

*U.S. DEPARTMENT OF COMMERCE*  
*Office of Inspector General*

---



**PUBLIC  
RELEASE**

*BUREAU OF THE CENSUS*

*Actions to Address the Impact on the  
Accuracy and Coverage Evaluation  
of Suspected Duplicate Persons  
in the 2000 Decennial Census*

*Inspection Report No. OSE-13812/March 2001*

*Office of Systems Evaluation*





UNITED STATES DEPARTMENT OF COMMERCE  
The Inspector General  
Washington, D.C. 20230

MAR 29 2001

**MEMORANDUM FOR:** William G. Barron, Jr.  
Acting Director, Bureau of the Census

**FROM:** Johnnie E. Frazier *Johnnie Frazier*

**SUBJECT:** *Actions to Address the Impact on the Accuracy and Coverage Evaluation of Suspected Duplicate Persons in the 2000 Decennial Census*  
Final Inspection Report No. OSE-13812

Our September 2000 report on the Master Address File raised concerns about potential duplicate addresses, which could result in persons being counted more than once in the 2000 Decennial Census.<sup>1</sup> As the census progressed, the bureau became increasingly concerned about address duplication and undertook a process to resolve the problem, with the goal of producing more accurate apportionment numbers. A consequence, however, was that some person records eventually included in the apportionment numbers were not included in key operations of the Accuracy and Coverage Evaluation (A.C.E.).

This report documents our evaluation of the bureau's handling of suspected duplicate enumerations as they may have affected the A.C.E. Specifically, the bureau had not resolved the status of suspected duplicate enumerations when the census results were needed for use in the tightly scheduled A.C.E. As a result, the bureau decided to exclude the suspected duplicate person records from the A.C.E. matching and follow-up operations.

Once the bureau decided which of the suspected duplicate person records to remove from the census, the remaining records were incorporated into the A.C.E. final calculations using the bureau's plan for dealing with late census data.<sup>2</sup> According to bureau officials, using this plan should introduce little or no error into the A.C.E. results if certain assumptions implicit in the approach hold true. Our major concern was whether these assumptions were valid for the reinstated person records.

<sup>1</sup>*A Better Strategy is Needed for Managing the Nation's Master Address File*, Office of Inspector General, OSE-12065, September 2000.

<sup>2</sup>*Treatment of Late Census Data for Accuracy and Coverage Evaluation Estimation*, DSSD Census 2000 Procedures and Operations Memorandum Series #Q-17, Memorandum to Howard Hogan, Division Chief, Decennial Statistical Studies Division, from Donna Kostanich, Assistant Division Chief for Sampling and Estimation, Decennial Statistical Studies Division, December 29, 1999.

We shared our concerns with senior bureau officials during meetings held in January and February 2001. We recommended that they provide an analysis of the likely impact of their method for handling the reinstated person records. We also recommended that they ensure the impact was considered in the bureau's process for reviewing the A.C.E. and census results for making a recommendation to the Secretary of Commerce about whether or not the census counts should be statistically adjusted for redistricting. The bureau was responsive to our recommendations. It has been preparing a chapter for the report supporting its recommendation regarding whether or not to adjust entitled, *Accuracy and Coverage Evaluation Survey: Effect of Excluding "Late Census Adds,"* and has already discussed the issue in the reports and analyses supporting its recommendation. Our report documents the issues and concerns discussed in those meetings and actions agreed to or taken by the bureau.

## BACKGROUND

The Constitution mandates that a census of the nation's population be taken every 10 years for the purpose of congressional apportionment. Census data is also used for state redistricting and the allocation of federal funds. Data from the decennial census provides official, uniform information about the nation's people and their social, demographic, and economic characteristics.

In counting the population, the bureau relies on its Master Address File (MAF) to identify where people reside. The address file built from the MAF for use in the census is termed the decennial MAF. The quality of addresses maintained by the MAF directly affects the accuracy, completeness, and cost of the decennial. To overcome the historic undercoverage of housing units for the 2000 census, the bureau devised an inclusive approach for retaining addresses in its file and also used a wider variety of sources to obtain these addresses. In our September 2000 report on the MAF, we found that this inclusive approach resulted in an unknown number of duplicate addresses, which could cause duplicate enumerations.<sup>3</sup>

During a decennial census, the Census Bureau attempts to count and gather information about every resident in the country. However, in any decennial, some enumerations that should have been excluded are included, and some portion of the population that should have been included is missed. The first source of error leads to an overcount; the second, to an undercount. Every census for which the effect of these errors has been systematically measured has shown a net undercount—that is, the number of residents missed is greater than the number counted in error. Studies going back to 1940 also show a net differential undercount for minority population groups, meaning that minorities are missed at a higher rate than the white population. For 2000, the bureau planned to measure coverage and reduce the differential undercount through the A.C.E.<sup>4</sup> The A.C.E. is a statistical methodology based on an independent sample of the

---

<sup>3</sup>A Better Strategy is Needed for Managing the Nation's Master Address File.

<sup>4</sup>"Accuracy and Coverage Evaluation; Statement of the Feasibility of Using Statistical Methods to Improve the Accuracy of the Census 2000," *Federal Register*, Department of Commerce, Bureau of the Census, June 20, 2000.

population, which is then compared or matched with the census records to determine persons missed and erroneous enumerations. The A.C.E. uses a process termed “dual system estimation” to estimate the net undercount of various demographic subgroups of the population (called post-strata) and to calculate the “coverage correction factors” which can be used to adjust the census counts.

In designing and conducting the census and A.C.E., the bureau strives to balance two types of error. One type, sampling error, occurs only in the A.C.E. Sampling error, which is quantified by sampling variance, occurs because a sample is used to represent a population. The other type, nonsampling error, occurs in both the census and the A.C.E. For nonsampling error, the bureau is particularly concerned with systematic errors or biases. The most serious source of bias in the census is coverage error, which results from people being missed or from erroneous enumerations (including duplicate enumerations). The most notable example of this is the net undercount, including the differential undercount. Bias can occur in the A.C.E. as a result of errors in matching, errors in accounting for missing information, or other systematic errors in collection or processing. Bias caused by systematically missing individuals in both the census and A.C.E. is termed correlation bias.<sup>5</sup>

The issue of whether statistical sampling could be used in the census was brought before the United States Supreme Court, which ruled in January 1999 that such sampling could not be used for congressional apportionment purposes. The Court did not prohibit the use of statistical sampling for other demographic purposes, including redistricting.

The Secretary of Commerce is required by law to report redistricting data to the states within one year after the decennial census date. For the 2000 census, the Secretary must report such data to the states by April 1, 2001. By the end of February, the bureau had completed its internal assessment of whether the A.C.E. should be used to adjust the census counts. An Executive Steering Committee for A.C.E. Policy (ESCAP), consisting of 12 senior career bureau officials, was responsible for reviewing census and A.C.E. data and preparing a report for the bureau director recommending whether the adjusted or unadjusted census data should be used. The ESCAP’s March 1, 2001, report recommended that the unadjusted data be released as the Census Bureau’s official redistricting data because the information available at that time was insufficient to conclude that the adjusted census data would be more accurate. The Acting Director of the Census Bureau concurred with the ESCAP recommendation, and on March 6, 2001, the Secretary decided that the unadjusted data would be used.

---

<sup>5</sup>“Accuracy and Coverage Evaluation; Statement of the Feasibility of Using Statistical Methods to Improve the Accuracy of the Census 2000,” *Federal Register*, Department of Commerce, Bureau of the Census, June 20, 2000.

## OBJECTIVES, SCOPE, AND METHODOLOGY

The objective of this evaluation was to determine whether the bureau's methodology for handling the reinstatement of 2.4 million person records into the decennial census minimized the impact on the accuracy and reliability of the A.C.E. This issue arose during our review of the process used in the A.C.E. for automated person matching.

During January and February 2001, we briefed six members of the ESCAP regarding our concerns and our recommendations for addressing them. Members whom we briefed included the Associate Director for Decennial Census, who was the ESCAP chair, and the Assistant Director for Decennial Census. Because these officials have already acknowledged our concerns and agreed with our recommendations, we are issuing this report in final. This report documents the issues and the bureau's approach to addressing them. Bureau officials were given the opportunity to review the report for technical accuracy. Where appropriate, their comments have been incorporated.

We conducted our fieldwork between November 2000 and March 2001. We did not evaluate the bureau's process for removing duplicates or thoroughly review the analysis regarding these issues in the documentation supporting the bureau's recommendation.

This evaluation was conducted in accordance with the *Quality Standards for Inspections* issued by the President's Council on Integrity and Efficiency and was performed under the authority of the Inspector General Act of 1978, as amended, and Department Organization Order 10-13, dated May 22, 1980, as amended.

## OBSERVATIONS AND CONCLUSIONS

### **The Bureau Addressed Our Concerns About the Effect of Reinstatements on the A.C.E.**

Because of its concerns about address duplication, the bureau monitored the number of housing unit addresses by comparing demographic benchmarks to the decennial MAF at several stages of the decennial. By June 2000, some counties continued to show higher than expected coverage when compared to the benchmarks. The bureau decided to conduct fieldwork in some of these counties to investigate the problem. Using the April 2000 version of the decennial MAF, bureau staff examined approximately 20,000 addresses and found that over 13 percent were either duplicated (11.6 percent) or nonexistent (1.5 percent). The bureau was sufficiently concerned about address duplication and the resulting potential duplicate enumerations that in July 2000 it began devising possible methods for identifying duplicates.<sup>6</sup>

Special bureau operations involving address and person matching were developed and implemented for the purpose of identifying and removing duplicate enumerations that remained on the decennial file at the end of the census. The process was termed the Duplicate Housing Unit Operations. Its design was not specified before the census, but rather was developed during

---

<sup>6</sup>*Overview of the Duplicate Housing Unit Operations*, Fay Nash, U.S. Census Bureau, November 7, 2000.

the census in response to the problem of duplication. The bureau acknowledged that this operation made mistakes of both exclusion and inclusion, but believed it was necessary to avoid seriously impairing the accuracy of the apportionment numbers.<sup>7</sup>

Through the rules of this process, the bureau was able to identify 6 million persons in 2.4 million housing units as potential duplicates. However, according to the bureau, there was not enough time to resolve the status of suspected duplicates before the census file was needed to begin the tightly scheduled A.C.E. matching and follow-up operations. In consultation with the bureau official responsible for conducting the A.C.E., the bureau decided to provide to the A.C.E. person matching operation a version of the census file that did not include the suspected duplicated person records. The bureau's rationale was that removing the suspected duplicates from the A.C.E. matching process was better than retaining them because the model used to calculate the adjustment works better with more accurate A.C.E. sample and Census 2000 counts. In addition, the bureau had a process for dealing with late census data and felt that any reinstatements could be handled through this process.<sup>8</sup>

After analyzing the suspected duplicates, the bureau removed approximately 1.4 million housing units and 3.6 million persons from the census and reinstated approximately 1 million housing units and 2.4 million persons into the census. Our review of the rules that the bureau used tends to suggest that the reinstatements comprised primarily duplicate persons residing in nonduplicated housing units.<sup>9</sup> The bureau's goal was to delete only duplicate housing units, and therefore duplicate persons in nonduplicated housing units were reinstated. The bureau worked under the assumption that these duplicate persons were similar to those residing in housing units known to be occupied but with unknown persons. In this instance, the bureau would have added persons to households occupied by unknown persons through a statistical process termed whole person imputation. In the case of the reinstated persons residing in nonduplicated housing units, the bureau reasoned that these persons served as the imputations. The bureau also noted that all of the reinstated person records represented less than 1 percent of the total population and that it expected such records to be represented in the A.C.E. in about the same proportion.

The bureau's memorandum documenting its process for incorporating late census data into the A.C.E. evaluated four options for handling late additions and recommended the option that the bureau deemed to be the most comprehensive and believed would avoid giving the appearance of data manipulation.<sup>10</sup> According to the bureau, this option would not cause the expected value of

---

<sup>7</sup>*Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*, U.S. Census Bureau, March 1, 2001, p. 12.

<sup>8</sup>*Treatment of Late Census Data for Accuracy and Coverage Evaluation Estimation*.

<sup>9</sup>*Specification for Reinstating Addresses Flagged as Deletes on the Hundred Percent Census Unedited File (HCUF)*, DSSD Census 2000 Procedures and Operations Memorandum Series #D-11, Memorandum for Susan Miskura, Chief Decennial Management Division, from Howard Hogan, Chief, Decennial Statistical Studies Division, November 7, 2000, and *Results of Reinstatement Rules for the Housing Unit Duplication Operations*, Memorandum for Preston J. Waite, Assistant to the Associate Director for the Decennial Census, from Susan M. Miskura, Chief, Decennial Management Division, November 21, 2000, Attachment 2.

<sup>10</sup>*Treatment of Late Census Data for Accuracy and Coverage Evaluation Estimation*, p. 5.

the results of dual system estimation to be biased if the reinstated person records met certain assumptions.<sup>11</sup> Important among these is that the match rate for correctly enumerated reinstated person records is the same as for the rest of the A.C.E.<sup>12</sup>

Bureau officials stated that they did not have enough time to test whether the reinstated person records conformed to the assumptions before they had to make the recommendation on whether or not to adjust the census data; consequently, they could not demonstrate that their approach caused little or no bias in the dual system estimates. The bureau acknowledged that the reinstatements would contribute to variance in the estimates, thereby increasing the risk of obscuring whether the A.C.E. estimates are more accurate than the census counts. Bureau officials told us that although the specific variance caused by the reinstatements would not be known by the ESCAP at the time it had to make its recommendation on whether or not to adjust, this additional variance would be reflected in the total variance, an indicator that would be known and that the ESCAP would consider in its deliberations.

Because the bureau did not plan to analyze the reinstated person records in any depth before the recommendation on whether or not to adjust and did not have previous analysis documenting the likely characteristics of such records, we became concerned about the bureau's ability to defend its assumptions about the effects of the reinstatements on the recommendation. Our concerns were heightened because, as noted previously, our review of the rules that the bureau used tends to suggest that the reinstatements comprised primarily duplicate persons residing in nonduplicated housing units and because bureau officials told us that they suspected the reinstatements were geographically clustered.<sup>13</sup> Because of these issues, we recommended that the bureau provide written analysis to support its position that the reinstatements would introduce little or no bias into the dual system estimate.

A related concern that we also raised was the number and impact of whole person imputations in the census. This concern is related because these imputations are treated in the same way as the reinstatements in the dual system estimate. At the end of the census, the bureau reported that there had been approximately 5.7 million people added to the census count by imputation, more than 2.5 times the number added in 1990. The imputations, along with the 2.4 million reinstatements, totaled 8.1 million persons, about 2.9 percent of the population count. Given the importance of an accurate census, we emphasized to senior census officials that the bureau should be able to demonstrate whether the reinstatements, along with the imputations, exacerbated overcoverage or masked undercoverage.

The Associate Director for Decennial Census agreed to ensure that the ESCAP would consider our concerns in making its recommendation. In addition, the bureau official responsible for the A.C.E., also an ESCAP member, agreed to conduct sensitivity testing of the assumptions and

---

<sup>11</sup>Expected value is an average of estimates derived from all possible samples.

<sup>12</sup>*Accuracy and Coverage Evaluation: Data and Analysis to Inform the ESCAP Report*, Howard Hogan, U.S. Census Bureau, March 1, 2001, p. 52.

<sup>13</sup>According to the bureau's demographic analysis, erroneous addresses occurred with greater frequency in certain counties.

document the results. This analysis has been completed and is undergoing internal review by the bureau. It will be published as a chapter of the report supporting the ESCAP's recommendation entitled, *Accuracy and Coverage Evaluation Survey: Effect of Excluding "Late Census Adds."*

Importantly, the issues that we raised pertaining to the reinstatements and imputations were considered in evaluation reports supporting the recommendation of the ESCAP, as well in the ESCAP report itself. The principal analysis and discussions are presented in two reports, *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy* and *Accuracy and Coverage Evaluation: Data and Analysis to Inform the ESCAP Report*.

According to these reports, the ESCAP reviewed the evaluation report data, as well as other information, and concluded that the key assumptions underlying the methodology for including the reinstated person records in the A.C.E., such as match ratios for the correct enumerations, could be expected to hold, although they would not hold perfectly. The ESCAP believed that the measures available for assessing the effects of sampling variance and correlation bias would include the effects of the treatment of late additions and whole person imputations.<sup>14</sup> However, the ESCAP was concerned that geographic clustering of the reinstated person records and imputations might have increased another type of error and further reviewed this effect.<sup>15</sup> The ESCAP concluded that the data did indicate some degree of geographic clustering within post-strata and noted that it took these findings into consideration when reviewing the results of the adjustments.

Although we have not had the opportunity to thoroughly review the analysis supporting the bureau's decision regarding error added by reinstatements, we believe that the bureau's actions were responsive to our concerns and recommendations in the short time frame available for reviewing the data and making the recommendation on whether or not to adjust. The bureau plans to perform further evaluation studies to assess the impact of the reinstatements and imputations.

We believe that such studies are appropriate to better understand both the impact on dual system estimation in the 2000 Decennial Census and the impact of similar late data requirements in future censuses and surveys since dual system estimation is an important bureau methodology for measuring data quality. Further, to help avoid similar problems with the address file in the future, we reaffirm the recommendations presented in our MAF report regarding approaches for addressing housing unit overcoverage and undercoverage.

Should you have any questions regarding this report, please contact me at (202) 482-4661 or Judith Gordon, Assistant Inspector General for Systems Evaluation, at (202) 482-5643. We would like to thank Census Bureau headquarters staff for the cooperation and courtesies extended to us during our review.

cc: Lee Price, Acting Under Secretary for Economic Affairs

---

<sup>14</sup>*Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*, p. 26.

<sup>15</sup>This type of error is referred to as synthetic error and is related to the distribution of the measured net undercount to local areas and demographic subgroups.