

149248

United States General Accounting Office

GAO

Testimony

Before the Subcommittee on Census, Statistics,  
and Postal Personnel  
Committee on Post Office and Civil Service  
House of Representatives

For Release on Delivery  
Expected at  
1:00 p.m. EDT  
Wednesday  
April 14, 1993

FEDERAL DATA  
COLLECTION

Measuring Race and Ethnicity  
Is Complex and Controversial

Statement of William M. Hunt  
Director, Federal Management Issues  
General Government Division



149248

057238 / 149248

**FEDERAL DATA COLLECTION: MEASURING RACE AND  
ETHNICITY IS COMPLEX AND CONTROVERSIAL**

**SUMMARY STATEMENT  
WILLIAM M. HUNT  
DIRECTOR, FEDERAL MANAGEMENT ISSUES**

GAO's work analyzing the experience of the decennial census and other federal agency measurement of race and ethnicity revealed a complex and controversial statistical issue.

The 1990 Decennial Census experience underscores the importance of developing consensus on race and ethnic questions as soon as possible to avoid disruption to 2000 census planning. However, delays in some scheduled activities mean the Bureau faces a tight time frame in which to improve the race and ethnic questions for the 2000 census.

In planning for the 1990 census, disagreement over the format of the race question led to a protracted debate and a last-minute decision on the final format that was contrary to the Bureau's initial recommendations. Therefore, the efforts expended during the Bureau's testing and consultation programs proved unsuccessful in achieving significant changes. Bureau evaluations suggested that the data from the 1990 race and Hispanic origin questions are generally of high quality, although some problems associated with collecting data on Hispanics continued to plague the 1990 census as they did the 1980 census.

In 2000 census planning, the Bureau faces a tight time frame in which to improve the race and ethnic questions because of delays in some scheduled activities. Advisory committees for minority populations, important to obtaining buy-in for test objectives, need to be rechartered. Because of research delays, GAO is concerned that the Bureau will not be prepared for 1995 tests of race and ethnic questions. GAO encourages Bureau census reform efforts that explore alternatives for producing better intercensal data on important population characteristics, such as race and ethnicity.

In looking more broadly at other federal agency data collection efforts, GAO's work focused on whether agencies adhered to standard race and ethnic definitions under OMB's Statistical Policy Directive No. 15, not on whether the definitions themselves were appropriate. Federal agencies generally use consistent race and ethnic definitions. Some potential for noncompliance exists because there is no statutory requirement or resources for OMB monitoring data collection projects through the collection and reporting phases. Even when the classification requirements of OMB Directive No. 15 are followed, inconsistent use of race and ethnic terms can occur when different methods are used to identify an individual (self versus observer identification) or when agencies use external sources, such as state-provided data.

Mr. Chairman, Mr. Petri, and Members of the Subcommittee:

I am pleased to be here today to report on the collection of race and ethnic<sup>1</sup> data. Race and ethnic data are among the most complex and controversial data collection efforts undertaken by the federal government. As the Subcommittee requested, I will address the Census Bureau's experience in measuring race and ethnicity in the 1990 census, the status of 2000 census research in this area, and federal agencies' experience in measuring race and ethnicity under the Office of Management and Budget's (OMB) current classification system. My comments today are based on two recently issued GAO reports<sup>2</sup> and our current monitoring of 2000 census redesign efforts as requested by this Subcommittee.

The 1990 Decennial Census experience underscores the importance of developing consensus on the race and ethnic questions as soon as possible to avoid disruption to 2000 census planning.

However, delays in some scheduled activities mean the Bureau faces a tight time frame in which to improve the race and ethnic questions for the 2000 census. In our examination of federal agency data collection efforts, our work focused on agency adherence to standard race and ethnic definitions under OMB's

---

<sup>1</sup>The terms "ethnic" and "ethnicity," as used in this testimony, refer to Hispanic origin.

<sup>2</sup>Census Reform: Early Outreach and Decisions Needed on Race and Ethnic Questions (GAO/GGD-93-36), Jan. 28, 1993); and Federal Data Collection: Agencies' Use of Consistent Race and Ethnic Definitions (GAO/GGD-93-25, Dec. 15, 1992).

Statistical Policy Directive No. 15, not on whether the definitions themselves were appropriate. We found that federal agencies generally used consistent race and ethnic definitions.

1990 DECENNIAL CENSUS EXPERIENCE  
WITH COLLECTION OF RACE AND ETHNIC DATA

The 1990 Decennial Census experience provided several valuable lessons. It demonstrated the need to develop consensus on the race and ethnic questions as early in the decade as possible. It also showed that the Bureau needs to continue efforts to improve race and ethnic data quality.

Early Consensus Needed on Race and Ethnic Questions

The race and ethnic questions in the census have been a source of controversy. Race and ethnicity are not objectively definable characteristics, which make measurement difficult. Moreover, the Bureau anticipates that as minority populations grow, including biracial and multiethnic children of intermarriages, the Bureau will be subject to increasing pressure from new groups and subgroups for identification on the census form and in census data products.

In the 1990 census, the Bureau failed to build consensus on its recommended version of the race question in spite of a special testing and consultation program. The focus of the debate was how Asian and Pacific islanders were to be represented in the race question. The shorter version of the question recommended by the Bureau contained a single space for Asian and Pacific islanders to write in their specific groups, as shown in figure 1. By relying on write-in responses, this version of the race question replaced the separate categories for Asian and Pacific islander subgroups that were used in 1980.

On the basis of its testing program, the Bureau concluded that this short version of the race question was likely to produce data on the Asian and Pacific islander population that was at least as good as other test versions of this question. In one test, the short version captured more persons who wrote in Asian and Pacific islander groups, such as Amerasian, Pakistani, and Sino-Vietnamese, which are not represented by the detailed categories of the longer version of the race question. However, in response to congressional direction and pressures from the Asian and Pacific islander community, the Bureau made the last-minute decision to include in the 1990 census a version of the race question with detailed Asian and Pacific islander categories (for example, Chinese, Japanese, Asian Indian), as shown in figure 2.

**Figure 1: Short Version of the Race Question**

<p><b>4. Race</b></p> <p>Fill ONE circle for each person.</p> <p style="text-align: center;"><b>AND</b></p> <p>If "Asian or Pacific Islander," print one group.</p> <p>If "Indian (Amer.)," print enrolled or principal tribe</p> <p>If "Other race," print race.</p> <p>Report the race the person considers him/herself to be.</p>	<p>1 <input type="radio"/> White</p> <p>2 <input type="radio"/> Black or Negro</p> <p>3 <input type="radio"/> Asian or Pacific Islander (Print one group, for example: Chinese, Asian Indian, Hawaiian, Laotian, Vietnamese, etc.)</p> <p style="text-align: center;">-----</p> <p>4 <input type="radio"/> Indian (Amer.) Print enrolled or principal tribe</p> <p style="text-align: center;">-----</p> <p>5 <input type="radio"/> Eskimo</p> <p>6 <input type="radio"/> Aleut</p> <p>7 <input type="radio"/> Other race—Print race</p> <p style="text-align: center;">-----</p>
--	---

Source: Census Bureau

**Figure 2: 1990 Census Race Question**

<p><b>4. Race</b></p> <p>Fill ONE circle for the race that the person considers himself/herself to be.</p> <p>If Indian (Amer.), print the name of the enrolled or principal tribe. -----&gt;</p> <p>If Other Asian or Pacific Islander (API), print one group, for example: Hmong, Fijian, Laotian, Thai, Tongan, Pakistani, Cambodian, and so on. -----&gt;</p> <p>If Other race, print race. -----&gt;</p>	<p><input type="radio"/> White</p> <p><input type="radio"/> Black or Negro</p> <p><input type="radio"/> Indian (Amer.) (Print the name of the enrolled or principal tribe.)</p> <p style="text-align: center;">-----</p> <p><input type="radio"/> Eskimo</p> <p><input type="radio"/> Aleut</p> <p style="text-align: center;"><u>Asian or Pacific Islander (API)</u></p> <p><input type="radio"/> Chinese                      <input type="radio"/> Japanese</p> <p><input type="radio"/> Filipino                        <input type="radio"/> Asian Indian</p> <p><input type="radio"/> Hawaiian                       <input type="radio"/> Samoan</p> <p><input type="radio"/> Korean                         <input type="radio"/> Guamanian</p> <p><input type="radio"/> Vietnamese                   <input type="radio"/> Other API</p> <p style="text-align: center;">-----</p> <p><input type="radio"/> Other race (Print race)</p>
---	---

Source: Census Bureau

The relatively late formation of advisory committees for minority populations was a major roadblock to the planning process. These committees were established by the Department of Commerce in 1986, after the objectives for the 1986 tests on race and ethnicity had already been decided. The results of the 1986 tests were central to the subsequent debate about the format of the race question. Bureau staff said these committees did not have sufficient time to reach out to the minority communities in order to gain buy-in for the test objectives. Representatives of the Asian and Pacific islander community also said that the Bureau had not solicited their participation in the early phases of redesigning the race question.

As a result, the final format of the race question was decided late in the decade after protracted debate and was contrary to the Bureau's initial recommendations. With this relatively late change in the race question format, the efforts expended during the Bureau's testing and consultation programs proved unsuccessful in achieving significant changes to the race question used in the 1980 census.

#### Improvements Needed in Data Quality

Bureau evaluations suggested that the data from the 1990 race and Hispanic origin questions are generally of high quality. However, these evaluations also suggested that some data quality

problems associated with collecting data on Hispanics that confronted the Bureau in the 1980 census continue to plague the Bureau. Given the promise the shorter version of the race question showed in tests leading up to the 1990 census as a way to measure the Asian and Pacific islander population, the Bureau should also continue to test the use of shorter write-in questions.

#### Collecting Data on Hispanics Poses Difficulties

Some Hispanics, particularly foreign-born Hispanics, equate their "Hispanicity" with race. According to Bureau evaluations, these Hispanics have difficulty classifying themselves by the standard race categories of White, Black, Asian and Pacific islander, Indian (American), Eskimo, and Aleut. Approximately 40 percent of Hispanics responded "other race," a special category permitted in the census race item. The Bureau determined that these persons were Hispanic because they indicated they were Hispanic in response to the Hispanic origin question or in the write-in space provided in the race question.

The growth of the "other race" category appears to be largely a reflection of the difficulties the race item poses for Hispanics. Almost 10 million people, close to 4 percent of the total U.S. population of nearly 248.7 million, were reported in the "other race" category in the 1990 census. This represents a 45-percent



growth between 1980 and 1990, a much faster growth rate than that of the total population--which was less than 10 percent.

According to the Bureau, 97.5 percent of the almost 10 million persons who were reported in the "other race" category in the 1990 census were persons of Hispanic origin. The Bureau, however, maintains the capability of redistributing persons in the "other race" category to one of the standard race groups. Persons in the "other race" category were assigned the same race of the nearest person outside of the household who gave an identical response to the Hispanic origin question.

Not only does the race question cause problems for some Hispanics, the Hispanic origin question also poses problems for non-Hispanics. Some non-Hispanics, having already responded to the race question, skip over the Hispanic origin question when they should indicate that they are not of Hispanic origin. The Bureau allocates responses for those persons who skip the Hispanic origin question. (Allocation refers to a method in which a characteristic is assigned to a respondent who failed to answer the question or who failed to answer the question in a complete or legible fashion).

The Hispanic origin question has the highest allocation rate (10 percent) of all population questions asked on the census short form. The allocation rates for all the other questions

were all lower than 3 percent. Allocation is an efficient, statistically based approach to filling in missing data compared to costly follow-up procedures. High allocation rates may not be problematic if Hispanic origin characteristics are allocated in an unbiased fashion. Preliminary analysis showed no evidence of bias, but Bureau staff said that further study is needed before any conclusions can be drawn.

#### Questions Asking for Write-in Responses Show Promise

Testing of the race question before the 1990 census showed that questions asking for write-in responses may produce data as good as that from questions with prespecified categories. Some evidence suggests that questions asking for write-in responses hold advantages in capturing groups not represented by the detailed categories of long versions of the race question. This advantage is particularly important given the likelihood that the Bureau will be subject to increasing pressure from other population groups for identification on the census form.

Replacing separate categories with a space for write-in response can also shorten the census questionnaire. Bureau research showed that shorter questionnaires can lead to better response rates.<sup>3</sup> Of course, the benefits of write-in questions must be

---

<sup>3</sup>Census Reform: Questionnaire Test Shows Simplification Holds Promise (GAO/T-GGD-92-59, July 1, 1992).

weighed against their higher processing costs. The Bureau's ability to process write-in data in a timely fashion should also be considered. Whatever the overall benefits of moving away from detailed race and ethnic categories, the Bureau must try to convince different race and ethnic communities that these benefits outweigh any real or perceived disadvantages or it may face opposition to its plans.

#### STATUS OF THE 2000 CENSUS PLANNING FOR RACE AND ETHNICITY

The Bureau has developed a plan for research and development of race and ethnic questions for the 2000 census. The plan shows an awareness of the major issues on race and ethnic questions needing attention.

As I noted when I testified before this Subcommittee last month, we are concerned that the Bureau is not making necessary progress in its preparations for the 1995 census test, the critical field test for the 2000 census.<sup>4</sup> Likewise, delays in important planning activities for race and ethnicity may impair the Bureau's preparations for the 1995 test. In order to take full advantage of the 1995 testing opportunity, the Bureau needs to begin consultations, complete critical 1990 census evaluations,

---

<sup>4</sup>Decennial Census: Fundamental Reform Jeopardized by Lack of Progress (GAO/T-GGD-93-6, Mar. 2, 1993).

and establish test objectives. However, the Bureau has experienced delays in each of these areas.

#### Minority Advisory Committees Need to Be Rechartered

The 1990 census experience revealed the importance of coming to a broad agreement with race and ethnic groups early in the process. As events are unfolding now, the Bureau may experience similar problems in the 2000 census. The charters for the minority advisory committees for the 1990 census expire at the end of fiscal year 1993, and no date for rechartering new committees for the 2000 census has been established.

Although the Bureau has begun informal consultations, Bureau staff said that minority advisory committees provide the stature and mechanisms needed for obtaining formal recommendations for its research and testing activities. If the minority advisory committees are not rechartered, the Bureau needs to develop another means of gaining broad understanding and acceptance of its proposals.

### Research Delays Could Lead to Lack of Preparedness for 1995 Test

The Bureau initially hoped to complete 1990 census evaluations by fiscal year 1991. However, Bureau staff told us that the final, comprehensive evaluation of the 1990 census will not be available until this summer. According to Bureau staff, pressures for issued the 1990 race and ethnic data products issued have constrained the Bureau's ability to do its evaluation. The Bureau must set priorities for its evaluations to ensure that critical data and insights are available when needed to support decision making. During preparations for the 1990 census, we repeatedly expressed concern about the pace of the Bureau's evaluation efforts. The Bureau's decision to allow a longer time frame to complete the race and ethnicity evaluations will allow little time for preparing for the census tests.

The Bureau's first field test of the race and ethnic questions was originally scheduled for fiscal year 1993, but the Bureau's latest version of its plan shows the test is now scheduled for fiscal year 1994. If resources permit, the Bureau hopes to conduct small-scale testing on the race and ethnicity questions in fiscal year 1994 and build special tests into the 1995 census test. Bureau staff said that although the schedule is tight, they believe they will have adequate time to incorporate 1994 test results into the 1995 test.

Alternatives for Providing More Current  
Race and Ethnic Data Need to Be Explored

A once-a-decade census of the nation's population cannot and should not be the primary vehicle to provide information on such rapidly changing phenomena as the racial and ethnic makeup of the U.S. population. For example, between 1980 and 1990 the Hispanic population in the United States grew just over 50 percent, from approximately 14.6 million to 22.4 million persons. Significant change is even more dramatic at subnational levels--state, county, city, census tract, and block. For example, in Arlington County, VA, the number of Hispanics grew by more than 150 percent, from almost 9,000 to more than 23,000 in that same 10-year period. Surveys now taken between censuses do not adequately capture these changes at such lower levels.

Current and accurate data are important and have practical implications. For example, under a Department of Health and Human Services initiative entitled "Healthy People 2000," the National Center for Health Statistics (NCHS) is responsible for tracking a wide range of health objectives by race and ethnic group at the national, state, and local levels. Between censuses, NCHS has difficulty obtaining geographically detailed population counts by race and ethnic group to produce these health statistics.

To try to meet this growing demand for more timely race and ethnic data, the Bureau is expanding its Intercensal Demographic Estimates program. However, Bureau staff said intercensal population estimates for race and ethnic groups are difficult to produce because the administrative records on which they are largely based generally capture less race and ethnic detail than the census and surveys. The Bureau is exploring options for producing better intercensal data on important population characteristics, such as race and ethnicity.

#### FEDERAL AGENCY EXPERIENCE WITH COLLECTION OF RACE AND ETHNIC DATA

Federal agencies generally used consistent race and ethnic definitions. We found problems, however, in agency data collection efforts when they use data based on different methods for identifying an individual's race or ethnicity or data based on state or local government administrative records not controlled by federal collection policy.

#### Standard Categories Exist to Guide Collection of Federal Race and Ethnic Data

Federal race and ethnic data collection activities are governed by OMB Statistical Policy Directive No. 15, which has been in effect since 1980. Directive No. 15 attempts to achieve

consistency in federal statistical data by providing standard classifications for race and ethnicity. The directive defines four racial groups (American Indian or Alaska native, Asian or Pacific islander, Black, and White) and one ethnic group (Hispanic) based on geographical or cultural rather than scientific origins. The Census Bureau was granted an exception to Directive No. 15 in order to offer the category of "other race" for those persons who do not identify with any of the race categories provided.<sup>5</sup>

The Paperwork Reduction Act requires OMB to approve federal data collection plans. Through this approval process, OMB reviews federal data collection instruments and methodologies for consistency with the requirements of Directive No. 15. However, OMB is not required under that Act to monitor agency data collection projects through the collection and reporting phases.

---

<sup>5</sup>Agencies must be able to aggregate all race and ethnic data to these basic categories, although they are also free to collect more detailed information within these categories, as is done in the decennial census. The Bureau maintains a separate file that reclassifies persons in the "other race" category into standard race categories to meet the needs of federal and state agencies and researchers.



Federal Agencies Generally Follow Directive No. 15

On the basis of a review of data collection instruments used to produce the major statistical reports of eight federal agencies,<sup>6</sup> we found that the standard definitions of Directive No. 15 were generally followed. While our review did not include every federal data collection effort, we examined 33 major surveys and corresponding reports of 8 agencies that, according to OMB, collect significant amounts of race and ethnic data.

The only example of noncompliance that we discovered in this review was a Federal Bureau of Investigation (FBI) data collection effort that did not include arrest information on persons of Hispanic origin. OMB officials were unable to clarify for us whether the data collection instrument submitted by the FBI for review contained the Hispanic origin category. They told us, however, that their approval of this data collection effort without a Hispanic origin category was an oversight and would be corrected in future surveys.

---

<sup>6</sup>Our sample included survey documents, data collector and respondent instructions, and reports from eight federal agencies. These sample documents were typical of the collection instruments supporting the agencies' major data systems. These agencies were the Department of Justice (Bureau of Justice Statistics), Department of Education (National Center for Education Statistics and Office for Civil Rights), Department of Health and Human Services (National Center for Health Statistics), Department of Labor (Bureau of Labor Statistics), Department of Defense (Defense Manpower Data Center), Department of Housing and Urban Development, Equal Employment Opportunity Commission, and Small Business Administration.

The potential exists, however, for other instances in which federal data collection efforts do not conform to standard race and ethnic definitions. OMB reviews forms and methodology, but it is not required by the Paperwork Reduction Act to monitor agency data collection projects through the collection and reporting phases.<sup>7</sup>

### Different Methods of Identification Cause

#### Inconsistent Use of Race and Ethnic Terms

Inconsistent use of race and ethnic terms can occur when different methods are used to identify a person's race or ethnicity, even when the classification requirements of Directive No. 15 are followed. While most federal data collection efforts determine race or ethnicity through self-identification,<sup>8</sup> some rely on observer identification. The preferred method for categorizing individuals (self versus observer identification) is not stated in Directive No. 15.

---

<sup>7</sup>OMB does not currently have the staff to perform this function centrally, and OMB officials told us the costs of such a review by OMB would be prohibitive. Another alternative to decrease the probability of noncompliance would be to create a more stringent monitoring effort by departmental forms clearance officers to determine conformity through the collection and reporting phases.

<sup>8</sup>Self-identification can also refer to the identification of a person's race or ethnic origin by family members.

A January 1992 study of vital statistics databases of birth and death certificates for infants, for example, disclosed that 43 percent of Asian and American Indian infants were classified by race differently at death from how they were classified at birth.<sup>9</sup> According to this study, many of these infant deaths were classified as white, thereby overstating white mortality somewhat, but greatly understating Asian and American Indian infant mortality. Officials from the Centers for Disease Control and Prevention (CDC) believe that death certificate preparers were classifying infants by observation and not by following CDC's policy of asking a family member to identify the appropriate race or ethnic background of the infant. CDC staff said they were aware of the problem before this study was released, and CDC has already made efforts to better enforce its own policy.

#### State-Provided Data Cause Inconsistent

#### Use of Race and Ethnic Terms

Whereas federal agencies must follow Directive No. 15 for race and ethnicity in designing their data collection instruments, state and local government data, and the administrative records on which these data are generally based, are not controlled by

---

<sup>9</sup>Robert A. Hahn, Joseph Mulinare, and Steven M. Teutsch, "Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in U.S. Infants," The Journal of the American Medical Association (Jan. 8, 1992), pp. 259-263).

federal collection policy. We found that as a result, state data were not consistently reported, causing problems for federal agencies that depend on these data. The impact can be demonstrated with Department of Justice data. Based on state-furnished data, Hispanic populations in the corrections system could be significantly understated because of the large percentage of the population for which ethnicity is unknown. Nineteen states did not report the ethnic composition of their probation populations in 1989. Three of these states (Illinois, California, and New Jersey) are among those in which the greatest numbers of Hispanics reside.

Inconsistencies also exist in the manner in which states classify persons of mixed race and ethnicity. A 1992 survey of 800 school districts by Education's Office for Civil Rights found that about 30 percent of them use a special category to classify people of mixed race and ethnicity. In addition, some states determine a student's race or ethnicity by that of the mother; others use that of the father. These state practices conflict with Directive No. 15, which calls for persons of mixed race or ethnic background to be reported in a standard category "which most closely reflects the individual's recognition in his community." Inconsistencies between states themselves, combined with those between federal and state policy, could become a greater problem in the future if the number of multiracial and multiethnic families grows as trends over the last 20 years suggest it will.

- - - - -

In summary, Mr. Chairman, measuring race and ethnicity is complex and controversial. The 1990 census experience proved the importance of timely planning and research coordinated with consensus-building efforts. There is no doubt that inadequacies and inconsistencies in federal race and ethnic data exist due to operational problems, but we found that agencies generally used consistent race and ethnic definitions in designing their data collection instruments and methodologies.

This concludes my prepared statement. My colleagues and I would be pleased to answer any questions.

(243046)

---

  

---

**Ordering Information**

**The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.**

**Orders by mail:**

**U.S. General Accounting Office  
P.O. Box 6015  
Gaithersburg, MD 20884-6015**

**or visit:**

**Room 1000  
700 4th St. NW (corner of 4th and G Sts. NW)  
U.S. General Accounting Office  
Washington, DC**

**Orders may also be placed by calling (202) 512-6000  
or by using fax number (301) 258-4066.**