of using the Bureau's data.

Scope and Methodology

As agreed with your offices, we assessed whether the program was consistently implemented across the country, paying particular attention to the extent to which the Bureau reviewed the census data for errors that could have been caused by broader, more systemic problems, such as shortcomings with a particular census-taking procedure. We also evaluated how well the Bureau transitioned from an earlier quality assurance program used in the 2000 Census, Full Count Review.

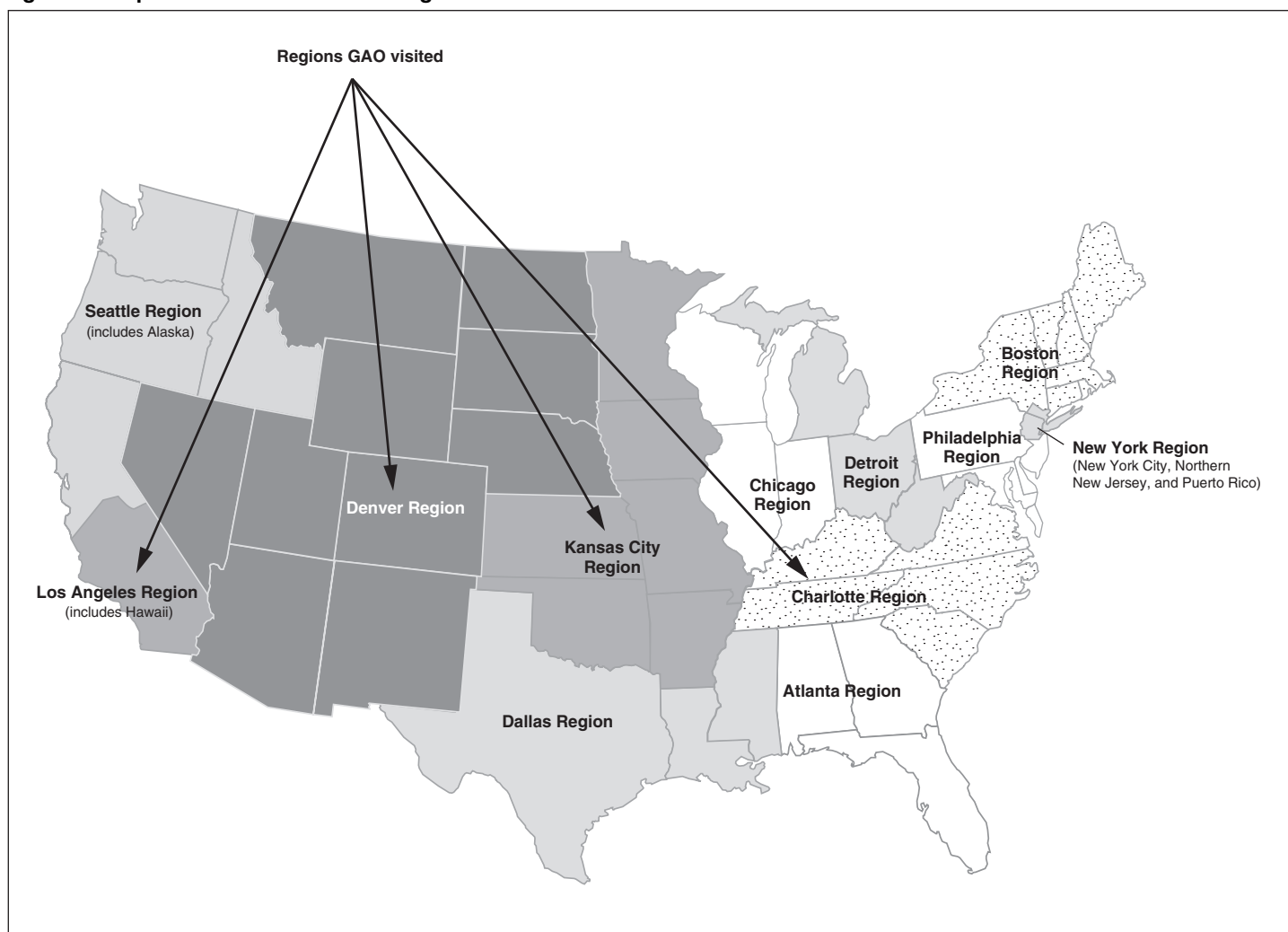
To assess the implementation of the CQR program, we obtained a headquarters perspective by reviewing program documents and case files at the Bureau's offices in Suitland, Maryland, as well as program results reported on the Bureau's Web site.² As part of this assessment, we reviewed the program's internal controls, especially those controls related to ensuring data quality. We also interviewed Bureau officials responsible for administering the program.

¹13 U.S.C. §§ 141(b) – (c).

²Based on our assessment of the data, we found the case file information and program results sufficiently reliable for our review.

To determine how CQR was implemented in the field, we visited 4 of the Bureau's 12 regional offices—Charlotte, North Carolina; Denver, Colorado; Kansas City, Missouri; and Los Angeles, California (see fig. 2).

Figure 2: Map of Census Bureau's 12 Regions



Source: GAO analysis of U.S. Census Bureau data.

We selected these regions because they included the six states and 10 governmental units within those states where the largest CQR count revisions occurred. We supplemented these cases by selecting an

additional seven states and 61 places within the four regions for further examination. The 61 localities were selected because they represented the full spectrum of CQR cases and were geographically diverse.

At each of the four regions, we reviewed regional case file information and interviewed Bureau personnel responsible for implementing CQR, such as program managers and geographers. We also made a site visit to at least one type of facility—including prisons, apartment buildings, and dormitories—in each region to understand firsthand the nature of the errors and the corrections made. To augment these four regional visits and obtain a more complete picture of how CQR was implemented, we used a structured telephone interview to elicit information from program officials at the Bureau's eight remaining regional offices that we did not visit in person.

To determine the extent to which the Bureau reviewed census data for systemic errors, and its procedures for folding unresolved cases from the Full Count Review program into CQR, we examined program manuals, memoranda, and other documents, and interviewed officials in the Bureau's headquarters and all of its regional offices. As part of this effort, we also analyzed the CQR case-tracking data in an attempt to determine the number of unresolved Full Count Review cases that were rolled into the CQR program. However, we were unable to do this because the tracking system did not contain information on which CQR cases originated as full count issues.

We requested comments on a draft of this report from the Secretary of Commerce. On May 20, 2005, we received the Acting Deputy Secretary's written comments and have reprinted them in appendix III; we address them in the Agency Comments and Our Evaluation section of this report.

CQR Program Corrected Numerous Data Errors, but More Consistent Implementation and Other Improvements Are Needed

Overall, the CQR program corrected data affecting over 1,180 of the nation's more than 39,000 governmental units. The revisions impacted a range of housing types including private homes with only a handful of residents, to college dormitories and prison cell blocks with populations in the thousands. At the same time, however, we identified several shortcomings with the CQR program, including inconsistent handling of internal cases by the Bureau's regional offices and inaccurate data being posted to the Bureau's public Web site. Moreover, while CQR found the counting of group quarters in their correct location—a problem known as geocoding error—to be particularly challenging, the Bureau did not

perform a nationwide review of these known trouble spots, and thus missed an opportunity to improve the accuracy of the data for these dwellings.

**CQR Program Corrected
Errors in Hundreds of
Governmental Units**

Nationwide, the CQR program corrected count errors involving governmental units in 47 states, Puerto Rico, and the District of Columbia (see fig. 3).³ Three states—Maine, New Hampshire, and Rhode Island—had no CQR corrections.

³The Bureau also made corrections to governmental units classified as American Indian/Alaska Native Areas.

Figure 3: CQR Revisions Affected Numerous Governmental Units in Most States



Source: GAO analysis of U.S. Census Bureau data.

^aThe Bureau made count changes in Hawaii and the District of Columbia, but these revisions were made at the census block level and did not change the state's governmental unit counts.

The corrections affected over 1,180 governmental units in the United States. Although this is a small percentage of the nation's more than 39,000 governmental units, the impact of those changes on local governments was, in some cases, substantial, and could have implications for federal assistance and state funding programs that use census numbers in their allocation formulas, as well as other applications of census data.

For example, officials in one Kentucky county challenged the geocoding of a housing unit located near new precinct and congressional district boundaries. They told the Bureau that the new boundaries split the county, and they were concerned that the geocoding error would affect where the housing unit's few occupants registered to vote. Because the housing unit was improperly geocoded, the Bureau corrected the data.

With respect to fiscal effects, the Controller of the State of California uses population figures as the basis for refunding a portion of state taxpayer fees—including automobile licensing fees—to cities and counties. Because of an error in the 2000 Census, Soledad, California, officials estimated it lost more than \$140,000 in state refunds when over 11,000 residents were incorrectly counted in two nearby cities' populations, according to city and state officials. Although CQR eventually corrected the error, Soledad did not recover the funds that went to the other cities.

Other examples of large CQR corrections include the following (See app. I for a complete list of state-level population changes):⁴

- North Carolina's population count was reduced by 2,828 people, largely because the Bureau had to delete duplicate data on almost 2,700 students in 26 dormitories (see fig. 4) at the University of North Carolina (UNC) at Chapel Hill. The erroneous enumerations occurred, in large part, because of mistakes that occurred in various preparatory activities leading up to the 2000 Census. (See app. II for a more detailed discussion of this incident.)

⁴Because a population increase in one government entity was typically offset by a loss in population in a neighboring entity (or vice-versa), there was generally little net change in population counts at the national and state levels as a result of the CQR program and no effect on apportionment.

Figure 4: Students in 26 UNC Dormitories Were Counted Twice in the Census



Source: GAO.

- The population count of Morehead, Kentucky, increased by more than 1,600 when CQR found that a large number of students from Morehead State University's dormitories were erroneously excluded from the city's population. During the 2000 Census, the Bureau had incorrectly identified the dormitories as being outside city limits and in an unincorporated area of Rowan County.
- The population count of Cameron, Missouri, was off by nearly 1,500 people when the Bureau found that the prison population of the state's Crossroads Correctional Center was inadvertently omitted from the town's headcount (see fig. 5). The correction to the town's population accounted for the entire 1,472 person increase in Missouri's total population under the CQR program.

Figure 5: Prisoners in Cameron, Missouri, Were Mistakenly Omitted From the Town's Population Count



Source: GAO.

- The population of the city of Waseca, Minnesota, increased by more than 1,100. The 2000 Census had mistakenly included the prison population for the Waseca Federal Correctional Institute in two surrounding townships in Waseca County. The CQR program resulted in the population being shifted to the city.
- The population of Colorado increased by more than 750 in large part because a processing error in counting housing units in Grand Junction initially excluded almost 700 people from the city's population total.
- The population of Denver and Arapahoe Counties in Colorado shifted by more than 900 because the Bureau had incorrectly assigned the location of two apartment complexes. As a result of CQR, the apartment complexes were incorrectly identified and counted as being in Denver but under CQR were later found to be in adjoining Arapahoe County.

CQR Program Was Unevenly Executed

The Bureau's 12 regional offices did not always adhere to the same set of procedures when developing internal cases, and this, in turn, produced uneven results. Importantly, the procedures used to execute public programs need to be well documented, transparent, and consistently applied in order to ensure fairness, accountability, defensible decisions, and reliable outcomes. To do otherwise could raise equity questions.

One variation in the way internal cases were handled was evident at the Bureau's Los Angeles Regional Office, which appeared to be the only region to do comprehensive methodical research to actively identify possible count errors beyond those that were submitted by governmental entities. According to the office's senior geographer, the geography staff developed a structured approach to systematically examine census data from all the prisons and colleges within the office's jurisdiction, because the data on both types of group quarters were known to be problematic. He added that the more problems they found, the more they were motivated to keep digging. The geographer noted that the in-depth review was possible because the Los Angeles region covers only the southern half of California and the state of Hawaii, and thus has fewer governmental units compared to the Bureau's other regional offices.

Had it not been for the Los Angeles region's self-initiated and systematic review, certain data errors would have gone uncorrected because they were not identified by the affected jurisdiction. For example, regional staff found instances where college dormitories were counted in the wrong geographic location, which, in turn, affected the population counts of their surrounding locales. Such was the case with California State University Monterey Bay (CSUMB) and the University of California at Santa Barbara (UCSB). As a result, the Bureau transferred a population of more than 1,400 between the towns in which they were initially counted and in which CSUMB is located and shifted a population of more than 2,700 between the city and the unincorporated area of the county in which UCSB is located.

The Bureau's Charlotte Region, while also more active than the Bureau's other offices that generated internal CQR cases, seemed to be less methodical and comprehensive than Los Angeles in its approach. For example, although Charlotte geographers detected the duplicate count of almost 2,700 students at the University of North Carolina mentioned in the previous section, their research was not the result of any systematic review. Rather, it came about largely because of the curiosity of key employees in the Charlotte office, who were also alumni of the school. (See app. II for more details on the circumstances surrounding the duplicate count.)

Better Guidance and Training Could Improve Implementation

Vague guidance was one reason for the disparate handling of internal cases. For example, the CQR procedural manual indicates that the Bureau's 12 regional offices were to research CQR cases "as appropriate." However, the manual did not define whether this meant that the regional offices

should initiate their own data reviews or merely verify CQR cases submitted by governmental units. More generally, numerous geographers we interviewed—the primary users of the manual—did not find it user-friendly, noting it was confusing, complex, or impractical. For example, a geographer pointed out that the manual did not have an index covering the eight chapters and 26 appendixes, which would have helped them more quickly find information and procedures. In addition, we found that the manual and other documents did not discuss how program staff were to address Full Count Review issues.

The Bureau's training was also problematic and likely added to the implementation disparities. For example, geographers in five regions told us that during training they were instructed or given the impression they were not to generate additional internal cases beyond the small number of count errors that had already been identified at the beginning of the program. Also, geographers in two of these regions told us they were specifically told not to investigate any count errors they found that were outside the scope of the cases that governmental units submitted. Conversely, geographers in the other seven regions said they were not restricted in any way.

There were other training issues as well. According to the Bureau's draft CQR program assessment, the final version of which is pending, some training materials were developed at the last minute and were never finalized, and training began before needed software was in place at all the research divisions. Proper training was particularly important because, as the draft evaluation notes, staff assigned to the CQR program had census experience but limited geographic and field operations knowledge. Others had limited or no Census 2000 software program experience.

Internal Control and Quality Assurance Problems Led the Bureau to Report Erroneous Data

Federal internal control standards call on agencies to employ edit checks and other procedures to ensure their information processing systems produce accurate and complete data.⁵ However, the Bureau's internal controls in reporting CQR results were insufficient in that we found, after a partial review, a number of instances where the Bureau disseminated inaccurate data on its Web site where it maintains an errata report that lists the CQR revisions to the 2000 Census data.

⁵GAO, *Internal Control Management and Evaluation Tool*, [GAO-01-1008G](#) (Washington, D.C.: Aug. 1, 2001).

Specifically, after comparing data from the errata report to the certified numbers in the CQR case files, we found errors with the reporting of CQR housing, group quarters, and population counts. Importantly, our review found that the revised, certified figures the Bureau provided to affected jurisdictions were correct. This is significant because the affected jurisdictions could use these updated numbers for revenue sharing and other programs that require census data. However, users who obtain information from the Bureau's errata report—these can be people in academia, government, and the private sector—would not have the most up-to-date information. For example:

- The original state-level total housing unit count for Delaware mistakenly excluded 30,000 housing units.
- The revised housing unit count for the Minnehaha County portion of Sioux Falls, South Dakota, was underreported by almost 47,000 housing units.
- The Burlington County, New Jersey, revised total housing count mistakenly excluded about 145,000 units.
- The errata report excluded 8 of the 12 American Indian and Alaska Native Areas that had revisions to their housing, group quarter, or total population counts.

The Bureau later corrected these errors after we brought them to its attention.

Although the Bureau had controls in place to ensure accurate research and reporting, the problems we found point to the need for the Bureau to tighten its procedures to prevent mistakes from slipping into its data products. For example, the CQR manual included some quality control steps, such as having headquarters divisions review field research and results. Further, field geographers told us they consulted one another about questions or procedures and checked each other's work, and Bureau program managers had procedures in place to review final revisions and certify them as correct.

Documents in the CQR case files we reviewed substantiate these practices. Also, managers told us they randomly checked data entered into the files that are the basis for the revisions posted to the errata report. Still, the number of errors we found after only a partial review of the errata files

raises questions about how effectively the Bureau implemented the quality assurance procedures, as well as the quality of the data we did not review. It also underscores the importance of adequate control activities to prevent these problems from recurring.

Web-based Errata Report Should Be More User- Friendly

Data users may have encountered problems trying to access certain information from the Web-based errata report. Additionally, because there was no link or cross-walk between some of the initial population data the Bureau released and the CQR revisions, users may have been unaware that some of the original numbers had been revised.

As shown in figure 6, the Bureau's Web-based errata report presented revised data for states, American Indian and Alaska Native Areas, and other jurisdictions at the state or similar geographic level. Although the table had embedded links that were supposed to take users to revisions at lower levels of geography, these links did not always work and produced error messages instead. We found that unless the users' software and Internet access paralleled the Bureau's, users could not access the more detailed data using the embedded links. Bureau staff involved with posting the data to the Web site stated in the summer of 2004 that they were aware of the problem, but as late as March 2005, the problem had yet to be fixed.

Figure 6: CQR Table in the Bureau's 2000 Notes and Errata Report Showing Faulty Links to Data

Corrected Census 2000 Total Population, Group Quarters Population, Total Housing Unit, and Vacant Housing Unit Counts for the United States and Puerto Rico

Note: Corrected counts are a result of the Count Question Resolution (CQR) Program.

Federal Register/Vol. 66, No. 130/Friday, July 6, 2001 (Summary): The CQR program is not a mechanism or process to challenge the March 6, 2001, decision of the Secretary of Commerce to release unadjusted numbers from Census 2000 for redistricting purposes; nor is it a mechanism or process to challenge or revise the numbers sent to the President on December 28, 2000, to be used to apportion the U.S. House of Representatives.

This table presents census counts only when there is a CQR change that affects the state or Puerto Rico level. States with no CQR changes (and no tables) are Maine, New Hampshire, and Rhode Island. Use GU link to get to governmental units pages, or T/B link to get to tract/block listings.

Last updated 1/12/2005

Number of governmental units affected by CQR = 1183

United States State and Puerto Rico	State FIPS	2000 Census Counts							
		Tabulation (Original)				Corrected (Revised)			
		Total Population	Group Quarters Population	Total Housing Units	Vacant Housing Units	Total Population	Group Quarters Population	Total Housing Units	Vacant Housing Units
United States		281 421 906	7 778 633	115 904 641	10 424 540	281 424 603	7 780 475	115 904 472	10 424 715
STATE									
Alabama	GU T/B 01	4 447 100	114 720	1 963 711	226 631	4 447 351	114 720	1 963 834	226 662
Alaska	GU T/B 02	626 932	19 349	260 978	39 378	626 931	19 349	260 963	39 364
Arizona	GU T/B 04								
Arkansas	GU T/B 05								
California	GU T/B 06	33 871 648	819 754	12 214 549	711 679	33 871 653	819 757	12 214 550	711 679
Colorado	GU T/B 08	4 301 261	102 955	1 808 037	149 799	4 302 015	102 955	1 808 358	149 819
Connecticut	GU T/B 09	3 405 565	107 939	1 385 975	84 305	3 405 602	107 939	1 385 997	84 312
Delaware	GU T/B 10								
District of Columbia	GU T/B 11								
Florida	GU T/B 12	15 982 378	388 945	7 302 947	965 018	15 982 824	388 945	7 303 108	965 033
Georgia	GU T/B 13	8 186 453	233 822	3 281 737	275 368	8 186 816	233 822	3 281 866	275 381
Hawaii	GU T/B 15								
Idaho	GU T/B 16	1 293 953	31 496	527 824	58 179	1 293 956	31 496	527 825	58 179
Illinois	GU T/B 17	12 419 293	321 781	4 885 615	293 836	12 419 647	321 773	4 885 744	293 845
Indiana	GU T/B 18	6 080 485	178 154	2 532 319	196 013	6 080 517	178 154	2 532 327	196 013
Iowa	GU T/B 19	2 926 324	104 169	1 232 511	83 235	2 926 382	104 169	1 232 530	83 237
Kansas	GU T/B 20	2 688 418	81 950	1 131 200	93 309	2 688 824	81 950	1 131 395	93 332
Kentucky	GU T/B 21	4 041 769	114 804	1 750 927	160 280	4 042 285	114 804	1 751 118	160 285
	GU T/B 22	4 468 976	135 965	1 847 181	191 128	4 468 958	135 965	1 847 174	191 129
	GU T/B 23								
	GU T/B 24	5 296 486	134 056	2 145 283	164 424	5 296 507	134 056	2 145 290	164 424
	GU T/B 25	6 349 097	221 216	2 621 989	178 409	6 349 105	221 216	2 621 993	178 410
	GU T/B 26	9 938 444	249 889	4 234 279	448 618	9 938 480	249 889	4 234 252	448 625
	GU T/B 27	4 919 479	135 863	2 065 946	170 819	4 919 492	135 862	2 065 952	170 819
	GU T/B 28	2 844 558	95 414	1 161 953	115 519	2 844 556	95 414	1 161 952	115 519
	GU T/B 29	5 595 211	162 058	2 442 017	247 423	5 596 683	163 534	2 442 003	247 425
	GU T/B 30								
Nebraska	GU T/B 31	1 711 263	50 818	722 668	56 484	1 711 265	50 818	722 669	56 484
Nevada	GU T/B 32								
New Hampshire	GU T/B 33								
New Jersey	GU T/B 34	8 414 350	194 821	3 310 275	245 630	8 414 347	194 821	3 310 274	245 630
New Mexico	GU T/B 35								
New York	GU T/B 36	18 976 457	580 461	7 679 307	622 447	18 976 821	580 833	7 679 307	622 447
North Carolina	GU T/B 37	8 049 313	253 861	3 523 944	391 931	8 046 485	253 881	3 522 330	391 959
North Dakota	GU T/B 38	642 200	23 631	289 677	32 525	642 200	23 631	289 678	32 526
Ohio	GU T/B 39	11 353 140	299 121	4 783 051	337 278	11 353 145	299 121	4 783 066	337 279
Oklahoma	GU T/B 40	3 450 654	112 375	1 514 400	172 107	3 450 652	112 375	1 514 399	172 107
Oregon	GU T/B 41	3 421 399	77 491	1 452 709	118 986	3 421 436	77 491	1 452 724	118 986
Pennsylvania	GU T/B 42	12 281 054	433 301	5 249 750	472 747	12 281 054	433 301	5 249 751	472 748
Rhode Island	GU T/B 44								
South Carolina	GU T/B 45	4 012 012	135 037	1 753 670	219 816	4 011 816	135 037	1 753 586	219 817
South Dakota	GU T/B 46								
Tennessee	GU T/B 47	5 689 283	147 946	2 439 443	206 538	5 689 267	147 946	2 439 435	206 537
Texas	GU T/B 48	20 851 820	561 109	8 157 575	764 221	20 851 790	561 109	8 157 557	764 221
Utah	GU T/B 49	2 233 169	40 480	768 594	67 313	2 233 198	40 480	768 603	67 313
Vermont	GU T/B 50								
Virginia	GU T/B 51	7 078 515	231 398	2 904 192	205 019	7 079 030	231 398	2 904 432	205 040
Washington	GU T/B 53	5 894 121	136 382	2 451 075	179 677	5 894 141	136 382	2 451 081	179 677
West Virginia	GU T/B 54	1 808 344	43 147	844 623	108 142	1 808 350	43 147	844 626	108 142
Wisconsin	GU T/B 55	5 363 675	155 958	2 321 144	236 600	5 363 715	155 958	2 321 157	236 601
Wyoming	GU T/B 56								
Puerto Rico	GU T/B 72	3 808 610	46 774	1 418 476	157 151	3 808 603	46 774	1 418 474	157 151
American Indian/Alaska Native Areas	T/B								

- represents zero

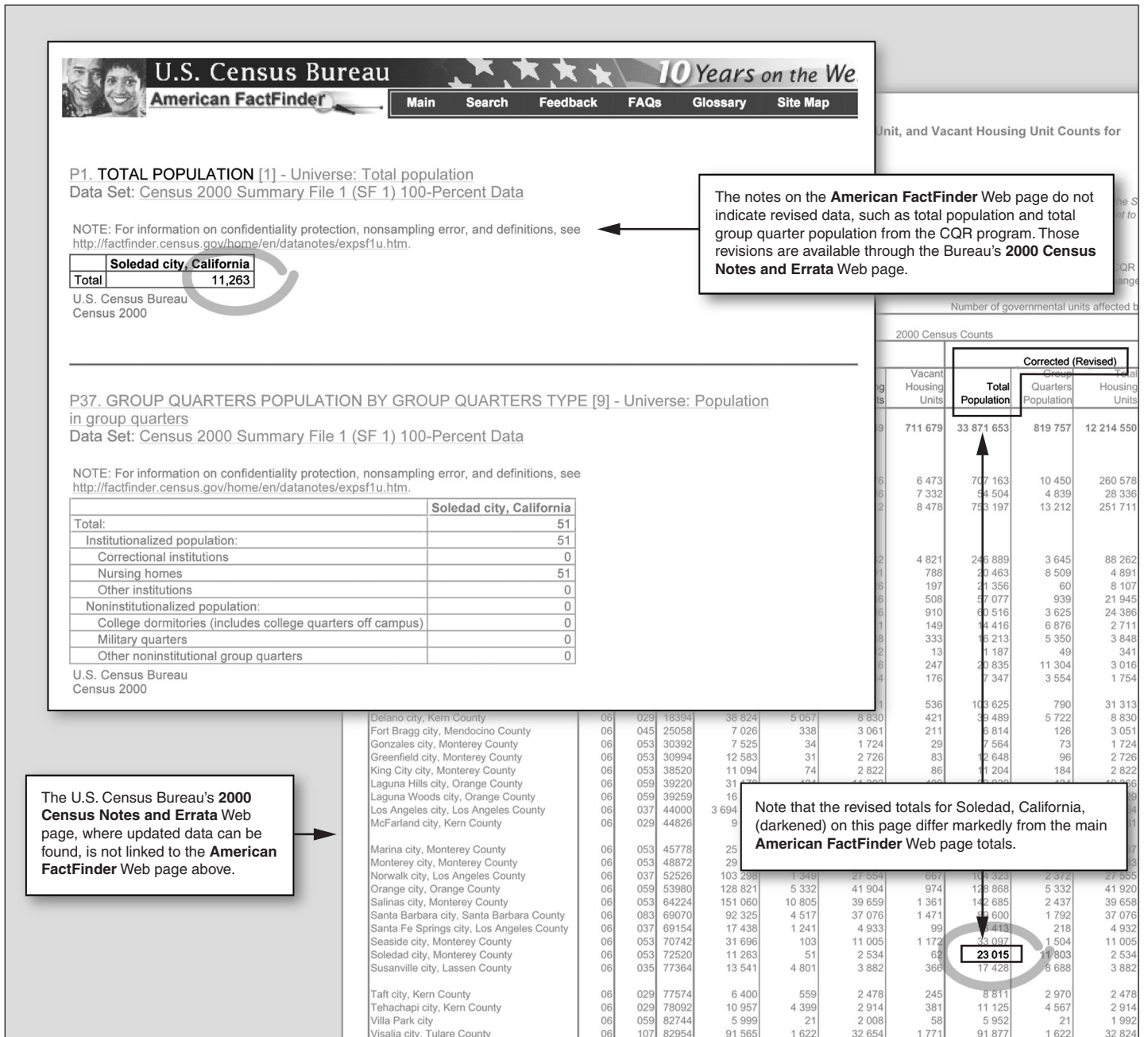
Source: GAO analysis of U.S. Census Bureau information.

At the same time, the CQR revisions may not be evident to users who access certain data from the 2000 Census data posted elsewhere on the Bureau's Web site. This is because these sites lack notes or flags informing users that updated figures are available in the census errata report.⁶ For example, the Bureau's American FactFinder Web site—the Bureau's primary electronic source of census data—does not inform users that revised data on group quarter counts, including the number of correctional institutions, as well as data on their associated populations, exist as part of the Bureau's notes and errata report.

American FactFinder presents data known as Summary File 1 (SF-1), which is the first data set the Bureau produces from the census, and is used for purposes of apportionment and redistricting. While the SF-1 data remain unchanged, other data users may find the revised numbers better suited to their needs. Figure 7 illustrates the existence of two sets of numbers without any explanation. Summary File data from American FactFinder show the population for Soledad, California, as 11,263. However, the Bureau's errata report, which reflects the CQR revisions, shows the Soledad population at 23,015. Because American FactFinder lacks notes or links that tell users about the revised data, users might inadvertently obtain erroneous information.

⁶The Bureau and American FactFinder home pages do not list or provide direct links to the 2000 Census notes and errata report or the CQR program Web site. However, the Bureau's Census 2000 Gateway Web site provides links to both of them and its American FactFinder Web site provides an indirect link to the notes and errata through that site's data sets.

Figure 7: Initial Census Data on Bureau Web Site Do Not Inform Users That Some Numbers Have Been Revised



According to Bureau officials, while they thought about adding notes directing users to the CQR revisions, they decided against it because they thought it would confuse more people than it would help. They reasoned that knowledgeable users, such as county planners and state data center staff, are likely aware of the CQR information and would therefore not need to be informed about the existence of the notes and errata Web site.

CQR Errors Highlight Problems with 2000 Census Address List Development Procedures

The errors uncovered by the CQR program highlight some of the limitations in the way in which the Bureau builds its address list for the decennial census, particularly in the procedures used to identify and locate group quarters. For the 2000 Census, the Bureau had three operations that were primarily designed to locate these types of dwellings. However, given the number of prisons and other group quarters geocoded to the wrong location, refinements are needed. Moreover, the Bureau's draft CQR program assessment found that the Bureau's Master Address File had numerous data entry errors including incorrect spellings, geocoding, and zip codes.

To its credit, the Bureau is planning several improvements for 2010, including integrating its housing unit and group quarter address lists. This could help prevent the type of duplicate counting that occurred at UNC when the same dormitories appeared on both lists. Likewise, the Bureau's planned use of a satellite-based navigational system could help census workers more precisely locate street addresses.

Better Strategic Planning and Other Actions Could Improve Future Count Correction Efforts

The CQR program was preceded by a quality assurance program called Full Count Review, which ran from June 2000 through March 2001 and, like CQR, was designed to find problems with the census data. However, although the Bureau planned to fold unresolved full count issues into CQR, many full count issues were rejected from CQR because the latter program had more stringent documentation requirements. As a result, the Bureau was unable to resolve hundreds of additional data issues.

Numerous Unresolved Full Count Issues Could Not Be Folded into CQR as Planned

Under the Full Count Review program, analysts were to identify data discrepancies to clear census data files and products for subsequent processing or public release. Analysts did so by checking the data for their overall reasonableness, as well as for their consistency with historical and demographic census data and other census data products. The types of

issues flagged during Full Count Review included potential discrepancies involving the counts and/or locations of group quarters, housing units, and individual households, among others.

As we noted in our July 2002 report, Full Count Review identified 4,809 potential data anomalies.⁷ However, of these, just five were corrected prior to the December 31, 2000, release of apportionment data and the April 1, 2001, release of redistricting data. The corrections included a military base, a federal medical center, and multiple facilities at two prisons and a college that were counted in the wrong locations. That the public law data were released with numerous data issues of unknown validity, magnitude, and impact, gave us cause for concern, and we noted that the Bureau missed an opportunity to verify and possibly improve the quality of the information.

When the Full Count Review program began, the Bureau planned to fold unresolved issues from that program into CQR. Indeed, according to a June 2000 memo on CQR policy agreements, “a by-product of [Full Count Review] is documentation of unresolved issues for potential use in CQR.” However, because the CQR program had more rigorous documentation requirements before it would accept a case compared to Full Count Review, a number of issues that were deemed suitable for Full Count Review but were unresolved, were rejected from CQR.

Of the 4,804 issues remaining after Full Count Review, 2,810 issues (58 percent), were not referred to CQR. Of the 1,994 issues (42 percent) that were referred to CQR, 537 were actually accepted by the program. The remaining 1,457 issues referred to CQR did not meet the Bureau’s CQR documentation requirements and, consequently, the Bureau took no further action on them.

The Full Count training materials we examined as part of our 2002 review did not provide any specific guidance on the type of evidence analysts needed to support data issues. Rather, the materials instructed analysts to supply as much supporting information as necessary. In contrast, the CQR program had more rigorous documentation requirements. Guidance available on the Bureau’s Web site required governmental units to supply maps and other evidence specific to the type of correction they were requesting, or the Bureau would not investigate their submissions.

⁷GAO, *2000 Census: Refinements to Full Count Review Program Could Improve Future Data Quality*, [GAO-02-562](#) (Washington, D.C.: July 3, 2002).

Simply put, Full Count Review identified hundreds of data issues but lacked the time to investigate the vast majority of them. Then, when the remaining cases were referred to CQR, most were rejected because they could not meet CQR's higher evidentiary bar.

A Mechanism for Correcting Public Law Data Will Be Critical for Future Enumerations

The Bureau lacked a program specifically designed to correct individual count errors contained in the apportionment and redistricting data. Because these numbers were later found to be flawed for some jurisdictions, as the Bureau proceeds with its plans for the 2010 Census, it will be important for it to explore options for reviewing and correcting this essential information before it is released.

Precision is critical because, in some cases, small differences in population totals could potentially impact apportionment and/or redistricting decisions. For example, according to an analysis by the Congressional Research Service, under the formula used to apportion seats in the U.S. House of Representatives, had Utah's population count been 855 persons higher, it would have gained an additional congressional seat and North Carolina would have lost a seat. However, had the duplicate UNC count and other errors detected by the CQR program as of September 30, 2003, been uncovered prior to the release of the public law data, the already narrow margin determining whether Utah gained a House seat would have dropped to 87 persons.⁸ Although in this particular instance there would not have been a change in congressional apportionment, it illustrates how the allocation of House seats can be determined by small differences in population counts.

Other Aspects of CQR Could Have Been Better Planned

Better planning could have improved the CQR program in other ways. For example, the Bureau's draft evaluation of CQR found, among other issues, that the three teams working on the planning and development phases of CQR should have tested implementation plans earlier in the process, and training materials were not based on the Bureau's experience in conducting the 1990 CQR program. Also, there was no mechanism to prioritize cases based on the magnitude of the error. As a result, regional offices wound up

⁸Congressional Research Service, *House Apportionment: Could Census Corrections Shift a House Seat?*, RS21638 (Washington, D.C.: Oct. 8, 2003).

expending considerable resources on CQR cases that only affected a handful of dwellings.

The draft evaluation also found that the two software applications the Bureau chose to administer and track CQR cases did not appear to be up to the task. Lost cases and documentation, poor integration with other applications, and the inability to produce reports were among the issues the evaluation cited.

More generally, the integration and coordination issues that affected CQR are not unique to that program; to the contrary, our past reports have found that other components of the 2000 Census were not well planned, which unnecessarily increased the cost and risk of the entire enumeration.⁹ The need for better strategic planning has been a consistent theme in many of our past recommendations to improve the Bureau's approach to counting the nation's population and represents a significant management challenge that the Bureau will need to address as it looks toward 2010.

The Bureau Is Making Count Correction Plans for 2010

The Bureau is beginning to develop plans for Full Count Review and CQR for the 2010 Census. As it does so, it will be important for it to develop an initiative or consolidated program that corrects both systemic and individual issues, and does so prior to the release of apportionment and redistricting data. Granted, this effort will be no simple task given the relatively short time between the closure of the local census offices and the need to release the public law data within the legally required time frames.

Still, there are steps the Bureau can explore to methodically check the data for nationwide systemic errors, obtain local input, and investigate any discrepancies, and do so in an expeditious manner. One approach might be to consolidate and leverage CQR, Full Count Review, and certain other Bureau programs.

Indeed, under the Full Count Review program, the Bureau obtained local input by contracting out some of the work to members of the Federal-State Cooperative Program for Population Estimates (FSCPE), an organization

⁹See for example, GAO, *2000 Census: Lessons Learned for Planning a More Cost-Effective 2010 Census*, [GAO-03-40](#) (Washington, D.C.: Oct. 31, 2002), 14 – 17, and GAO, *2010 Census: Cost and Design Issues Need to Be Addressed Soon*, [GAO-04-37](#) (Washington, D.C.: Jan. 15, 2004), 25 – 31.

composed of state demographers that has worked with the Bureau since 1973 to ensure accurate population counts. The Bureau worked with FSCPE, in part, because it lacked sufficient staff to complete the review on its own, but also because the Bureau believed that the members' knowledge of the demographic characteristics of their states could help the Bureau examine data files and products, including public law data. FSCPE members reviewed data for 39 states and Puerto Rico; Bureau employees reviewed data for the remaining states and the District of Columbia without FSCPE representation in Full Count Review. Both sets of analysts checked the data for their overall reasonableness, as well as for their consistency with historical and demographic data, and other census data products.

Bureau staff from its regional offices reviewed the data as well. They focused on identifying inconsistent demographic characteristics and did not necessarily concentrate on any one particular state or locality. Thus, the Bureau obtained local input that focused on individual states and smaller jurisdictions, and also performed its own, broader review.

Verifying any data discrepancies could be accomplished by beginning the count correction effort as local census offices complete nonresponse follow-up, when enumerators are still available to investigate issues. In fact, the Bureau is already planning to do this to some degree in 2010 under another operation called Coverage Improvement Follow-up (CIFU), where the Bureau is to call or visit housing units that have been designated as vacant or nonexistent but not confirmed as such by a second source. In the 2000 Census, CIFU began June 26, 2000, and ended on August 23, 2000. During that time, enumerators contacted 8.9 million housing units and counted 5.3 million people, according to the Bureau.

The Bureau could explore adding the count correction workload to enumerators' CIFU assignments, which would enable the agency to reconcile possible data errors, as well as add any housing units and group quarters the Bureau missed during the initial enumeration (As noted in the background section, CQR could not add any residences that existed on Census Day but the Bureau had failed to count.).

Further, the Bureau could help automate the count correction process by using computers to flag any data that exceed any predetermined tolerances. The Bureau could also develop a system to prioritize count correction issues to help manage its verification workload.

Importantly, to the extent the Bureau reviews and corrects census counts prior to the release of the public law data, the Bureau might not need separate Full Count Review and CQR programs; a consolidated effort might be more cost effective. At a minimum, to the extent a separate CQR program is needed, it may not have to be as large or last as long because presumably the earlier program would have caught the bulk of the problems.

Regardless, given the possibility that similar data errors might again occur during the 2010 Census, exploring options for resolving them prior to the release of public law data would be a sound investment. Reapportionment and redistricting data would be more accurate; the Bureau's credibility would be enhanced; and the need for a large-scale count correction program along the lines of CQR could be reduced or eliminated.

Conclusions

The CQR program played an important role in improving the quality of data from the 2000 Census. Although the net changes in housing and population counts from the program were small on a national scale, in a number of instances, they were substantial at the local level, and could affect various revenue sharing formulas and other programs that use decennial census data.

Because the program functioned as a safety net—a final opportunity to catch and correct mistakes that occurred along the chain of events that led to, and extended beyond Census Day 2000—the results shed light on the performance and limitations of certain upstream census operations, and areas where the Bureau should focus its efforts as its plans unfold for 2010. In this regard, the following is clear: although the Bureau puts forth tremendous effort to ensure a complete and accurate census, its numerous procedures and quality assurance operations will be challenged to stay ahead of the increasing difficulties associated with enumerating a population that is growing larger, more diverse, and increasingly hard to locate and count.

The timing of any count correction effort will also be critical. Indeed, we are concerned that key decisions using data from the 2000 Census employed figures that, for a number of jurisdictions, were later found to be flawed. As a result, it will be important for the Bureau to consider developing a count correction initiative that can complete its work in time to correct the public law data before that information is released.

Moreover, beyond the inherent demographic obstacles to a successful census, the results of our CQR review echo several of our past reports on other aspects of the census, which note that some of the Bureau's difficulties stem from a lack of adequate strategic planning and other management challenges. Ultimately, the success of the 2010 Census will hinge on the extent to which senior Bureau leadership resolves these challenges.

With this in mind, resolute action is needed across three fronts. First, it will be important for the Bureau to ensure, via thorough field testing, that its planned changes to its address list development procedures help resolve the geocoding and other operational problems revealed by CQR. Second, it will be important for the Bureau to improve its count correction efforts by designing a program that can systematically and consistently review the public law data and make any corrections prior to the release of those figures. Third, it will be important for the Bureau to address persistent strategic planning challenges.

Recommendations for Executive Action

To help ensure the nation has the best possible data for purposes of apportionment, redistricting, and other uses of census data, we recommend that the Secretary of Commerce direct the Bureau to improve its count correction efforts for the 2010 Census by taking such actions as:

1. Thoroughly testing improvements to the Bureau's group quarters and other address list development activities to help ensure the Bureau has resolved geocoding and other problems with its master address file.
2. Consolidating Full Count Review and CQR into a single program that systematically reviews and corrects any errors in the public law data prior to their release.
3. Expediting count correction efforts by initiating data reviews toward the end of nonresponse follow-up, when the Bureau starts getting complete data for geographic entities, and enumerators are available to help investigate any discrepancies. As part of this effort, the Bureau should consider using computers to systematically search for possible errors nationwide by checking data at the appropriate level of geography to ensure population, housing unit, and group quarter counts, as well as demographic characteristics, appear reasonable and are consistent with population estimates. Those areas that are outside of predetermined tolerances should be flagged for further review. The

Bureau should also pay special attention to ensure group quarters are properly geocoded and counted.

4. Prioritizing the investigation of errors based on the magnitude of the suspected error or similar triaging formula.
5. Ensuring that instructions on the Bureau's Web site make it clear that updated information exists and that users can readily access this information.
6. Improving the Bureau's quality assurance procedures to help ensure there are no mistakes in the data the Bureau posts on its Web site.
7. Enhancing the training and guidance provided to regional offices to help ensure they share the same understanding of their roles and responsibilities and will implement the program consistently.
8. Addressing persistent strategic management challenges, in part, through early testing to help ensure information systems, training, and other activities are fully integrated.

Agency Comments and Our Evaluation

The Acting Deputy Secretary of Commerce provided written comments on a draft of this report on May 20, 2005, which are reprinted in appendix III. Commerce stated that "the report provides a good overview of program results and makes several useful observations and recommendations," and specifically agreed with our finding that the process for conducting internal reviews was not consistently implemented. More generally, however, Commerce believes the shortcomings we describe reflect "a fundamental misunderstanding of the goals of the CQR program," and noted that our observations and recommendations indicate we believe that CQR should have been designed to correct the public law data before they were released during the 2000 Census.

Our concern over the CQR program centers on the way it was implemented in 2001, rather than the fact that the Bureau did not design the program to correct the apportionment and redistricting numbers. We agree with Commerce that this was not the intent of CQR and, as Commerce notes, we acknowledge this in our report. At the same time, based on the lessons learned from the 2000 Census, enumeration errors are almost inevitable. Thus, our recommendations focus on the future, and specifically, the importance of developing mechanisms for the 2010 Census to review and

correct errors in the public law data to the greatest extent possible before they are released. We have clarified the report to better reflect GAO's position.

Commerce specifically addressed two of our eight recommendations, disagreeing with both of them. With respect to our recommendation to consolidate Full Count Review and CQR into a single program for the 2010 Census, Commerce noted that preliminary counts at the census tract or block level are needed to conduct an effective CQR program, and that information is not available until close to the deadline for releasing the apportionment data. Commerce maintains there would be little opportunity for local entities to review the counts and document potential problems and even less time for the Bureau to conduct the necessary research and field work.

Our report recognizes that it would be a challenge for the Bureau to review and correct census figures and still release the public law data by the legally required time frames. Still, as we note in our report, we believe the Bureau could expedite the process by taking such steps as (1) using computers to check census data for their overall reasonableness and flagging areas that exceed predetermined tolerances; (2) focusing on known trouble spots such as group quarters; and (3) beginning the review process earlier, such as, when local census offices complete their nonresponse follow-up efforts.

Moreover, as we state in the report, during the 2000 Census, the Bureau already had programs in place that obtained local input on the census numbers before the release of the public law data (Full Count Review), and conducted extensive field operations to investigate certain discrepancies (Coverage Improvement Follow-up). We believe that it will be important for the Bureau to not simply replicate these programs for the 2010 Census or make incremental improvements, but to see whether these programs could be better leveraged and be more strategically employed to improve the accuracy of the apportionment and redistricting data.

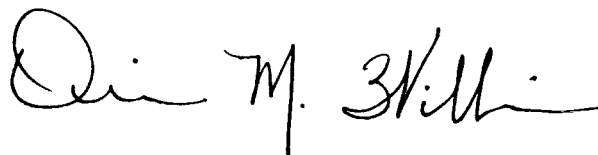
The other recommendation that Commerce specifically addressed was our call for the Bureau in 2010 to prioritize the investigation of errors based on the magnitude of the suspected problem. Commerce maintains that the Bureau's policy in 2000 was to handle cases in the order they were received from local jurisdictions, and asserts this was a fair and reasonable practice. While this practice is not unreasonable, we continue to believe that it would be more cost-effective for the Bureau to give priority to those cases

where it could achieve a greater return on its investment in resources (especially given our findings involving group quarters such as prisons and college dormitories that affected relatively large population clusters). Our recommendation echoes the Bureau's draft evaluation of the CQR program, which noted that regional offices expended considerable resources on CQR cases that affected only a handful of dwellings. Moreover, as we state in our report, prioritizing the Bureau's workload could help expedite the count correction process.

Commerce's comments also included some technical corrections and suggestions where greater clarity was needed. We revised the report as appropriate.

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>.

If you or your staff have any questions about this report, please contact me at (202) 512-6806 or williamso@gao.gov. Contact points for our Office of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made major contributions to this report are listed in appendix IV.

A handwritten signature in black ink, appearing to read "Orice M. Williams". The signature is fluid and cursive, with the first name "Orice" being the most prominent part.

Orice M. Williams
Director
Strategic Issues

Change in State Populations As a Result of Count Question Resolution Program

State	2000 Census total population	CQR total population	Total population change
U.S. Total	281, 421,906	281, 424,603	2,697
Alabama	4,447,100	4,447,351	251
Alaska	626,932	626,931	-1
Arizona	5,130,632	5,130,632	0
Arkansas	2,673,400	2,673,400	0
California	33,871,648	33,871,653	5
Colorado	4,301,261	4,302,015	754
Connecticut	3,405,565	3,405,602	37
Delaware	783,600	783,600	0
District of Columbia	572,059	572,059	0
Florida	15,982,378	15,982,824	446
Georgia	8,186,453	8,186,816	363
Hawaii	1,211,537	1,211,537	0
Idaho	1,293,953	1,293,956	3
Illinois	12,419,293	12,419,647	354
Indiana	6,080,485	6,080,517	32
Iowa	2,926,324	2,926,382	58
Kansas	2,688,418	2,688,824	406
Kentucky	4,041,769	4,042,285	516
Louisiana	4,468,976	4,468,958	-18
Maine	1,274,923	1,274,923	0
Maryland	5,296,486	5,296,507	21
Massachusetts	6,349,097	6,349,105	8
Michigan	9,938,444	9,938,480	36
Minnesota	4,919,479	4,919,492	13
Mississippi	2,844,658	2,844,656	-2
Missouri	5,595,211	5,596,683	1,472
Montana	902,195	902,195	0
Nebraska	1,711,263	1,711,265	2
Nevada	1,998,257	1,998,257	0
New Hampshire	1,235,786	1,235,786	0
New Jersey	8,414,350	8,414,347	-3
New Mexico	1,819,046	1,819,046	0
New York	18,976,457	18,976,821	364
North Carolina	8,049,313	8,046,485	-2,828

Appendix I
Change in State Populations As a Result of
Count Question Resolution Program

(Continued From Previous Page)

State	2000 Census total population	CQR total population	Total population change
North Dakota	642,200	642,200	0
Ohio	11,353,140	11,353,145	5
Oklahoma	3,450,654	3,450,652	-2
Oregon	3,421,399	3,421,436	37
Pennsylvania	12,281,054	12,281,054	0
Rhode Island	1,048,319	1,048,319	0
South Carolina	4,012,012	4,011,816	-196
South Dakota	754,844	754,844	0
Tennessee	5,689,283	5,689,267	-16
Texas	20,851,820	20,851,790	-30
Utah	2,233,169	2,233,198	29
Vermont	608,827	608,827	0
Virginia	7,078,515	7,079,030	515
Washington	5,894,121	5,894,141	20
West Virginia	1,808,344	1,808,350	6
Wisconsin	5,363,675	5,363,715	40
Wyoming	493,782	493,782	0
Puerto Rico	3,808,610	3,808,603	-7

Source: GAO analysis of U.S. Census Bureau data.

Human Error and Other Factors Contributed to University of North Carolina Counting Errors

The duplicate counting of nearly 2,700 students at the University of North Carolina (UNC) at Chapel Hill during the 2000 Census resulted from a combination of factors. The incident is interesting because it shows how the various safety nets the Bureau has built to ensure an accurate count can be undermined by human error, the limitations of census-taking operations, and other events that in some cases occur years before Census Day (April 1, 2000).

The duplicate count was discovered after CQR began when the director of the Charlotte regional office (a UNC graduate), asked one of her geographers (also a UNC graduate), to see whether the UNC dormitories were counted in their correct locations. According to the geographer, the director's curiosity was aroused after the CQR program found problems with the geocoding of dormitories at other schools in the Charlotte region. The geographer told us he initiated an internal CQR case in the summer of 2001 after discovering that two UNC dormitories were geocoded to the wrong census block. Upon further research, where he reviewed information from the census address file and the UNC Web site, the geographer concluded that, in addition to the geocoding error, a large number of dormitories and their occupants were counted in error.

Ultimately, by matching census records, the Bureau determined that 1,583 dormitory rooms in 26 buildings—and the 2,696 students who had resided in them—were included twice in the 2000 Census. On the basis of our interviews with Bureau staff and review of pertinent documents, the following sequence of events led to these erroneous enumerations:

The Bureau divides the places where people live into two broad categories: group quarters, which include prisons, dormitories, and group homes; and housing units, which consist of single family homes, apartments, and mobile homes. During the 2000 Census, the Bureau had distinct procedures for building its group quarters and housing unit address lists and enumerating their residents. For example, the Bureau typically enumerates college dormitories by working with schools to distribute census questionnaires to students. Conversely, the Bureau enumerates residents of housing units by delivering questionnaires directly to them through the mail. In the UNC situation, the 26 UNC dormitories were listed correctly in the Bureau's group quarters database and incorrectly in the Bureau's housing unit database.

Concerned there could be systemic issues with the Bureau's address list, staff at the Bureau's headquarters investigated the source of the problem

following the initial discovery by Charlotte employees. The headquarters review found that the dormitories were improperly included in the U.S. Postal Service's address file, which it initially shared with the Bureau in November 1997 and continued to update through early 2000. The Bureau uses this database to help build its housing unit address list. Specifically, the Bureau discovered that the data field that normally contains a street address erroneously contained a unit number and the name of a UNC dormitory. The Bureau had no explanation for how the dormitory names got into the U.S. Postal Service's address file.

Other procedures designed to verify census addresses produced conflicting results, compounding the problem. One procedure in 1998 mistakenly confirmed the dormitories as housing units, while another procedure—called block canvassing—correctly flagged the addresses for deletion from the Bureau's housing unit address list. However, under the Bureau's protocols, to ensure an address was not improperly removed from the census, an address had to be flagged twice to be deleted. During nonresponse follow-up in 2000, where enumerators visited housing units that failed to send back the questionnaires that were mailed to them, the Bureau had a third opportunity to uncover the error. Because the enumerators involved in this operation provided inconsistent information, the Bureau ultimately did not delete any housing units included in the initial census.

As part of the CQR case analysis, staff in the Bureau's Decennial Statistical Studies Division checked the Bureau's address file for any records that contained the word "dorm" in the address field to determine whether a similar duplication occurred at other schools. This would have picked up the word "dormitory" and its variants. On the basis of this search, the Bureau concluded that a similar issue was not problematic elsewhere in the country.

Comments from the Department of Commerce



THE DEPUTY SECRETARY OF COMMERCE
Washington, D.C. 20230

May 18, 2005

Ms. Orice M. Williams
Director
Strategic Issues
U.S. Government Accountability Office
Washington, DC 20548-0001

Dear Ms. Williams:

The U.S. Department of Commerce appreciates the opportunity to comment on the U.S. Government Accountability Office draft report entitled *Data Quality: Improvements to Count Correction Efforts Could Produce More Accurate Census Data (GAO-05-463)*. I enclose the Department's comments on this report.

Sincerely,

A handwritten signature in black ink, appearing to read "David A. Sampson", is written over the word "Sincerely,".

David A. Sampson
(Acting)

Enclosure

U.S. Department of Commerce

Comments on U.S. Government Accountability Office Draft Report Entitled
*Data Quality: Improvements to Count Correction Efforts
Could Produce More Accurate Census Data*
GAO-05-463

The Department of Commerce has the following general and specific comments on this draft audit report.

1. According to this report, the Government Accountability Office (GAO) was asked to review the results of the Count Question Resolution (CQR) program and to assess whether the program was consistently implemented across the country. While the report provides a good overview of program results and makes several useful observations and recommendations, we believe almost all of the shortcomings described by the GAO illustrate a fundamental misunderstanding of the goals of the CQR program rather than criticisms of its results or implementation.

The Census 2000 CQR program was not designed or publicized as a mechanism to correct the census results for apportionment or redistricting purposes. This was stated explicitly in the January 22, 2001, *Federal Register* notice that announced this program. While the report acknowledges this fact at the outset (see page 2), it then goes on to make various observations and recommendations that indicate the GAO believes the program should have been designed to achieve this goal for Census 2000 and that the U.S. Census Bureau should establish this as a goal for the 2010 Census.

We strongly disagree with such conclusions and recommendations concerning the Census 2000 CQR program. The Census Bureau does the best job it can to correctly enumerate the population within very strict and stringent time constraints imposed by federal law (13 U.S.C. §141 and P.L. 94-171). At the end of that effort, errors inevitably will remain in the results. Nonetheless, at specified points in time (December 31 of the census year for apportionment, April 1 of the following year for redistricting), the census results must be reported so that the Congress, the states, and local and tribal governments can use them for the purposes prescribed in those same laws. The Census Bureau does not have the authority to delay the reporting of those results, or to continue correcting those results over some period of time after the fact.

The Census Bureau does, however, have the authority and responsibility to evaluate the results of the decennial census and to provide whatever information it can to assist all data users in understanding the limitations of census data. Further, it has the authority to conduct Special Censuses and other efforts aimed at providing more current data about the population of any particular jurisdiction. As you note, apportionment and redistricting are not the only uses of decennial census results. Over the course of the decade, decennial census results are used to distribute hundreds of billions of dollars in federal, state, local, and tribal funding. While many

of the laws that govern these programs specify the use of the most recent decennial census results, some of them permit the use of other official Census Bureau data. This is why the Census Bureau offers the Special Census program and why it produces intercensal estimates and projections of the population. And, it was to this end that the Census Bureau created the CQR program to provide local governments a mechanism to identify potential problems in Census 2000 data that might have resulted from census processing errors and that might have affected the jurisdiction's funding streams over the decade.

If any updates were made to a jurisdiction's population and housing unit counts as a result of the CQR program, a statement documenting and certifying the updated figures was sent from the Director of the Census Bureau to the highest elected official of the jurisdiction. A copy of the statement also was sent to each affected jurisdiction's Secretary of State and to other state officials. The statement included the following language:

"Census counts used for Congressional apportionment and legislative redistricting, and Census 2000 data products, will remain unchanged. The Census Bureau will include the corrections in the errata information to be made available via the Internet on the *American FactFinder* system and used specifically to modify the decennial census file for use in yearly postcensal estimates beginning in December 2002."

While the GAO may not be suggesting that the Census Bureau should have delayed (or issued updated) counts for apportionment and redistricting based on the Census 2000 CQR program, it seems clear the GAO believes the Census Bureau should have designed the 2000 program so that it could have been completed before those counts were issued. In addition, the report recommends that the Census Bureau pursue such a design for the 2010 Census. Although the Census Bureau is exploring ways to improve the CQR program for the 2010 Census, it does not believe such a design is feasible:

- By its very nature, the CQR program requires census counts for very small areas. The Census Bureau believes it would be infeasible and unproductive to provide preliminary counts only at the city or county level and then expect jurisdictions to detect and pinpoint problems that could be reviewed in any systematic or timely fashion.
- The preliminary counts at the census tract or block level that are needed to conduct an effective CQR program are not available until the apportionment deadline is fast approaching. This leaves virtually no time for local jurisdictions to review the counts and assemble documentation supporting any potential problems, and even less time for the Census Bureau to conduct necessary research and field work to determine what corrections are needed.

2. The report does not acknowledge until page 13 that the number of potential count problems identified by local jurisdictions was extremely small compared to the total number of jurisdictions (1,180 out of about 39,000—only about 3%) and that the total population change resulting from all CQR cases was even smaller relative to the total U.S. population (less than 3,000 persons out of 281 million—barely one one-thousandth of one percent). The Census Bureau agrees, as also stated on page 13, that some of these changes may have had relatively large effects for a particular location. However, the summary of findings on the cover page of the report does not provide any of this context. Further, the first sentence of the report (the CQR program) says “corrected numbers affecting 47 states and over 1,180 governmental units” and then goes on to cite a few cases that involved relatively large count changes. Most count changes resulting from CQR were very small. For all of these reasons, we believe the summary of findings presented on the cover page of the report is very misleading as to the magnitude of problems revealed by this effort.
3. Based on its review, the GAO concluded that different Census Bureau regional offices implemented the program to different degrees. While we agree this is a valid criticism with respect to the internal review process the Census Bureau conducted in parallel with the CQR, we do not believe it is valid with respect to the way the Census Bureau handled external CQR challenges from local jurisdictions. For those external challenges, the Census Bureau believes that all regions followed the same procedures for their investigations.

The internal review process involved Census Bureau staff, the Federal-State Cooperative Program for Population Estimates, and others who were checking Census 2000 data for reasonableness, internal and intra-product consistency, and consistency with historical and external data sources. Reviewers identified, addressed, and/or explained issues or problems related to coverage, content, processing, and geocoding. Unresolved potential problems were forwarded to the CQR staff for additional analysis. Changes made as a result of this internal review and/or research were incorporated into the CQR process and documented the same way as changes based on external CQR challenges from jurisdictions.

The Census Bureau agrees the process for conducting internal reviews was not planned or implemented as systematically as the review of external challenges submitted by jurisdictions.

4. The report also criticizes the Census Bureau for not prioritizing CQR cases based on the size of the problem or some other measure of criticality. The Census Bureau’s policy was to handle the cases in the order they were received from local jurisdictions, and we believe that was fair and reasonable for a program where cases were submitted over a two-year period. The Census Bureau also gave higher priority to external challenges from jurisdictions than to internal review cases.

5. Page 6 of the report states that CQR was one of several Census Bureau quality assurance activities intended to improve the accuracy of Census 2000 data. We believe this statement is misleading because it implies the program was part of those coverage improvement operations built into the decennial census operations that produced the final apportionment and redistricting data. As stated earlier, this is not the case.

The last paragraph of page 12 includes a statement that there were no CQR count corrections for the District of Columbia. However, while the CQR program did not change the population count for the District of Columbia, the program did identify some geocoding errors within the District of Columbia. Figure 3 also should be revised to include the District of Columbia.

6. The paragraph in the middle of page 30 should be revised to state that the Full Count Review identified hundreds of potential data issues.

GAO Contact and Staff Acknowledgments

GAO Contact

Orice Williams, (202) 512-6806

Acknowledgments

In addition to the contact named above, Robert Goldenkoff, Keith Steck, Timothy Wexler, Robert Parker, Michael Volpe, Andrea Levine, and Elena Lipson made key contributions to this report.

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

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548