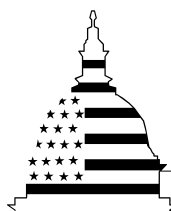


January 2003

DECENNIAL CENSUS

Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement



G A O

Accountability ★ Integrity ★ Reliability



Highlights of [GAO-03-228](#), a report to Congressional Requesters

Why GAO Did This Study

To help boost response rates of both the general and Hispanic populations, the U.S. Census Bureau (Bureau) redesigned the 2000 questionnaire, in part by deleting a list of examples of Hispanic subgroups from the question on Hispanic origin. While more Hispanics were counted in 2000 compared to 1990, the counts for Dominicans and other Hispanic subgroups were lower than expected. Concerned that this was caused by the deletion of Hispanic subgroup examples, congressional requesters asked us to investigate the research and management activities behind the changes.

What GAO Recommends

GAO recommends that the Bureau

- implement its plans to conduct further research on the Hispanic question, taking steps to properly test the impact of any changes on the quality of data on Hispanic subgroups and Hispanics overall, and
- develop agencywide protocols that provide guidelines for Bureau 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 to delay or suppress the data.

The Bureau agreed with our recommendations, but took exception to our findings concerning the adequacy of its data quality guidelines.

www.gao.gov/cgi-bin/gettrpt?GAO-03-228.

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

DECENNIAL CENSUS

Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement

What GAO Found

In both the 1990 and 2000 censuses, Hispanics could identify themselves as Mexican, Puerto Rican, Cuban, or other Hispanic. Respondents checking off this latter category could write in a specific subgroup such as “Salvadoran.” The “other” category in the 1990 Census included examples of subgroups to clarify the question. For the 2000 Census, the Bureau removed the subgroup examples as part of a broader effort to simplify the questionnaire and help improve response rates. The Bureau removed unnecessary words and added blank space to shorten the questionnaire and make it more readable.

Although the Bureau conducted a number of tests on the sequencing and wording of the race and ethnicity questions, and sought input from several expert panels, no Bureau tests were designed specifically to measure the impact of the questionnaire changes on the quality of Hispanic subgroup data. According to Bureau officials, because federal laws and guidelines require data on Hispanics but not Hispanic subgroups, the Bureau targeted its resources on research aimed at improving the overall count of Hispanics. Bureau evaluations conducted after the census indicated that deleting the subgroup examples might have confused some respondents and produced less-than-accurate subgroup data. A key factor behind the Bureau’s release of the questionable subgroup data was its lack of adequate guidelines governing the quality needed before making data publicly available. As part of its planning for the 2010 Census, the Bureau intends to conduct further research on the Hispanic origin question, including a field test in parts of New York City. However, until research on a new version of the question is finalized, Bureau officials said that other census surveys will continue to use the 2000 Census format of the Hispanic origin question.

Enumerator Administers Census Questionnaire



Source: U.S. Census Bureau.

Contents

Letter	1
Results in Brief	2
Background	4
Objectives, Scope, and Methodology	5
Efforts to Simplify Questionnaire Led Bureau to Delete List of Example Hispanic Subgroups	6
The Bureau Plans to Conduct Targeted Research on Hispanic Subgroups in the Future	15
Conclusions	17
Recommendations for Executive Action	18
Agency Comments and Our Evaluation	18

Appendix	
Appendix I: Comments from the Department of Commerce	20

Related GAO Products	25
----------------------	----

Figures	
Figure 1: Evolution of the Hispanic Question from the 1970 Census to the 2000 Census	7
Figure 2: The Bureau Simplified the 2000 Census Questionnaire	9
Figure 3: The 2000-Style Questionnaire Produced Lower Subgroup Counts than Those from a Test Using the 1990-Style Questionnaire	15

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. It may contain copyrighted graphics, images or other materials. Permission from the copyright holder may be necessary should you wish to reproduce copyrighted materials separately from GAO's product.



United States General Accounting Office
Washington, D.C. 20548

January 17, 2003

The Honorable Danny K. Davis
Ranking Minority Member
Subcommittee on Civil Service,
Census and Agency Organization
Committee on Government Reform
House of Representatives

The Honorable Wm. Lacy Clay
The Honorable Charles A. Gonzalez
The Honorable Carolyn B. Maloney
House of Representatives

Collecting data on race and ethnicity is among the federal government's most complex and controversial data collection efforts. The decennial census has collected these data in various forms beginning with the very first national headcount in 1790. Since the 1960s, race and ethnicity data have been used to monitor and enforce compliance with a number of civil rights laws, including those governing equality in employment, voting, housing, mortgage lending, health care services, and education. Over time, in response to changing federal mandates, demographics, and its own operational requirements, the U.S. Census Bureau (Bureau) has changed the format and sequence of the race and ethnicity questions. The Bureau made one such change for the 2000 Census when, in an effort to improve the count of Hispanics and simplify the questionnaire, it redesigned the question on Hispanic origin and dropped a list of examples of Hispanic subgroups.

As soon as the Hispanic and Hispanic subgroup data from the 2000 Census were released in May 2001, questions were raised about the counts for specific Hispanic subgroups. For example, the reported count of Dominican Hispanics was significantly lower than the counts reported in other Bureau surveys. Concerned that the lower-than-expected Hispanic subgroup counts were the result of dropping the list of example write-in Hispanic subgroups from the 2000 questionnaire, you asked us to investigate the research and management activities behind this change. As agreed with your offices, we reviewed (1) the decision-making process behind the Bureau's removal of the example subgroups, (2) the research the Bureau conducted to aid in that decision, and (3) the Bureau's future plans for collecting Hispanic subgroup data.

This report parallels our recent study addressing congressional concerns about how the Bureau reported data on people counted at emergency and transitional shelters, a segment of the population that includes, among others, the homeless.¹ Both reports are part of our ongoing series on lessons learned from the 2000 Census that can help inform the planning effort for 2010. (See the Related GAO Products section for the reports issued to date).

Results in Brief

The Bureau removed examples of Hispanic subgroups from the census question on Hispanic origin as part of an effort to make the questionnaire more “respondent-friendly.” The Bureau’s evaluations of the 1990 Census indicated that deleting unnecessary words and adding more white space, among other changes, could help improve response rates. The Bureau also modified the wording and format of the Hispanic question in order to improve Hispanic participation in the census.

Throughout the 1990s, the Bureau conducted a number of tests to determine the impact that these and other changes had on the overall count of Hispanics. However, because Office of Management and Budget standards governing the collection of race and ethnic data do not require data on Hispanic subgroups, the Bureau did not specifically design any tests to determine the likely effect of the changes on the quality of Hispanic subgroup data.

Although the Bureau did not test the likely impact of questionnaire changes on the Hispanic subgroup data, it released subgroup counts along with the overall Hispanic data in May 2001. Immediately following the release of these data, local government officials and representatives of Hispanic subgroups raised questions about the accuracy of specific subgroup counts. Bureau evaluations conducted following the census suggest that dropping the examples of Hispanic subgroups confused some respondents and produced less-than-accurate subgroup data. For example, in one experiment, the Bureau mailed a 1990-style questionnaire (which included subgroup examples) to a sample of individuals as part of the 2000 Census. The Bureau found that 93 percent of Hispanics given the 1990-style form reported a specific subgroup, compared to 81 percent of Hispanics given

¹U.S. General Accounting Office, *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).

the 2000-style form. Thus, while the Bureau reported what respondents marked on their questionnaires, because of respondents' confusion over the wording of the question, the subgroup data could be misleading.

The Bureau has made improving the quality of the Hispanic question a focus for the 2010 Census and intends to test questionnaire changes aimed at improving the quality of its overall count of Hispanics and its counts of Hispanic subgroups. In 2003, the Bureau is to begin testing the Hispanic question, and as part of a field test in 2004, the Bureau plans to administer the questionnaire in parts of the New York City borough of Queens. Any changes to the census questionnaire will also affect other Bureau surveys, such as the proposed American Community Survey (ACS), which the Bureau designed in part to replace the census long-form questionnaire. Bureau officials said that the ACS will continue to use the 2000 Census Hispanic question until research and testing on a new version is complete.

A key factor behind the Bureau's release of apparently less-than-accurate Hispanic subgroup data appears to be a lack of adequate guidelines governing decisions on quality considerations that should be addressed before making data publicly available. Had such guidelines been in place prior to releasing the Hispanic subgroup data, they could have prompted the Bureau to apply more rigorous quality checks on the accuracy of the Hispanic subgroup data; provided a basis for either releasing, delaying, or suppressing the data; and informed decisions on how to describe any of their limitations.

The lack of data quality guidelines resulted in similar difficulties when the Bureau initially decided not to release data on the homeless and others without conventional housing. In our companion report, we recommended that the Secretary of Commerce ensure that the Bureau develop agencywide guidelines governing the level of quality needed to release data to the public, when and how to characterize any limitations, and when it is acceptable to suppress data. Because these incidents, if repeated, could erode public confidence in the data, it will be important for the Bureau to implement these recommendations. Additionally, with respect to the Hispanic subgroup data, we are recommending that the Bureau take steps to properly test the impact that any changes to the Hispanic origin question have on the quality of Hispanic data, and the quality of Hispanic subgroups in particular.

The Secretary of Commerce forwarded written comments from the U.S. Census Bureau on a draft of this report (see app. I). The Bureau agreed

with our conclusions and recommendations and is taking steps to implement them, but took exception to our findings concerning the adequacy of its data quality guidelines.

Background

While the decennial census has long collected data on race and ethnicity,² a specific question on Hispanic origin was first added to the 1970 Census in response to the 1965 Voting Rights Act, which required the data to ensure equality in voting.³ Today, antidiscrimination provisions in a number of statutes require census data on race and Hispanic origin in order to monitor and enforce equal access to housing, education, employment, and other areas. The Office of Management and Budget (OMB), through its Federal Statistical Policy Directive No. 15, sets the standards governing federal agencies' collection and reporting of race and ethnicity data.

At least seven cabinet-level government departments, the Federal Reserve, every state government, and a number of public and private organizations use Hispanic data. Although not required by federal legislation or OMB standards, Hispanic subgroup data are also used for many of these same purposes. In addition, subgroup data are especially important to communities with rapidly growing and diverse Hispanic populations.

Collecting data on race and ethnicity has been a persistent challenge for the Bureau. Race and ethnicity are subjective characteristics, which makes measurement difficult. Moreover, the Bureau has found that some Hispanics equate their ethnicity—Hispanic—with race, and thus find it difficult to classify themselves by the standard race categories that include, for example, white, black, and Asian.

The Bureau's preparations for the 2000 Census included an extensive research and testing program to improve the Hispanic count. In 1990, the

²The Bureau, in accordance with Office of Management and Budget Federal Statistical Policy Directive 15, *Race and Ethnic Standards for Federal Statistics and Administrative Reporting*, collects data on two ethnicities: Hispanic origin and not of Hispanic origin. We use the same definition in this report. Additionally, the standards call for self-reporting of race and ethnicity rather than identification based on scientific or anthropological standards. The standards also cover reporting on race and ethnicity in administrative reports and for civil rights monitoring. They also specify that the data are not to be used for determining program eligibility.

³42 U.S.C. 1973aa-1a.

Bureau estimated that it did not enumerate 5 percent of the Hispanic population. Further, the ethnicity question, which was posed to all respondents, appeared to confuse both Hispanics and non-Hispanics. For example, many non-Hispanics, thinking the question only pertained to Hispanics, did not answer the question. Overall, 10 percent of respondents failed to answer the 1990 Hispanic question—the highest of any short form item in 1990. As a result, the Bureau made improving the Hispanic count a major priority for the 2000 Census.

Objectives, Scope, and Methodology

Our objectives were to review (1) the Bureau's decision-making process that led to its dropping the list of subgroup examples from the Hispanic question on the 2000 Census form, (2) the research conducted by the Bureau to aid in this decision, and (3) the Bureau's future plans for collecting Hispanic subgroup data.

To address each of these objectives, we interviewed key Bureau officials and examined Bureau, OMB, and other documents, including planning materials and internal memos. To obtain a local perspective of how municipal governments and community leaders use Hispanic subgroup data, we met with data users in New York City, including representatives of the New York Department of Planning and the Dominican and Puerto Rican communities. We also attended a meeting of the Dominican American National Round Table, a Dominican American advocacy group that discussed issues relating to the 2000 Census count of Dominican Hispanics. We also attended meetings of the Census Advisory Committee on Race and Ethnicity that addressed the issue of the quality of the Hispanic subgroup data.

Finally, to examine the research behind the Bureau's decision to remove the example subgroups from the 2000 questionnaire, we reviewed the results of the Bureau's National Content Survey, Targeted Race and Ethnicity Test, and other research conducted throughout the 1990s in preparation for the 2000 Census. Additionally, we reviewed information from the Bureau's meetings with its Advisory Committee on the Decennial Census and its Advisory Committee on Race and Ethnicity. We also examined relevant materials from OMB's Interagency Committee for the Review of the Racial and Ethnic Standards.

To review the Bureau's future plans for collecting Hispanic subgroup data, we attended meetings of the National Academy of Science Panel on Future Census Methods, the Decennial Census Advisory Committee, and the

Census Advisory Committee on Race and Ethnicity. We also discussed these plans with Bureau officials.

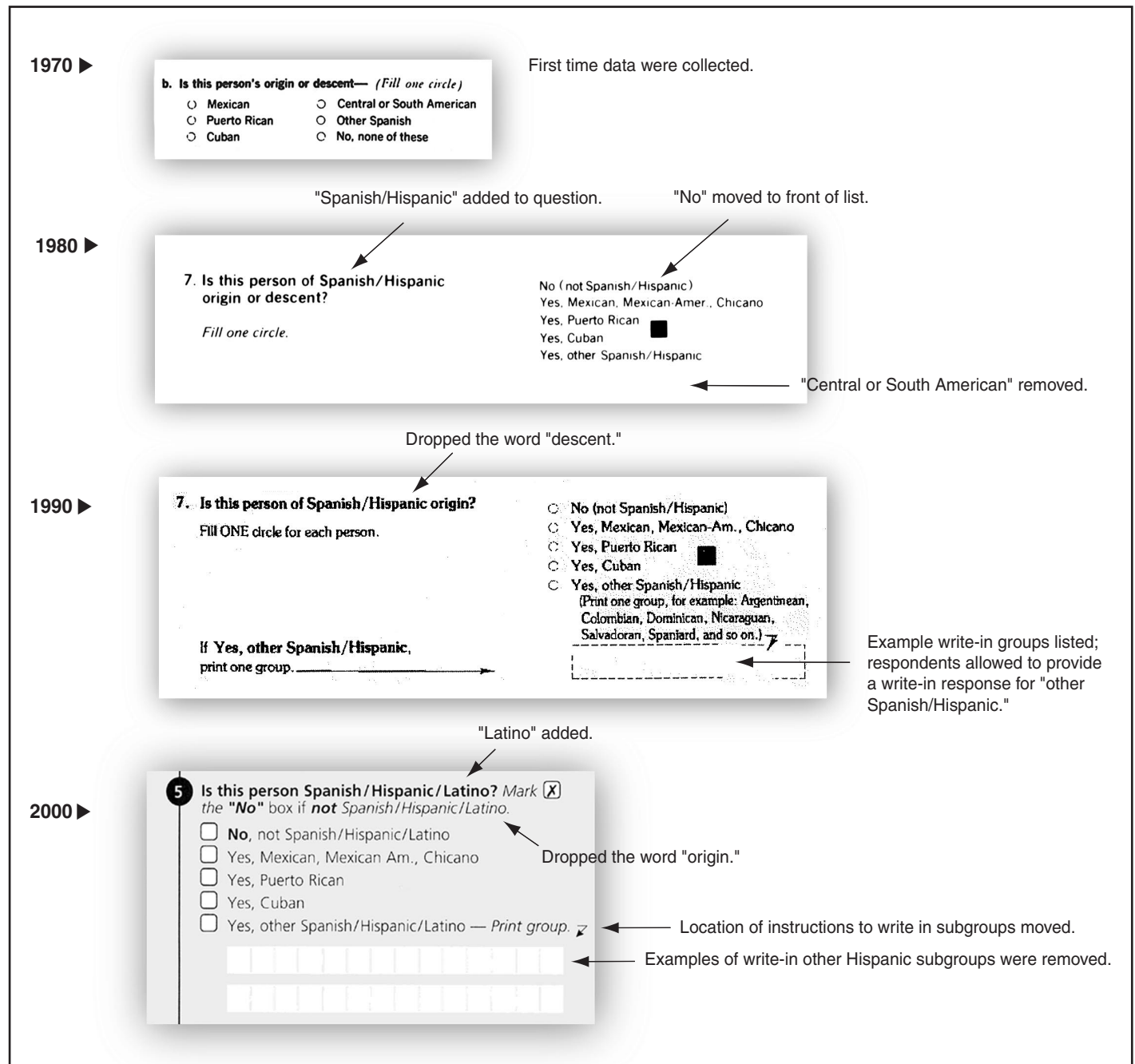
Our audit work was conducted in New York City and Washington, D.C., and at the Bureau's headquarters in Suitland, Maryland, from January through September 2002. Our work was done in accordance with generally accepted government auditing standards.

We requested comments on a draft of this report from the Secretary of Commerce. On November 27, 2002, the Secretary forwarded the U.S. Census Bureau's written comments on the draft. The comments are reprinted in appendix I. We address these comments at the end of this report.

Efforts to Simplify Questionnaire Led Bureau to Delete List of Example Hispanic Subgroups

Collecting accurate ethnic data has challenged the Bureau for over 30 years. Since the 1970 Census, when the Bureau first included a question on Hispanic origin, every census has had comparatively high Hispanic undercounts that reduced the quality of the data. As a result, the Bureau has modified the Hispanic question on every census since then as part of a continuing effort to improve the Hispanic count. (See fig. 1.) In addition, a Spanish language version of the census form has been available upon request since 1980.

Figure 1: Evolution of the Hispanic Question from the 1970 Census to the 2000 Census

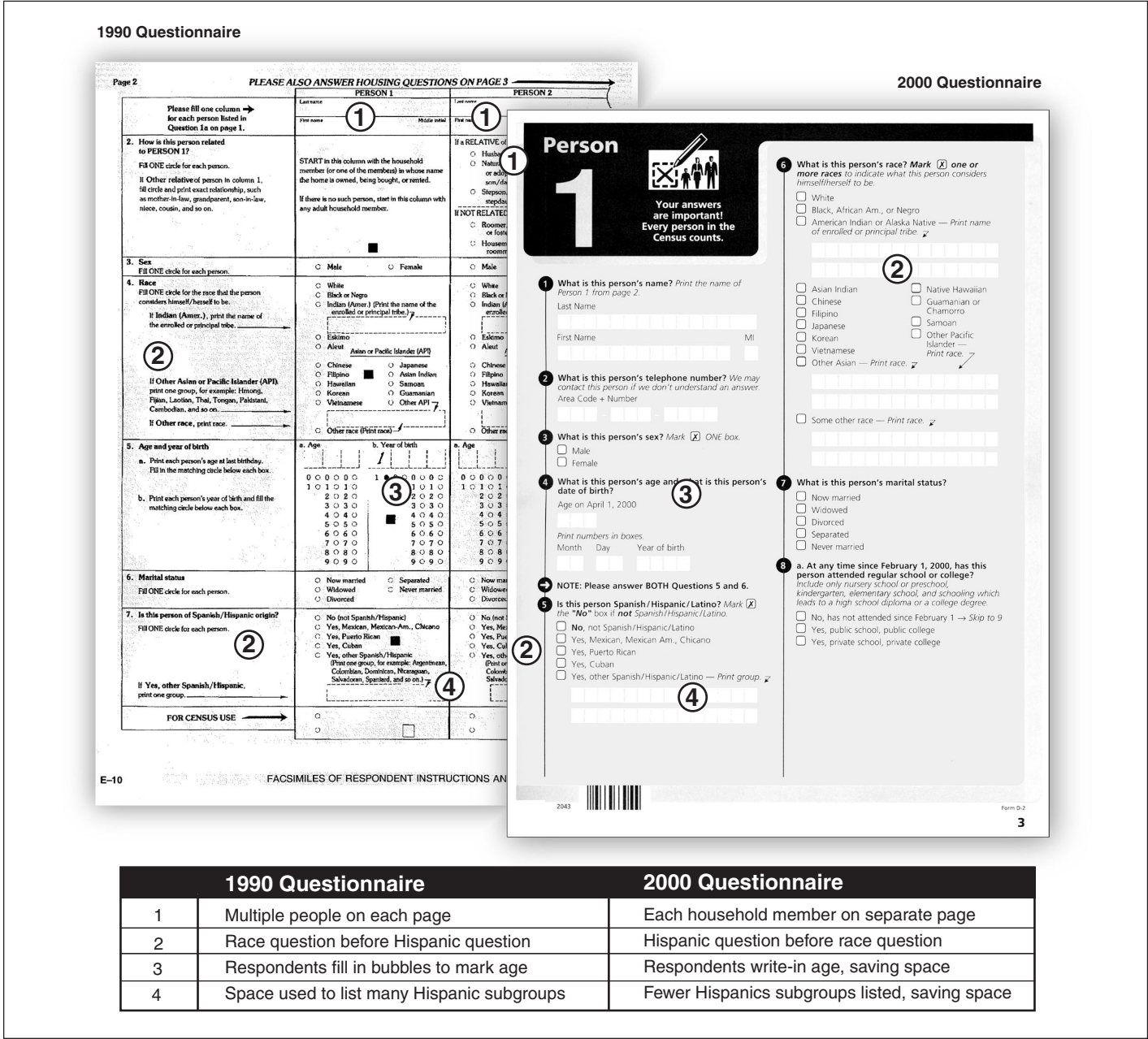


Source: U.S. Census Bureau and GAO analysis.

For the 2000 Census, Hispanics could identify themselves as Mexican, Puerto Rican, Cuban, or “other Spanish/Hispanic/Latino.” Respondents who checked off this last category could write in a specific subgroup such as “Salvadoran.” Although this approach was similar to that used for the 1990 Census, as shown in figure 1, the “other” category in the 1990 Census included examples of other Hispanic subgroups. The Bureau deleted these examples as one of several changes to the Hispanic question for the 2000 Census. Other changes included (1) adding the word “Latino” to the designation Spanish/Hispanic, (2) dropping the word “origin” from the question, and (3) moving the location of instructions on writing in an unlisted subgroup. According to Bureau officials, these latter three changes were made to improve the Hispanic count.

The Bureau removed the subgroup examples as part of a broader effort to simplify the questionnaire and thus help reverse the downward trend in mail response rates that had been occurring since 1970. Indeed, evaluations of the 1990 Census indicated that the overall design of the form was confusing to many and contributed to lower response rates, particularly among some hard-to-enumerate groups such as Hispanics. In redesigning the questionnaire, the Bureau added as much white space as possible, and removed unnecessary words to make the questionnaire shorter and more readable. As shown in figure 2, the 2000 questionnaire appears more “respondent-friendly” compared to the 1990 questionnaire.

Figure 2: The Bureau Simplified the 2000 Census Questionnaire



The Bureau initially proposed removing the example write-in subgroups during 1990 through 1992. A first version of the questionnaire without the example subgroups was used in the 1992 National Census Test. However, as discussed in the next section, testing continued from 1992 to 1996 to ensure that removing the write-in example groups did not harm the overall count of Hispanics. From 1995 to 1997, after testing showed that removal of the write-in example groups would not harm the overall Hispanic count, the Bureau finalized its decision to remove the example subgroups.

Although federal law and OMB standards⁴ only require information on whether an individual is Hispanic, Bureau officials told us they collect subgroup data to help improve the overall Hispanic count. According to the Bureau, many Hispanics do not view themselves as Hispanic, but identify instead with their country of origin or with a particular Hispanic subgroup. State and local governments, academic institutions, community organizations, and marketing firms, among other organizations, also use Hispanic subgroup data for a variety of purposes. For example, officials in the New York City Department of Planning told us that they need accurate information on the number and distribution of Hispanic subgroups in planning the delivery of numerous city services.

According to a Bureau official, no data are available on the precise impact the questionnaire redesign had on overall response rates in part because it was made in conjunction with other efforts to improve the response rate, such as a more aggressive outreach and promotion campaign. However, the initial mail response rate was 64 percent, 3 percentage points higher than the Bureau's expectations, and comparable to the similar 1990 mail response rate.

⁴Public Law 94-311 requires the collection of data on "Americans of Spanish origin or descent." OMB Federal Statistical Policy Directive 15 states that collection of data on Hispanic subgroups is optional, as long as the collection of these data does not harm efforts to collect accurate data on the number of Hispanics.

Moreover, evaluations conducted since the 2000 Census by the Bureau indicate that the Bureau obtained a more complete count of Hispanics in the 2000 Census than it did in 1990. For example, Bureau data show that the 2000 Census missed an estimated 2.85 percent of the Hispanic population compared to an estimated 4.99 percent in 1990—a 43 percent reduction of the undercount.⁵ The Bureau credits the improvement in part to the changes it made to the questionnaire. However, as discussed in the next section, removing the examples of Hispanic subgroups may have reduced the completeness of data on individual segments of the Hispanic population.

No Bureau Tests Were Designed Specifically to Measure the Impact of Questionnaire Changes on Hispanic Subgroup Data

Bureau guidance requires that any changes to the census form must first be thoroughly tested. For example, according to Bureau officials, before changing a question, the Bureau must first conduct research studies, cognitive tests, and field tests to determine how best to sequence and word the question, and to see if the proposed changes are likely to achieve the desired results. Additionally, the census questionnaire is to be reviewed by a variety of census advisory groups, OMB, and Congress before it is finalized.

Nevertheless, while the Bureau conducted a number of tests of the sequencing and wording of the race and ethnicity questions, according to Bureau officials, it did not specifically design any tests to determine the impact of the changes on the quality of Hispanic subgroup data.⁶ Because OMB standards do not require data on Hispanic subgroups, Bureau officials said that the Bureau targeted its resources on testing and research aimed at improving the overall count of Hispanics.

⁵These figures represent the net Hispanic undercount, which is the difference between the estimated Hispanic population per the Bureau's Accuracy and Coverage Evaluation Survey and the census count.

⁶The Census Bureau did look at the impact of changes on Hispanic subgroups. However, the sample size in the test was not large enough to detect statistically significant differences for the Hispanic subgroups that constitute the "Other Spanish/Hispanic/Latino" population. Additionally, the test was not designed to detect the impact of each change to the question separately.

Throughout the 1990s, in revising the race and ethnicity questions, the Bureau sought input from several expert panels, including the Interagency Committee formed by OMB⁷ and the Census Advisory Committee on Racial and Ethnic Populations, one of several panels with which the Bureau consulted to help it plan the 2000 Census. In addition, the Bureau conducted several tests of the questionnaire to assess respondents' understanding of the questions and their ability to complete them properly. They included the

- 1992 National Census Test, which field tested potential questions for the 2000 Census questionnaire;
- 1996 National Content Survey, which examined a number of issues to improve race and ethnic reporting; and
- 1996 Race and Ethnic Targeted Test, which tested alternative formats for asking race and ethnic questions.

In addition, the Bureau analyzed the results of Hispanic data from the 1990 Census (which led to its conclusions about the undercount), but did not conduct any specific evaluations of the quality of the 1990 Hispanic subgroup data. The consultation, research, and testing played a key role in the Bureau's decisions to place the ethnicity question before the race question and make several other changes discussed earlier in this report.

The test results also indicated that the example subgroups could produce conflicting results. On the one hand, the Bureau found that providing the example subgroups could help prevent respondents' confusion over how to describe their ethnicity. On the other hand, the Bureau found that removing the example subgroups could help reduce the bias caused by the example effect, which occurs when a respondent erroneously selects a response because it is provided in the questionnaire.

Although the Bureau conducted a dress rehearsal for the 2000 Census in 1998 in order to test its overall design, the dress rehearsal did not identify any problems with the Hispanic subgroup question. According to Bureau officials, this could have been because none of the three test sites—the city of Sacramento, California; Menominee County, Wisconsin, including the

⁷A group of more than 30 agencies that represent the many and diverse federal needs for data on race and ethnicity, including statutory requirements for such data.

Menominee American Indian Reservation; and the city of Columbia, South Carolina, and its 11 surrounding counties—had a large and diverse enough Hispanic population for the problems to become evident.

Questions Raised about the Quality of Reported Hispanic Subgroup Data

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, called the SF-1 file. 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 SF-1 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.

Following the initial release of the Hispanic data, local government officials and Hispanic advocacy groups raised questions about the accuracy of the counts of Hispanic subgroups listed as examples on the 1990 census form, but not the 2000 form. The 2000 Census showed lower counts of several Hispanic subgroups than analysts had expected based on their own estimates using a variety of information sources such as vital statistics, immigration statistics, population surveys, and other data. In New York City, local government officials and representatives of Hispanic subgroups who partnered with the Bureau to improve the enumeration of Hispanics told us that they were particularly concerned about low subgroup counts in their communities in part because they needed accurate numbers to plan and deliver specialized services to particular subgroups. Moreover, they said that because “official census numbers” are often considered definitive, problems with the released Hispanic subgroup numbers could lead to faulty decision making by data users.

Questionnaire Modifications May Have Led to Problems with Hispanic Subgroup Data

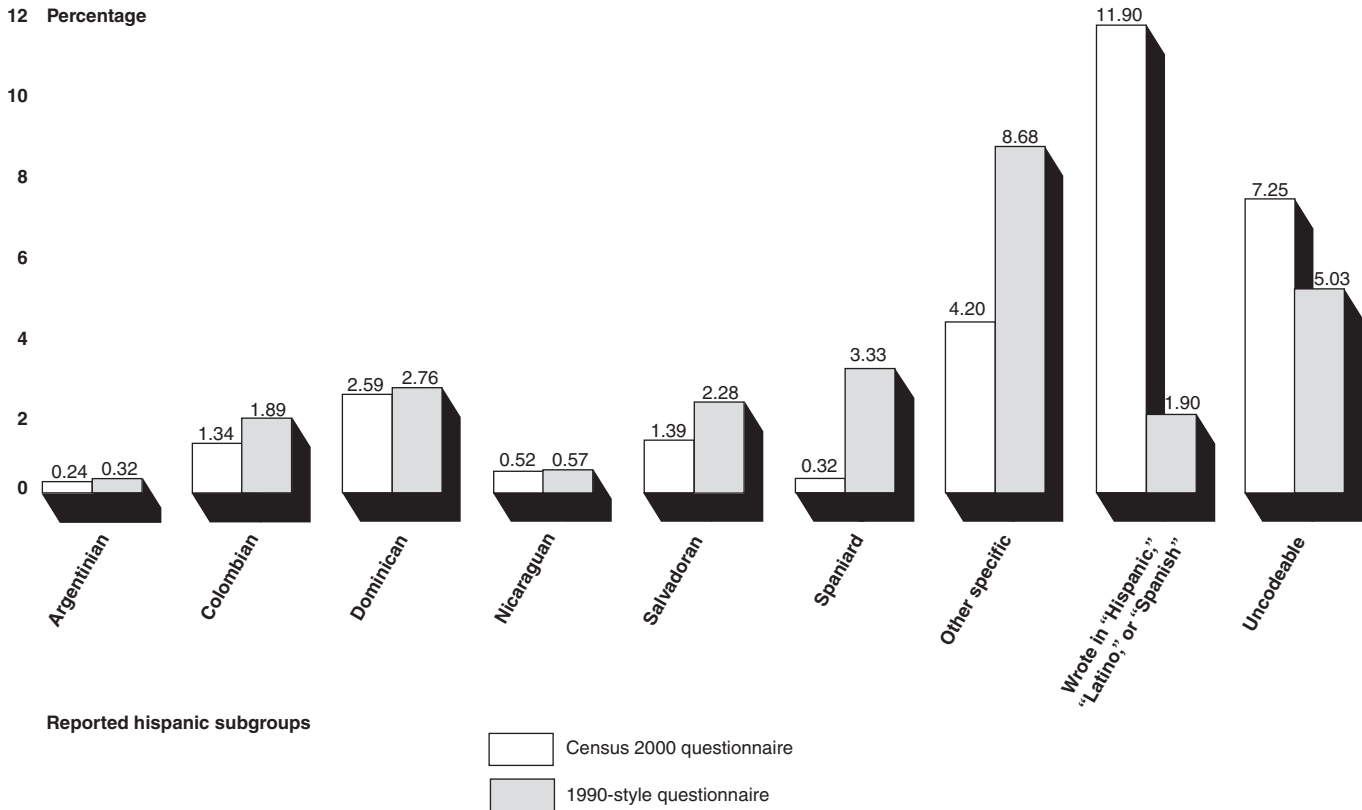
Since the release of the 2000 Census Hispanic data, the Bureau has conducted evaluations of the data that provided more information on how removing the subgroup examples may have affected the quality of Hispanic subgroup data. One key evaluation was the Alternative Questionnaire Experiment, in which the Bureau sent out 1990-style census forms to a sample of individuals as part of the 2000 Census. As shown in figure 3, the Bureau's research indicates that the 1990-style form elicited more reports of specific Hispanic subgroups than the 2000-style questionnaire.⁸ Indeed, 93 percent of Hispanics given the 1990-style form reported a specific subgroup, compared to 81 percent of Hispanics given the 2000-style form. Moreover, virtually every subgroup reported in the 2000-style form composed a smaller percentage of the overall Hispanic count than the 1990-style form. Thus, while the Bureau reported what respondents checked off on their questionnaires, because of respondents' confusion over the wording of the question, the 2000 subgroup data could be misleading.

Figure 3 also suggests that one possible reason for this might be that many respondents did not understand what they were supposed to write in, as many more people on the 2000-style form wrote in "Hispanic," "Spanish," or "Latino" (as opposed to a specific subgroup) compared to the 1990-style questionnaire. Additionally, a higher percentage of the respondents did not provide codeable (useable) responses.

Moreover, based on its analysis of the Census 2000 Supplementary Survey—an operational test for collecting long-form-type data based on a nationwide sample of 700,000 households—the Bureau estimated that there were about 150,000 more Dominican Hispanics than were counted in the 2000 Census. Some attribute the discrepancy to the fact that many respondents to the supplementary survey provided their answers by telephone, where enumerators were able to help them better understand the question on Hispanic subgroups.

⁸This study was conducted in English only. Because a sizable number of Hispanics only speak Spanish, the results of this study cannot be generalized to the Hispanic population at large.

Figure 3: The 2000-Style Questionnaire Produced Lower Subgroup Counts than Those from a Test Using the 1990-Style Questionnaire



Source: U.S. Census Bureau and GAO analysis.

The Bureau Plans to Conduct Targeted Research on Hispanic Subgroups in the Future

Because of concerns relating to the 2000 Census counts of Hispanic subgroups, Bureau officials said that they plan to focus testing and research on these questions in preparation for the 2010 Census. In particular, they stated that the Bureau would examine the likely impact of including Hispanic subgroup examples in the question again, as well as other aspects of the question that caused problems for some respondents. Before deciding on a new version of the Hispanic question, the Bureau must finish evaluating the results of the 2000 Census, conduct a number of cognitive tests, and field-test proposed changes to the question. The Bureau plans to begin testing the Hispanic question in 2003 and, as part of a field test in 2004, to administer the questionnaire in parts of Queens, New York, which the Bureau selected for its racial and ethnic diversity. The

Bureau intends to complete its testing and decide on changes to the Hispanic question from 2006 through 2008.

Any changes to the Hispanic question are relevant not only for the 2010 Census, but also for other Bureau questionnaires, such as the proposed ACS.⁹ Bureau officials told us that they expect that the ACS will continue to use the 2000 Census Hispanic question until research and testing on a new version is complete.

The Bureau Lacks Clearly Written, Transparent Guidelines for Releasing Data

While continued research could help the Bureau *collect* better-quality Hispanic subgroup data, it will also be important for the Bureau to address what led it to *release* data that could mislead users. A key factor in this regard is that the Bureau lacks adequate guidelines for making decisions about how data quality considerations affect the release of data to the public. Had such guidelines been in place prior to releasing the Hispanic subgroup data, they could have (1) prompted the Bureau to apply more rigorous quality checks on the Hispanic subgroup data, (2) provided a basis for either releasing, delaying, or suppressing the data, and (3) informed decisions on how to describe any limitations to data released.

This is not the first time that the lack of Bureau-wide guidelines on the level of quality needed for census results to be released to the public has created difficulties for the Bureau and data users. As we noted in our companion report¹⁰ on the Bureau's methods for collecting and reporting data on the homeless and others without conventional housing, one cause of the Bureau's shifting position on reporting those data and the resulting public confusion appears to be its lack of documented, clear, transparent, and consistently applied guidelines on the level of quality needed to release data to the public. With the Hispanic subgroup data, the Bureau released the information as planned before it could properly assess its quality, identify problems, and report its limitations. More rigorous guidelines could help ensure that decisions about the quality of all census data the Bureau releases are more consistent and better understood by the public.

⁹The ACS is designed to provide annual data for areas with populations of 65,000 or more and multiyear averages for smaller geographic areas. The ACS is also intended to replace the long-form Census questionnaire.

¹⁰[GAO-03-227](#).

In 2000, the Bureau initiated a program aimed at documenting Bureau-wide protocols designed to ensure the quality of data it collected and released. Because this effort is still in its early stages, we could not assess it. However, Bureau officials believe that the program is a significant first step in addressing the Bureau's lack of data quality guidelines. As the Bureau develops its protocols further, it will be important that they be well documented, transparent, clearly defined, consistently applied, and properly communicated to the public.

Conclusions

Throughout the 1990s, the Bureau went to great lengths to improve response rates to the 2000 Census in general, and participation of Hispanics in particular. Although the unique contributions of the individual components of the Bureau's efforts cannot be determined, the mail response rate was similar to the 1990 level, and the Bureau's preliminary data suggest that the 2000 Census count of Hispanics was an improvement over the 1990 count. However, the counts of Hispanic subgroups do not appear to have been improved and, in fact, there is concern that some of these subgroup counts may be less accurate than the 1990 counts. Moreover, the Bureau's experience in simplifying the questionnaire in part by removing the examples of the Hispanic subgroups shows the challenge the Bureau faces in trying to improve one component of the census count without adversely and unintentionally affecting other aspects of the census count. In light of these findings, it will be important for the Bureau to continue with its planned research on how best to enumerate Hispanic subgroups.

The Bureau's release of Hispanic subgroup numbers raised questions about the quality of the reported data and the Bureau's decision to report these data as a part of its release of the SF-1 data. Although the specific questions about the Hispanic subgroup data differed from those identified in our review of the Bureau's efforts to collect and report data on the homeless and others without conventional housing, a common cause of both sets of problems was the Bureau's lack of agencywide guidelines for its decisions on the level of quality needed to release data to the public. As we recommended in our report on homeless counts, the Bureau needs to develop well-documented guidelines that spell out how to characterize any limitations in the data, and when it is acceptable to suppress these data. The Bureau should also ensure that these guidelines are documented, transparent, clearly defined, consistently applied, and properly communicated to the public.

Recommendations for Executive Action

To ensure that the 2010 Census will provide public data users with more accurate information on specific Hispanic subgroups, we recommend that the Secretary of Commerce ensure that the Director of the U.S. Census Bureau implements Bureau plans to research the Hispanic question, taking steps to properly test the impact of the wording, format, and sequencing on the completeness and accuracy of the data on Hispanic subgroups and Hispanics overall. In addition, as we also recommended in our companion report on the homeless and others without conventional housing, we recommend that the Bureau develop agencywide guidelines governing the level of quality needed to release data to the public, when and how to characterize any limitations, and when it is acceptable to delay or suppress data.

Agency Comments and Our Evaluation

The Secretary of Commerce forwarded written comments from the U.S. Census Bureau on a draft of this report (see app. I). The Bureau agreed with our conclusions and recommendations and, as indicated in the letter, is taking steps to implement them. However, it expressed several general concerns about our findings. The Bureau's principal concerns and our response are presented below. The Bureau also suggested minor wording changes to provide additional context and clarification. We accepted the Bureau's suggestions and made changes to the text as appropriate.

The Bureau took exception to our findings concerning the adequacy of its data quality guidelines noting that it "conducted the review of the data on the Hispanic origin population using standard review techniques for reasonableness and quality." We do not question the Bureau's commitment to presenting quality data. Rather, our point is that the Bureau needs to translate its commitment to quality into well documented, transparent, clearly defined guidelines to provide a basis for consistent decision making on the level of quality needed to release data to the public, and on when and how to characterize any limitations. During our review, Bureau officials, including the Associate Director for Methodology and Standards, told us that the Bureau had few written guidelines, standards, or procedures related to the quality of data released to the public.

A second general concern expressed by the Bureau dealt with our characterization of problems with the Hispanic subgroup counts. The Bureau said that the data met an acceptable level of quality because they accurately reflect what people reported and therefore cannot be characterized as erroneous. We agree with the Bureau on this specific

point. However, we take a broader view of data quality. Specifically, we believe that questions about the accuracy of the Hispanic subgroup data must also take into account problems that the respondents had in understanding the meaning of the question. The Bureau challenged our assertion that the wording of the question “confused” some respondents, preferring to say that some respondents may have “interpreted” the question wording, instructions, and examples differently than expected. We agree with the Bureau that additional research will be required to understand the extent of this problem. Nevertheless, we believe there is sufficient evidence from the Bureau’s subsequent research and from analysis of trends in the data to support our concerns about the accuracy of Hispanic example subgroup counts in the 2000 Census.

As agreed with your office, unless you publicly announce 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, Christopher Miller, Elizabeth Powell, Timothy Wexler, Ty Mitchell, Benjamin Crawford, James Whitcomb, Robert Parker, and Michael Volpe.



Patricia A. Dalton
Director
Strategic Issues

Comments from the Department of Commerce



THE SECRETARY OF COMMERCE
Washington, D.C. 20230

NOV 27 2002

Ms. Patricia A. Dalton
Director, Strategic Issues
General Accounting Office
Washington, DC 20548

Dear Ms. Dalton:

The Department of Commerce appreciates the opportunity to comment on the General Accounting Office draft report entitled *Decennial Census: Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement*. The Department's comments on this report are enclosed.

Sincerely,

A handwritten signature in black ink, appearing to read "D. L. Evans", is positioned above the printed name of the Secretary of Commerce.

Donald L. Evans

Enclosure

Comments from the U.S. Department of Commerce
U.S. Census Bureau

U.S. General Accounting Office draft report entitled *Decennial Census: Methods for Collecting and Reporting Hispanic Subgroup Data Need Refinement*

General Comments on the Report

While the U.S. Census Bureau agrees with the General Accounting Office's (GAO) recommendations in this report, we take exception to the GAO's suggestion that decisions regarding the release and characterization of data on detailed Hispanic origin groups were based on anything other than our consistent commitment to clearly presenting data that conform with established guidelines for data quality.

The Census Bureau conducted the review of the data on the Hispanic-origin population using standard review techniques for reasonableness and quality. These quality decisions are based upon comparisons to independent work and findings from experts outside the Census Bureau, other surveys, analysis of trends, literature reviews, and consultations with experts (both public and private) throughout the decade. When data do not meet an acceptable level of quality, the Census Bureau will consider various options for modifying publication plans and determine the most appropriate way to disseminate these data. With regard to the data on detailed Hispanic-origin groups, we determined that it was entirely appropriate to present these data in our data products. Those products accurately reflect what people reported on their forms or to a census enumerator.

Also, it should be noted that data obtained from the census question on ethnicity are the result of self-identification and, therefore, should not be characterized as "erroneous" (as compared with results from the 1990 census), nor should they be subject to suppression, except under highly unusual circumstances that are clearly not present here. Additional research will be required to understand the extent to which the question wording and format influenced some people to report a more general response rather than a specific Hispanic ethnicity. But it is important to acknowledge that, in Census 2000, more people of Hispanic ethnicity may have preferred to identify generally as Hispanic, Spanish, or Latino than in previous censuses. Furthermore, to understand the reasons for differences in totals for detailed Hispanic groups between the 1990 and 2000 censuses, results from both censuses must be analyzed. For example, the use of examples in 1990 may have influenced more people to report in the groups that were listed and fewer to report in other detailed groups. Alternatively, those whose groups were not listed may have reported more generally as Hispanic. The appropriate conclusion is that the results of the two censuses are different, not that one is more accurate than the other.

The Census Bureau is undertaking a review of its data quality guidelines, independent of the GAO's findings in this report.

Comments on the Text of the Report

1. Section: *Highlights* page: “A key factor behind the Bureau’s release of the questionable subgroup data was its lack of adequate guidelines governing the quality needed before making data publicly available.”

Comment: As noted above, the Census Bureau conducts its data reviews using standard review techniques for reasonableness and quality. When data do not meet an acceptable level of quality, the Census Bureau will consider various options for modifying publication plans and determine the most appropriate way to disseminate the data. When we publish the data, we note any deficiencies and cautions in a section of the product documentation called “User Updates” and/or on our Web site.

2. Section: Page 3, second paragraph, third and sixth sentences: “Bureau evaluations conducted following the census show that dropping the examples of Hispanic subgroups confused some respondents and produced less-than-accurate subgroup data.”
“ . . . because of respondents’ confusion over the wording of the question, the subgroup data could be misleading.”

Comment: In some cases, respondents may have *interpreted* the question wording, instructions, and examples differently than we might have expected. This does not mean the respondents were confused, but would indicate that additional research and testing will be required to more fully understand these interactions.

3. Section: Page 6, first paragraph, first sentence: “Although not required by OMB standards, Hispanic subgroup data are also used for many of these same purposes.”

Comment: The sentence should be revised as follows: “Although not required by OMB standards or federal legislation, Hispanic subgroup data”

4. Section: Page 9, heading: “Efforts to Simplify Questionnaire Led Bureau to Delete List of Hispanic Subgroups.”

Comment: Heading should read “Efforts to Simplify Questionnaire Led Bureau to Delete Examples of Hispanic Subgroups,” because we use three specific subgroups (Mexican, Puerto Rican, and Cuban) as response categories.

5. Section: Page 15, first paragraph, last part of the first sentence: “. . . it did not specifically design any tests to determine the impact of the changes on the quality of Hispanic subgroup data.”

Comment: The Census Bureau did look at the impact of changes on Hispanic subgroups. However, the sample size in the test was not large enough to detect statistically

significant differences for the Hispanic subgroups that comprise the “Other Spanish/Hispanic/Latino” population. Additionally, the test was not designed to detect the impact of each change to the question separately.

6. Section: Page 15, first bullet: “1992 National Census Test, which was a field test of the 2000 Census questionnaire;”

Comment: This test was not a test of the actual questionnaire(s) used in Census 2000. The bullet item should be revised to indicate that this was a test of potential Census 2000 questionnaires.

7. Section: Page 15, last part of the last sentence: “. . . but did not conduct any specific evaluations of the quality of the 1990 Hispanic subgroup data.”

Comment: The Census Bureau did examine the data for those Hispanic subgroups that were response categories on the 1990 census questionnaire.

8. Section: Page 17, first paragraph, third sentence: “For the first time, the Bureau released data on Hispanic subgroups as a part of its release of SF-1 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.”

Comment: This sentence appears to be erroneous and should be deleted. The Census Bureau released data on detailed Hispanic subgroups in the sample 1990 summary files. (The data for detailed subgroups were coded only from the sample forms in 1990.) We conducted extensive testing of the wording for this question, including the instructions and examples, prior to Census 2000. Further, our review of these data from Census 2000 did not indicate any evidence of an “error” (for example, a data processing or data collection error) that would have precluded their dissemination. Subsequent evaluations have shown that additional research is needed to study how individuals choose the responses they write in.

9. Section: Page 18, first paragraph, last sentence: “Thus, while the Bureau reported what respondents checked off on their questionnaires, because of respondents’ confusion over the wording of the question, the 2000 subgroup data could be misleading.”

Comment: Same comment as in Item 2 above: In some cases, respondents may have *interpreted* the question wording, instructions, and examples differently than we might have expected. This does not mean the respondents were confused, but would indicate that additional research and testing will be required to more fully understand these interactions.

10. Section: Page 22, entire page.

Comment: Regarding the issues addressed on this page, we refer the reader to our general comments on the report and also to our response to Recommendation 2.

Responses to GAO Recommendations

Recommendation 1: The Census Bureau should implement its plans to conduct further research on the Hispanic question, taking steps to properly test the impact of any changes on the quality of data on Hispanic subgroups and Hispanics overall.

Census Bureau Response: The Census Bureau concurs with this recommendation. This work is underway as part of the research and testing program for the 2010 census.

Recommendation 2: The Census Bureau should develop agency-wide protocols that provide guidelines for bureau 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 to delay or suppress the data.

Census Bureau Response: The Census Bureau concurs with this recommendation. In order to continue to maintain its long tradition of producing high-quality data, the Census Bureau has asked the Methodology and Standards Council to review our statistical and quality guidelines for surveys and censuses and codify them in one place.

Related GAO Products

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

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

2000 Census: Coverage Evaluation Matching Implemented As Planned, but Census Bureau Should Evaluate Lessons Learned. [GAO-02-297](#). Washington, D.C.: March 14, 2002.

2000 Census: Best Practices and Lessons Learned for a More Cost-Effective Nonresponse Follow-Up. [GAO-02-196](#). Washington, D.C.: February 11, 2002.

2000 Census: Coverage Evaluation Interviewing Overcame Challenges, but Further Research Needed. [GAO-02-26](#). Washington, D.C.: December 31, 2001.

2000 Census: Analysis of Fiscal Year 2000 Budget and Internal Control Weaknesses at the U.S. Census Bureau. [GAO-02-30](#). Washington, D.C.: December 28, 2001.

2000 Census: Significant Increase in Cost Per Housing Unit Compared to 1990 Census. [GAO-02-31](#). Washington, D.C.: December 11, 2001.

2000 Census: Better Productivity Data Needed for Future Planning and Budgeting. [GAO-02-4](#). Washington, D.C.: October 4, 2001.

2000 Census: Review of Partnership Program Highlights Best Practices for Future Operations. [GAO-01-579](#). Washington, D.C.: August 20, 2001.

Decennial Censuses: Historical Data on Enumerator Productivity Are Limited. [GAO-01-208R](#). Washington, D.C.: January 5, 2001.

2000 Census: Information on Short- and Long-Form Response Rates. [GAO/GGD-00-127R](#). Washington, D.C.: June 7, 2000.

GAO's Mission

The General Accounting Office, the 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 the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

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. General Accounting 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

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Official Business
Penalty for Private Use \$300**

Address Service Requested

<p>Presorted Standard Postage & Fees Paid GAO Permit No. GI00</p>
--

