U.S. DEPARTMENT OF COMMERCE
Office of Inspector General

BUREAU OF THE CENSUS

Selected Aspects of Census 2000 Accuracy and Coverage Evaluation Need Improvements Before 2010

Final Inspection Report No. IG-14226-2/March 2002

PUBLIC RELEASE

Office of Audits and System Evaluations

TABLE OF CONTENTS

EXECUTIVE SUMMARY i

INTRODUCTION 1

OBJECTIVES, SCOPE, AND METHODOLOGY 2

BACKGROUND 3

FINDINGS AND CONCLUSIONS 6

I. The A.C.E. P-Sample May Have Missed People Residing in Retirements Homes 6

II. A.C.E. Requirements Process Should Be Improved 8

A. A.C.E. and Census Joint Requirements Need to Be Better Planned 9
B. Computer Matching Requirements Should Be More Completely Specified . . . 10
C. Laptop System Provided Minimal Benefit for PFU Users 12

III. Use of Automated Tools Facilitated Person Matching, but Better Testing, Documentation, and Quality Assurance Needed 13

A. Computer Matcher Software Tested but Results Not Documented 13
B. Output Not Adequately Reviewed After Changes to Forms Printing Software . 14

IV. PFU Quality Assurance Program Needs to Be Automated 15

RECOMMENDATIONS 16

APPENDIX A. Terms and Acronyms
APPENDIX B. Bureau Response

EXECUTIVE SUMMARY

On December 28, 2000, the Census Bureau released the nation's population count: 281,421,906. This number resulted from the information collected when approximately 77 million households returned their census forms and approximately 42 million households were subsequently visited by enumerators to collect census information. Census data is used for many critical purposes including congressional apportionment, state redistricting, and the allocation of federal funds.

During a decennial census, the Census Bureau attempts to count and gather information about every resident in the country. However, in any census 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 errors leads to an overcount; the second, to an undercount. To measure these coverage errors, the Census Bureau conducted a separate and independent quality check survey known as the Accuracy and Coverage Evaluation (A.C.E.). A.C.E. uses a statistical methodology based on obtaining an independent sample of the population. The A.C.E. population sample is matched with the census population to identify persons missed or incorrectly counted in the census. Correctly matching A.C.E. person records with census person records is important because even small mistakes can affect the bureau's ability to measure the over and undercount precisely.

The issue of whether statistical sampling could be used to adjust the census was brought before the United States Supreme Court. In January 1999, the Supreme Court ruled that sampling could not be used for congressional apportionment purposes but was silent on its use for state redistricting and the allocation of federal funds.

The deadline for providing states with redistricting data was April 1, 2001. On March 1, 2001, the Acting Director of the Census Bureau, in consultation with the bureau's Executive Steering Committee for A.C.E. Policy, recommended using the unadjusted counts rather than the statistically adjusted counts for redistricting. The primary reason given by the bureau for arriving at this conclusion was the disparity in population estimates from two different sources: the recently completed A.C.E. and the bureau's demographic benchmark analysis. Since the decision on redistricting, the bureau has analyzed this disparity and has further recommended using the unadjusted counts for allocating federal funds as well.

While A.C.E. involved many different operations, our review focused on two of them, person computer matching and person follow-up (PFU). Person computer matching, which started October 5, 2000, and completed on October 25, 2000, and PFU, which started on October 19, 2000, and completed on November 21, 2000, matched the majority of individual census and A.C.E. person records. Computer matching involves using a software program that, with the aid

of human experts, assigns a match status to the records. PFU is a field operation that gathers additional information for persons assigned a status of non-match or possible match.

The OIG conducted this review to determine how well systems supporting person computer matching and PFU field operations worked and to assess the efficiency and effectiveness of these operations. The bureau is evaluating the 2000 decennial processes to make future improvements. Thus, the lessons learned from the 2000 A.C.E. are important inputs to planning the next decennial. Our review reached the following conclusions:

- **The A.C.E. Independent Sample May Have Omitted Residents of Retirement Homes**. A.C.E. design included people living in housing units, such as houses, apartments, or mobile homes, and excluded people living in group quarters, such as nursing homes, dormitories, and prisons. However, for the elderly, A.C.E. and census definitions of group quarters versus housing units differed which, we believe, may have resulted in the A.C.E. not measuring census coverage of retirement home residents. This difference came to our attention in discussions with PFU field staff who identified two retirement homes as examples of census housing units that were excluded from the P-sample. In addition, we found a third example while reviewing some PFU forms. Because the A.C.E. may have systemically failed to measure census coverage of the elderly population living in retirement homes, we believe that the bureau should review A.C.E. data to determine whether these omissions were systemic and, if so, what impact this would have on estimates of the elderly population. (See page 6.)

- **A.C.E. Requirements Process Should Be Improved.** Many systems supporting various headquarters and field operations were defined and developed for A.C.E. Although the bureau met the overall schedule for this complex program, we did find some areas where improvements could be made. Specifically, A.C.E. requirements relying on census input need more planning among divisions, computer matching requirements should be more clearly defined and documented, and laptop systems provided to field supervisors should better meet user needs. To address these issues, the bureau should ensure that A.C.E. requirements are carefully planned, defined, and documented, and that they support user needs. (See page 8.)

- **Use of Automated Tools Facilitated Person Computer Matching, but Better Testing, Documentation, and Quality Assurance Needed.** The bureau's use of the computer matcher, a trusted, much-used software tool, and a questionnaire printing system enhanced the efficiency of bureau matching and production of PFU questionnaires. We could not assess the adequacy of the testing of the computer matcher as used for A.C.E. person matching because the bureau did not document the test cases and results. Although the computer matcher was not newly developed software, it was being applied differently and

the results of its testing should have been documented. With respect to the questionnaire printing system used for PFU, errors occurred in production that we believe resulted from not sufficiently reviewing the accuracy of printed questionnaire data after last minute software changes. To better ensure data accuracy and completeness, the bureau should document testing and perform thorough quality assurance. (See page 13.)

- **PFU Quality Assurance Program Needs to Be Automated.** The bureau conducted quality assurance on selected PFU questionnaires to detect interviewing problems, including data falsification. We commend the bureau for adding a quality assurance process for PFU in response to our recommendation made during the decennial dress rehearsal. However, according to personnel interviewed at several regional offices, tracking the quality assurance workload was paper- and labor-intensive and the process could be better planned. Bureau officials told us that the PFU operation was intended to be conducted using a computer-assisted person interview system similar to the system used in the person interview operation, but the bureau decided during dress rehearsal that not enough resources were available to develop and appropriately test such a system for PFU. As a result, the bureau made the decision to use the questionnaire printing system instead. For 2010, we believe that quality assurance should be integrated as part of an automated PFU operation. (See page 14.)

We recommend that the Acting Director of the Census Bureau direct senior management responsible for evaluating the A.C.E. and planning the quality check activities of the 2010 census to take the necessary actions to:

1. Evaluate whether residents of retirement homes were systemically omitted from the A.C.E. independent sample and if so, what impact this had on estimates of the elderly population.

2. Ensure that definitions for identifying housing units, special places, and group quarters, as well as other definitions as appropriate, are consistent for both the census and the A.C.E.

3. Document and clearly communicate requirements when planning systems that have to meet requirements of both the census and the A.C.E.

4. Ensure that computer matching requirements are fully developed and documented.

5. Ensure that computer systems are developed and modified in accordance with rigorous, documented system and software engineering standards that, at a minimum, address requirements specification, design and development, testing, documentation, and quality assurance.

6. Integrate the PFU quality assurance process as part of an automated PFU operation to ensure efficiency.

In responding to our draft report, the bureau generally agreed with all but the first recommendation. The bureau stated that all A.C.E. evaluations requiring field work had been completed and that additional evaluation work to satisfy this recommendation would be difficult because such housing units were not characterized as being retirement homes in the census data. Furthermore, two years after Census Day, residents may no longer live in the housing units in question. The bureau also stated that it is considering including group quarters in the coverage measurement survey for Census 2010, which should reduce the potential for error caused by confusing housing units with group quarters. We still maintain that not knowing how much missed coverage of retirement homes impacted this population's net undercount calculation diminishes confidence in the A.C.E. results for this population group. We believe that analyzing the results of matching the P-sample and E-sample to identify high concentrations of elderly within A.C.E. clusters that were not matched in the P-sample would indicate whether this omission was systemic. If the omission is found to be systemic, then the bureau could estimate the impact such an omission would have on estimates of populations of the elderly. The bureau agreed with the second recommendation that the definitions for housing units, special places, and group quarters should be consistent for A.C.E. and census.

The bureau, although agreeing with the third recommendation, also stated that a very tight schedule after the Supreme Court's decision on the use of sampling for producing the apportionment counts made it difficult to document and communicate requirements for systems supporting both the census and the A.C.E. While we recognize that planning was delayed because of the sampling decision, the examples we cited were clearly more of a problem with documenting and communicating all known requirements not impacted by the sampling decision. Also, while agreeing that, in general, the bureau needs to document requirements more completely, the bureau claims that the automated matching requirements, which the fourth recommendation addressed, were fully developed and understood by those who conducted this operation. We do not believe that was the case, as the criteria for including records as possible matches were ambiguous.

The bureau's response to our fifth recommendation, to ensure that systems are modified in accordance with rigorous documented system and software engineering standards, referred to its response for the third and fourth recommendation and, therefore, was not specific to the intent of this recommendation. We are asking the bureau to do more than improve requirements definition;

as our findings pointed out and other reviews have found, the bureau needs to improve all phases of its system and software engineering processes, not just requirements definition. Finally, with regard to the sixth recommendation, we recommend that the quality assurance process be integrated as part of a fully automated PFU operation for the 2010 census. The bureau agreed that the PFU quality assurance process is important and stated that it will be studied during the planning for the next decennial.

The bureau's response letter is found in Appendix B.

INTRODUCTION

The Constitution mandates that a census of the nation's population be taken every 10 years for the purpose of congressional apportionment. Data from the decennial census provides official, uniform information on the social, demographic, and economic characteristics of the nation's population. During a decennial census, the Census Bureau attempts to count and gather information for every resident in the country. However, in any decennial census, some enumerations that should have been excluded are included, and some portion of the population that should have been included is missed. For 2000, the bureau measured coverage and planned to adjust the census counts through the Accuracy and Coverage Evaluation (A.C.E.).

A.C.E. uses a statistical methodology based on obtaining an independent sample of the population, which is then matched with the census records to try to identify persons missed or erroneously counted. The A.C.E. methodology uses a process called dual system estimation to estimate the net undercount of various demographic subgroups of the population (called post-strata). These estimates, if considered to be sufficiently reliable, can then be used to adjust the census counts accordingly.

The issue of whether statistical sampling could be used to adjust the census was brought before the United States Supreme Court. In January 1999, the Supreme Court ruled that sampling could not be used for congressional apportionment purposes but was silent on the issues of state redistricting and the allocation of federal funds.

The deadline for providing states with redistricting data was April 1, 2001. On March 1, 2001, the Acting Director of the Census Bureau, in consultation with the bureau's Executive Steering Committee for A.C.E. Policy, recommended using the unadjusted counts rather than the statistically adjusted counts for redistricting. The primary reason given by the bureau for arriving at this conclusion was the disparity in population estimates from two different sources: the recently completed A.C.E. and the bureau's demographic benchmark analysis. Since the decision on redistricting, the bureau has analyzed this disparity and has further recommended using the unadjusted counts for allocating federal funds.

The Office of Inspector General has completed a review of person computer matching and person follow-up (PFU), two operations that were integral to matching A.C.E. and census person records. Correctly matching A.C.E. persons with census persons is important because matching errors in even a small percentage of cases can affect the precision of the undercount or overcount estimates. In addition, the bureau is evaluating the 2000 decennial processes to make future improvements. Thus, the lessons learned from the 2000 A.C.E. are important inputs to planning the next decennial.

OBJECTIVES, SCOPE, AND METHODOLOGY

The purpose of this evaluation was to assess the person computer matching and PFU operations. Person computer matching, which started October 5, 2000, and was completed on October 25, 2000, and PFU, which was started on October 19, 2000, and completed on November 21, 2000, were critical in matching the majority of A.C.E. and census person records. For computer matching, we assessed how the bureau ensured that matching software performed to an expected level of accuracy. For the PFU operation, we assessed its efficiency and effectiveness, and sought to identify any problems that may have affected the quality of the A.C.E.

Our methodology for reviewing software primarily involved reviewing bureau requirements and testing documentation and interviewing the responsible bureau officials. Specifically, to assess whether the requirements specification was appropriate, sufficiently defined, and clearly communicated, we obtained an understanding of matching requirements by attending bureau matching training at the National Processing Center (NPC) in Jeffersonville, Indiana and observing computer matching at bureau headquarters in Suitland, Md. We also reviewed specifications to assess whether they (1) were written in a clear, unambiguous fashion and followed a software engineering standard; (2) defined an acceptable threshold for errors; and (3) appropriately addressed concerns raised by bureau evaluations, including the ability to match inaccurate, incomplete, and conflicting data.

To determine if a comprehensive set of test cases was used to test the requirements, we determined whether the bureau had test plans, test cases, and test results and whether the testing addressed diverse response data to ensure that the matching software could reliably handle a wide array of potential input data and that the accepted error threshold was met.

We also evaluated the PFU operation in the field at selected locations. We reviewed policies and procedures and examined selected PFU questionnaires and other relevant documentation. We also spoke with NPC staff about the PFU operation. In addition, we observed PFU interviews in Philadelphia, Pennsylvania, Winchester, Virginia, and Las Vegas, Nevada, and we interviewed field staff, including interviewers and field supervisors, at these locations concerning the efficiency and effectiveness of the PFU operation. Finally, we interviewed bureau headquarter officials in Suitland, Maryland and regional personnel in Seattle, Denver, Charlotte, and Philadelphia.

This evaluation was conducted in accordance with the Inspector General Act of 1978, as amended, and the *Quality Standards for Inspections*, March 1993, issued by the President's Council on Integrity and Efficiency.

BACKGROUND

Many Census Bureau offices were involved in A.C.E. The Decennial Statistical Studies Division (DSSD) was responsible for design and management of all A.C.E. operations. These included questionnaire design, development of requirement specifications for matching software and operations, software testing and implementation, and designing and monitoring quality control of matching operations. DSSD also worked with the field and other offices in testing computer-assisted person interview instruments. Twelve Field Division A.C.E. regional offices were responsible for staffing, training, and conducting the PFU operation. Field Division headquarters provided oversight to the regional offices and developed training materials, budget and schedule estimates, and progress reports. NPC conducted clerical matching operations and provided the forms and maps for conducting PFU. The Technologies Management Office designed and installed system hardware and application software, created management reporting systems, and provided laptops to the field. The Decennial Systems and Contracts Management Office (DSCMO) performed headquarters data processing for the decennial and was responsible for processing census enumeration data. Other offices have been involved in conducting evaluations and providing overall management support.[1]

The bureau conducted A.C.E. on a sample of about 315,000 housing units in various areas of the fifty states, the District of Columbia, and Puerto Rico. The sample consisted of block clusters,[2] which are geographic areas that on average contained about 30 housing units. The bureau selected sample block clusters from areas that would represent the nation as a whole and considered such characteristics as size of block cluster, American Indian Reservation, type of enumeration area, and race/ethnicity composition.[3] Once the block clusters were selected, the bureau developed an address list of housing units in selected clusters independent of the census address listing activity. This was followed by a housing unit matching operation between the A.C.E. and census listing with a follow-up operation in the field to resolve discrepancies. The resulting A.C.E. address list was used to conduct computer-assisted person interviews, on the telephone or in person, at each known housing unit within the sample. Interviewers asked questions to determine whether residents should have been counted at the sample address on Census Day or whether the residents counted on Census Day had since moved. Person records

---

[1]DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-TL-06, October 18, 1999, Draft 2 and Census 2000 A.C.E. Person Followup - Program Master Plan, May 22, 2000.

[2]Block clusters contain one or more census blocks, which are small geographic areas bounded by physical features and jurisdictional boundaries and are used to organize data collection.

[3]*Accuracy and Coverage Evaluation: The Design Document*, DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, January 11, 2000.

collected during A.C.E. are referred to as the P-sample. The A.C.E. also used an E-sample that included persons enumerated by the census in housing units from the A.C.E. block clusters.

To obtain the data used to estimate the net undercount for each post-strata, the bureau compared the P-sample records to the census data for households in the same vicinity. Census person records needed for this comparison were selected from the Hundred-percent Census Unedited File (HCUF). Generally speaking, the primary objective of this process was to determine for each A.C.E. block cluster if each record matched a record in the other sample.

The first phase of this person record comparison is computer matching and employs the SRD[4] Record Linkage System. The version used for A.C.E. was similar to the one used in the 1990 decennial. To use the SRD Record Linkage System, bureau officials had to designate which fields within each A.C.E. and census record pair were to be compared, the method by which they would be compared, and the scores that would be assigned based on the level of agreement or disagreement between data contained in each field. Fields compared included first name, last name, middle initial, month of birth, date of birth, age, gender, and unit designation. Before the records were submitted to the matcher, first names, street names and suffixes, and apartment subunit designations were standardized. For example, "First Street" was converted to "1st St," and "apartment A" was converted to "apartment 1." Other data, such as first name, race, and relationship, was edited into standard designations as well.

As census data became available, the bureau divided 702,465 P-sample records and the census person records into 22 match groups consisting of an average of about 31,930 P-sample person records. Each match group was submitted to the matcher separately, and matching occurred only for person records within the same block cluster. The person records were matched in two passes. For the first pass, only persons whose surname started with the same character were compared; this restriction was lifted for the second pass, where records unmatched after the first pass were compared. Certain fields, such as date of birth and middle initial, had to match exactly to add their agreement weights to the matching score. Names were compared using a "string comparator," which allows for typographical and scanning errors while scoring according to the degree to which the data agreed. The SRD Record Linkage System paired each record from the P-sample with a census person record and sorted the records in descending order with respect to score. For example, in the first pass scores ranged from 13.0566 (indicating a definite match) to −3.8929 (indicating a definite non-match).

Use of the SRD Record Linkage System also involved expert human intervention to delineate which linked records were matches (linked records that identified the same person with a high

[4]SRD stands for the bureau's Statistical Research Division, which created the computer matcher. See William E. Winkler, *Advanced Methods for Record Linkage,* Bureau of the Census, September 1994.

degree of certainty); possible matches (linked records that identified the same person with less certainty); and non-matches (linked records that were clearly not the same person and needed to be unlinked). After the first pass, the expert matcher running the job would inspect the linked records. The expert supplied a high cutoff, ranging from a score of 3.151 to 2.618, to the computer for linked records that were matches. The expert supplied a lower cutoff, ranging from a score of 1.393 to .6472, for linked records that were possible matches. Any records with scores below the lower cutoff would be unlinked and matched again in the second pass. After the second pass produced the file of linked records, the expert matcher would repeat the procedure, again supplying the computer a cutoff score for matches and a cutoff score for possible matches. When the second pass of computer matching was completed, the match group results were sent to NPC's clerical matching operation. Computer matching averaged match rates of 69.6 percent for person records coded as matches, 9.9 percent as possible matches, and 19.9 percent as not matched.[5]

Clerical matchers reviewed possible matches, judging whether to change them to matches or non-matches. Clerical matchers also reviewed unmatched P-sample and E-sample records, attempting to find more matches or possible matches. The records that the clerical matchers were unable to match with certainty and E-sample records that remained unmatched were sent to the PFU operation.

For PFU, NPC printed a questionnaire for each household with at least one person requiring follow-up. Each questionnaire was shipped to one of the 12 A.C.E. regional offices around the country. The regions prepared the questionnaires and assigned them to field interviewers, who completed each questionnaire at the address printed on it. Once completed, questionnaires were checked for accuracy and completeness by a field supervisor and then were returned to the regional office. Field supervisors were assigned laptop systems to manage the PFU workload.

Upon receipt, the regional office performed an edit of each questionnaire to ensure that all entries were legible, that the skip pattern[6] had been followed, and that the field supervisor had initialed the form. The edit also determined if the PFU case needed a quality assurance reinterview. The questionnaires were then shipped to NPC, where they were reviewed by clerical matchers. With the new information obtained during PFU, the clerical matchers assigned final codes to the non-matched P- and E-sample persons to indicate if they should have been counted within the block cluster on census day.

---

[5]A small percentage of the person records, .6 percent, were not submitted to matching because they lacked sufficient information. The bureau used the numbers of these excluded records in calculating the matching percentages, which effectively decreased them.

[6]Skip patterns are directions printed on the questionnaire that tell interviewers which questions to ask next depending on the interviewee's previous answer.

FINDINGS AND CONCLUSIONS

I. The A.C.E. P-Sample May Have Missed People Residing in Retirements Homes

The A.C.E. P-sample may have systemically omitted residents of retirement homes. We believe these people were excluded because the A.C.E. staff responsible for completing the independent A.C.E. housing unit list may have excluded retirement homes thinking that they did not meet the definition of a housing unit. Because this problem started with the housing unit list used to create the P-sample, we are concerned that A.C.E. may have systemically missed people residing in retirement homes, which could have contributed to errors in estimating the net undercount. We believe that the bureau should assess whether retirement homes were systemically omitted and if so, determine the impact of missing residents of these homes.

By design, A.C.E. did not include group quarters on the independent housing list used to construct the P-sample. Group quarters were defined as special places where occupants share certain residential facilities, such as a cafeteria, and included dormitories, nursing homes, orphanages, prisons, and halfway houses. Instead, the A.C.E. P-sample included person data only for people living in a housing unit, which was defined as a house, an apartment, a group of rooms, or a single room whose occupants live separately from other persons and have direct access from the outside of the building or through a common or public hall.

We found that A.C.E. and census guidance for identifying group quarters for the elderly differed and may have led to the exclusion of elderly people residing in retirement homes from the P-sample. A.C.E. address listing guidance instructed field staff *not* to include "homes for the aged" on the address list because they were considered group quarters. However, the census definition of group quarters used the term "nursing homes or convalescent homes" rather than the term "homes for the aged," making it confusing about how to treat retirement homes. According to bureau officials, any retirement home fitting the definition of a housing unit is not a group quarter and should have been included in the address listing and person data collection activities for the A.C.E.

E-sample records that did not match P-sample records were sent to PFU to be confirmed as correctly enumerated by the census. We discovered the problem with retirement homes in our discussions with PFU field staff members in different parts of the country responsible for conducting these interviews who told us that they were surprised to be conducting interviews in entire retirement homes that they believed were group quarters that should not have been included in the E-sample. The staff identified two retirement homes where this occurred, and we identified another during our review of some of the PFU forms. Upon reviewing the addresses of these

housing units using A.C.E. and census data, we confirmed that the A.C.E. P-sample did not include them or the 188 people residing in them. These persons were coded as correctly enumerated, correctly counted by the census, and omitted from the P-sample. Because this problem can be traced back to how a P-sample housing unit was defined, we believe that this problem could be systemic and are concerned about the impact on estimating the net undercount because the ratio of missed people in this group of the population would not be measured.

We discussed the correctly enumerated status for the people living at these housing units with bureau officials who provided two explanations. One, A.C.E. listers may have overlooked the housing units, or, two, census could have misclassified the address as a housing unit when it really should have been a group quarter. However, we believe that the P-sample missed these housing units because the A.C.E. group quarter definition, as related to housing for the elderly, was broader than the census definition. Thus, the A.C.E. listers considered the three retirement homes as group quarters, not housing units, and therefore residents were not included in the P-sample.

We also discussed missed P-sample coverage by A.C.E. and the discrepancy in the group quarter definition with members of the Executive Steering Committee on A.C.E. Policy.[7] They stated that it is not always clear as to whether an elderly residential living situation is a group quarter, and they agree that the Census and A.C.E. definitions for special places and group quarters need to be consistent for 2010. Furthermore, bureau officials contend that A.C.E. P-sample coverage does not have to be perfect, just independent, and dual system estimation works even if some of the P-sample is missed. We understand that the P-sample does not need to be perfect; however, we are concerned that the P-sample omissions we have identified are specific to a certain group of people — the elderly population living in retirement homes. We believe the bureau should evaluate how extensively this population was omitted from the P-sample and the extent to which the coverage measured for this group would change as calculated by dual system estimation. With an increasingly aging population, we believe this analysis is necessary.

The bureau is evaluating the quality of the A.C.E. to improve it for the future. As part of its work, it is evaluating whether definitional problems led to group quarter classification errors. However, this evaluation will not help determine for this decennial if elderly citizens residing in retirement homes were systemically excluded from the P-sample. Without such an assessment, the Bureau cannot determine the validity of its estimates of populations of the elderly.

---

[7]The Steering Committee, which consisted of 12 senior career bureau officials, was responsible for reviewing census and A.C.E. data and preparing a report in February, 2001, for the bureau director recommending whether the adjusted or unadjusted census data should be used for redistricting.

We discussed conducting additional evaluation work with the bureau. However, bureau staff do not believe that more evaluation work is necessary for several reasons. First, more evaluation work would be difficult because the A.C.E. data is not characterized in a manner that would allow them to identify retirement homes. Second, they do not believe they missed all of the population living in retirement homes because various A.C.E. listers may have interpreted the definition differently. Finally, the bureau contends that the post-strata for the population that would be living in retirement homes was "50 years and older" and that missing some of the population in retirement homes would not be sufficient to have an impact on the estimation of this post-strata. However, we believe that the bureau should review census person records coded as correct enumerations (i.e., correctly enumerated in the census but not in the P-sample); assess whether these correct enumerations included high concentrations of elderly in certain clusters, thus possibly indicating missed retirement homes; and explain what impact this has on the estimation. More importantly, the lessons learned here will be useful as the bureau plans for 2010.

II. A.C.E. Requirements Process Should Be Improved

Many computer systems supporting various headquarters and field operations were defined and developed for A.C.E. We found some areas where improvements could be made to ensure that all requirements were appropriate, sufficiently defined, and clearly communicated. The bureau has standards to aid in the development of software requirements; however, we found that they were not always used. Instead, bureau staff responsible for generating requirements specifications sometimes relied on dress rehearsal specifications that did not completely define all requirements. For the PFU laptop system, the bureau relied on procedures that had worked for other surveys and censuses. We believe that requirements definition suffered because not enough time was spent on planning activities. The problems cited in this report should not detract from the successes that the bureau had in managing and implementing many system developments and enhancements, sometimes in very short time frames. However, as discussed in other OIG reports issued for this decennial, in the future the bureau would benefit from using a well-managed software development process that allows sufficient time for planning, developing, testing, and implementing all of its systems. Our recommendations in those reports hold true for A.C.E.[8]

---

[8] *A Better Strategy Is Needed for Managing the Nation's Master Address File,* OSE-12065, September 2000; *PAMS/ADAMS Should Provide Adequate Support for the Decennial Census, but Software Practices Need Improvement*, OSE-11684, March 2000; *Improvements Needed in Multiple Response Resolution to Ensure Accurate, Timely Processing for the 2000 Decennial Census,* OSE-10711, September 1999; and *Headquarters Information Processing Systems for the 2000 Decennial Census Require Technical and Management Plans and Procedures*, OSE-10034, November 1997.

A. A.C.E. and Census Joint Requirements Need to Be Better Planned

Over a year before the PFU operation was to start, the Field Division requested that Decennial Statistical Studies Division (DSSD) prioritize the receipt of each regional office's PFU workload prior to the respective matching operations. This prioritization would allow the regions to influence the order in which the work would arrive, so they could plan staffing and training, and complete A.C.E on schedule.[9] Bureau officials told us that they ranked each block cluster based on weather and other operational concerns. Higher-ranked clusters' PFU questionnaires were to be delivered to the regional office first. DSSD agreed and used the ranking to create a national list of 11,802 prioritized block clusters. DSSD planned to follow the national priority list based on the availability of the census data. While the bureau was able to deliver all PFU questionnaires on schedule, the order in which they were delivered had significant deviations from the priority ranking. For example, California had 3 block clusters ranked within the top 50 of the 11,802 block clusters but was one of the last states to receive its PFU questionnaires.

DSSD staff responsible for implementing the requirement explained that they did not have a complete specification from DSCMO which should have defined and documented how the census data would be created. Because census data availability drove when person matching would occur, the extent to which the priority requirement could be met relied on how census data would be delivered. Absent the complete specification, DSSD assumed that census block cluster data would be available as soon as it was aggregated by local census office, as it was during the 1998 dress rehearsal.[10] DSSD officials furthered explained that, although they completed the specification for prioritizing field PFU work on July 14, 1999, they did not learn until about a year later that the headquarters data processing operation would release the data when each state was ready, rather than by local census office.

Headquarters field staff stated that, despite the confusion, the PFU workload was completed on schedule for several reasons: the weather was mild, the workload was almost a third less than planned, and extra staff were used to cope with the change in plans. Once all involved offices were aware of the requirement to obtain the census data in a priority order, DSCMO, DSSD, and the Field Division worked to correct the problem. The Field Division created a new priority list by state, and DSSD and DSCMO worked to meet this requirement by accelerating high priority states when possible. Although the work was completed on schedule, some field staff stated their hope that block cluster prioritization would be workable in the future. We believe that meeting

---

[9]A.C.E. Regional Office Memorandum No. 99-05 and DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-FT-02, *Accuracy and Coverage Evaluation Housing Unit Follow-up and Person Follow-up Prioritization Specifications,* July 14, 1999.

[10]In the dress rehearsal, DSCMO delivered enumeration data for the DSSD coverage measurement processing operation by local census office because only three local census offices were involved.

such a requirement would improve efficiency and reduce the risk of schedule and cost overruns to the PFU operation.

When planning systems that have to meet the requirements of both the census and the A.C.E., completed specifications would ensure that all requirements are documented and better communicated. In this case, having documented requirements well before the start of both operations would have facilitated better planning to meet the Field Division's needs and resulted in a more efficient PFU operation.

B. Computer Matching Requirements Should Be More Completely Specified

By most measures, computer matching went well. It was completed very close to schedule, the match rate of 69.6 percent was higher than expected, and clerical matchers reviewing the results stated that they were highly accurate. Nevertheless, we found that the requirement for determining the lower cutoff score for possible matches could be better defined, and the requirements specification should have been more completely documented. In the future, the bureau may not have key staff members who supported Census 2000 and would therefore benefit from clear and complete requirements specifications.

The person computer matching phase involved linking person records from the E-sample to person records from the P-sample, assigning match codes to each linked pair, and submitting these results to clerical review. The goal of computer matching was to label linked records that matched with a high degree of certainty as matches and those that matched with a lower degree of certainty as possible matches. To do this, the SRD Record Linkage System assigned a match score ranging from 13.0566 (indicating a definite match) to -3.8929 (indicating a definite non-match) to each linked pair.

In our observation of the expert matcher review of one match group, we found that the way that the possible match cutoff score was chosen was not clear. Also, our assessment of the clerical match process led us to believe that it would be easier to complete the matching process if possible matches remained linked, at least until clerical matchers had a chance to review them. However, DSSD staff believed that NPC managers indicated that linking too many records that were obviously not matches could cause problems for clerical staff, who would have to spend time unlinking these records and thus might lose confidence in the computer matcher results. One possible approach may be to provide the clerical matchers with the linked records, noting those that fall below the lower cutoff score, and not requiring them to unlink them. While the expert matchers were clear about where to mark matches with a high degree of certainty, they were less definitive about where to delineate the possible matches. For example, in the match group we observed, when we requested that the possible match cutoff be lowered, the expert matcher complied and stated that this cutoff point was flexible.

We reviewed the requirements specification for person computer matching to assess its completeness and to ascertain what the requirements were for designating possible matches. We found that the requirements specification defined such items as address and person standardization, insufficient or incomplete information for matching, match codes to be assigned to each record, and what should be included in the P-sample and E-sample match universe. However, the specification did not designate which fields were to be compared, how they should be compared, and what scores should be assigned based on the level of agreement of the data such as cases with conflicting data. Nor did the specification define cutoff scores for delineating matches or possible matches or records that were to be unlinked prior to clerical matching. In addition, the specification did not indicate a target level of accuracy for matches or possible matches, even though the expert matcher indicated that the goal was to achieve a 99.5-percent level of accuracy for matches.

DSSD managers decided that the computer specification was to include only the information on how the input to be used in the computer matcher would be handled and standardized, and not provide a description of the computer matcher software. Bureau staff responsible for writing the specification further explained that it was based on the one for the dress rehearsal, that they relied on the judgment of the expert matchers, and that the specification would be updated to reflect the experience gained from the matching operation.

We believe that the specification should be complete, unambiguous, and present requirements in a fashion that can be tested. In addition, while bureau officials stated that they planned to refine requirements after the operation was completed, it remains unclear whether the data will be available and whether the people with the appropriate expertise will be involved in the process. For example, when we went back to obtain data about the consistency of matches by match group, we were told that new software would be required to process our request. In addition, cutoff scores are contained in an undocumented log file, and examples of linked person records above and below these cutoffs are not readily available, making it difficult to determine the basis for the cutoff scores.

The expert matchers were familiar with using the SRD Record Linkage System and could conduct the computer matching with an incomplete specification by using prior experience and setting cutoffs by viewing the data as it was being matched. However, relying on matcher expertise in conducting key operations such as A.C.E. computer matching created unnecessary risk to the successful completion of the operation and may have left insufficient information to be used in developing procedures for the next decennial. In the future, the bureau needs to ensure that user requirements, such as what constitutes a possible match, are clearly understood and that requirements specifications are complete, unambiguous, clearly communicated, documented, and testable.

C. Laptop System Provided Minimal Benefit for PFU Users

Laptops previously used to collect A.C.E. person interview data were provided to field supervisors to transmit management reports and e-mail messages during PFU. According to the requirements specification, the laptop was designed as a tool for managing computer-assisted person interviews in the field. The system included a case management component, a telecommunication system, a training function, and a field supervisor function, which included management reporting. During the 1998 dress rehearsal, we reported that the management reports were not always available on the laptop because of software delays and technical problems and that management reports had to be faxed or mailed to field supervisors daily.[11] As a result, during the dress rehearsal, the data was outdated and could not be relied on to manage the operation. We recommended that for future laptop use, management reporting software be corrected and tested to ensure that accurate and timely information would be available to field managers.

During A.C.E. PFU operations, we asked field supervisors whether the laptop management reporting problems identified during dress rehearsal had been resolved. Supervisors again stated that management reports were not useful because they were not timely. Similar to dress rehearsal, the reports lagged a day or two behind the operation. Given the fast pace of the PFU operation, a day or two lag in data rendered the reports useless to the supervisors. As a result, many supervisors kept their own records to track the number of questionnaires completed by interviewers. In hindsight, headquarters field staff attributed the reporting delay to the time it took to get the questionnaires shipped from the field to the regional office and then scanned into the computer for check-in purposes.

Field supervisors were also dissatisfied with the e-mail system. For example, supervisors could send e-mails only to the regional office and not to one another. In addition, supervisors did not know how to save messages, and since the laptops did not have printing capability, supervisors did not have a record that a message had been sent. Headquarters staff explained that the laptops did not have a traditional e-mail system. Instead, the mail system developed for the laptop was designed only to allow users to send messages to their regional office.

In the exit conference, senior managers stated their belief that not all user requirements, such as e-mail, need to be met and noted that the extent to which requirements are fulfilled is a management decision. While those statements may be accurate, we found that, in providing the laptops to PFU field staff, the bureau deployed hundreds of laptops that were of minimal use to field staff managing the PFU operation. Originally, the system was designed as an automated tool to collect and transmit person interview data, functionality not used for PFU where interviews

---

[11] *Dress Rehearsal Quality Check Survey Experience Indicates Improvements Needed for 2000 Decennial,* Office of Inspector General, ESD-11449-9-0001, September 1999.

were conducted with pen and paper. Functionality such as traditional e-mail, that would have made management more efficient, was not on the laptops. Therefore, the laptops did not meet user needs which implies a questionable use of funds in providing these tools. To improve management of its operations in the future, the bureau needs to ensure that systems meant to facilitate operations better satisfy user requirements.

III. Use of Automated Tools Facilitated Person Matching, but Better Testing, Documentation, and Quality Assurance Needed

The bureau's use of the computer matcher and a Docuprint system enhanced the efficiency of bureau matching operations and the printing of PFU questionnaires. Although the computer matcher was not newly developed software, it was being applied differently and the test plan and results should have been documented. With respect to the Docuprint system, it appears that errors occurred in production because of a failure to sufficiently review the accuracy of the printed questionnaire data after last minute changes to the system.

A. Computer Matcher Software Tested but Results Not Documented

Before the 2000 A.C.E., the SRD Record Linkage System had been used in numerous operations, including quality checks of the 1990 census, and the 1998 dress rehearsal, as well as the identification of multiple responses submitted by households in the 1998 dress rehearsal and the 2000 decennial. Its accuracy and performance have been documented, and it is viewed as a reliable software component.[12] However, when the SRD Record Linkage System is applied to specific data, choices have to be made as to how the data is to be grouped for matching, which fields to match on, and how those fields are to be weighted to determine the match score.[13] For example, the A.C.E. person matching operation required that matching be done in two passes, with different fields used in each pass, and that special weighting be given to certain types of matches. Changes such as these made to tailor the matcher to a specific operation should be tested to ensure that the results meet accuracy and consistency requirements.

DSSD officials told us that they conducted independent testing of the SRD Record Linkage System by using the new computer match process to match person records from dress rehearsal and then comparing the results to the final matched results from dress rehearsal. The bureau had not documented this testing, but clerical matchers responsible for reviewing the A.C.E. match data

---

[12] Matthew A. Jaro, *Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida*, Journal of American Statistical Association, June 1989.

[13] *User Documentation for the Census SRD Record Linkage System*, Bureau of the Census (undated).

during the census clerical matching operation did not report any problems with matcher accuracy. Nevertheless, we are concerned that the bureau did not have a documented test plan demonstrating that the test cases used were comprehensive and its results were in agreement with the requirements. We believe that to diminish risk and ensure optimal accuracy, the bureau should adopt a formal testing process that includes test plans, a demonstrably comprehensive set of test cases, and documented results.

B. Output Not Adequately Reviewed After Changes to Forms Printing Software

For the PFU field operation, a questionnaire was printed at NPC for each household requiring follow-up. The questionnaire contained pre-filled items that provided household and address information and was designed to collect additional information about persons for whom the bureau did not have enough information to determine their Census Day address. The questionnaire could have up to 12 sections but typically had 5 or 6.

We found that a programming change had been made to the software during production that resulted in the printing of incorrect data on thousands of questionnaires by the NPC's Docuprint system. The programming change was made to the software during the first several days of PFU questionnaire printing. An unintended side effect of this change caused the software to select and print the same last name for every person listed on the questionnaire even if they had a different last name. This error went undetected for four days and affected thousands of questionnaires that contained multiple people with different last names. The bureau was not able to determine at what point during the printing process the error occurred or how many forms were affected. The greatest impact was at the regional level, where the forms were immediately shipped after printing.

Regional staff in Seattle, Denver, Philadelphia, and Charlotte told us that it took a substantial amount of time to review every form to ensure that the names were correct before they were sent out into the field. For example, Charlotte staff lost one day of production while they reviewed every form. Fortunately, the error only affected the inside sections of the questionnaire and not the names printed on the rosters on the front cover. Therefore, regional staff were able to use the last name on the cover to correct the last name inside the questionnaire.

Although this software error caused additional work for the regions and delayed field production, PFU was completed on schedule. The regional offices stated that, because this error occurred at the start of PFU, they had time to recover and the NPC clerical matchers who reviewed the data did not believe that the quality was impacted.

Quality assurance procedures are designed to ensure that output from processes such as the printing of PFU forms are accurate and complete. Although the bureau had quality assurance

procedures for reviewing questionnaire output, they needed to be more specific on how to verify that the content was accurate. We believe that the quality assurance procedures should be revised to include steps that ensure data output is accurate and complete.

IV. PFU Quality Assurance Program Needs to Be Automated

During our review of the PFU operation in the 1998 dress rehearsal, we found that there was no quality assurance program in place and recommended that such a program be added for the decennial.[14] We commend the bureau for adding a quality assurance process in response to our recommendation made during the decennial dress rehearsal. During the decennial, we were informed by bureau personnel that the quality assurance program worked and had caught interviewers falsifying data on PFU forms. While this program was effective, personnel at the regional office explained that tracking the quality assurance workload was very paper- and labor-intensive.

During several phases in the quality assurance program, regional personnel had to manually track the progress of each questionnaire and hand-carry it from phase to phase. The phases included selecting questionnaires; determining eligibility; performing telephone quality assurance and, if necessary, assigning the form to a field quality assurance checker; and when the interview was completed, sending selected PFU questionnaires to NPC for processing and interview worksheets to headquarters.

The process was tedious because the outcome for each phase had to be recorded, the questionnaire and quality assurance interview worksheet had to be placed back into an envelope, and then the envelope had to be put into the appropriate bin for pickup by the next phase. Regional personnel believed that the overall flow was too segmented and that some phases could have been combined.

Because the PFU quality assurance operation was added after the 1998 dress rehearsal, there was no time to operationally test it before the decennial. In addition, bureau officials told us that the PFU operation was intended to be conducted by a computer-assisted person interview system similar to the person interview operation, but the bureau decided during dress rehearsal that not enough resources were available to develop and appropriately test this system for PFU. As a result, the bureau then made the decision to use the Docuprint system for PFU. For 2010, we believe that quality assurance should be integrated as part of an automated PFU operation.

---

[14] *Dress Rehearsal Quality Check Survey Experience Indicates Improvements Needed for 2000 Decennial,* Office of Inspector General, ESD-11449-9-0001, September 1999.

RECOMMENDATIONS

We recommend that the Acting Director of the Census Bureau direct senior management responsible for evaluating the current A.C.E. and planning the quality check of the 2010 census to take the necessary actions to:

1. **Evaluate whether residents of retirement homes were systemically omitted from the A.C.E. independent sample and if so, what impact this had on estimates of the elderly population.**

   In its response, the bureau stated that all A.C.E. evaluations requiring field work had been completed and that additional evaluation work to satisfy this recommendation would be difficult because such housing units were not characterized as being retirement homes in the census data. Furthermore, two years after Census Day, residents may no longer live in the housing units in question. The bureau also stated that it is considering including group quarters in the coverage measurement survey for Census 2010, which should reduce the potential for error caused by confusing housing units with group quarters.

   We maintain that not knowing how much missed coverage of retirement homes impacted this population's net undercount calculation diminishes confidence in the A.C.E. results for this population. Even without additional field work, analyzing the results of matching the P-sample and E-sample to identify high concentrations of elderly within A.C.E. clusters that were not matched in the P-sample would indicate whether this omission was systemic. Without such a determination, the bureau cannot determine the validity of its estimates of populations of the elderly.

2. **Ensure that definitions for housing unit, special place, and group quarters, as well as any other definitions, as appropriate, are consistent for both the census and the A.C.E.**

   The bureau agreed with this recommendation.

3. **Document and clearly communicate requirements when planning systems that have to meet requirements of both the census and the A.C.E.**

   In its response, the bureau stated that a very tight schedule after the Supreme Court's decision on the use of sampling for producing the apportionment counts made it difficult to document and communicate requirements for systems supporting both the census and the A.C.E.

We believe that documenting and communicating these requirements would have largely been the same task with or without sampling.

4. Ensure that computer matching requirements are fully developed and documented.

In its response, the bureau claims that the automated matching requirements were fully developed and understood by those who conducted this operation.

We believe that the criteria for including records as possible matches were ambiguous because setting the lower cutoff was not completely specified or documented.

5. Ensure that computer systems are developed and modified in accordance with rigorous, documented system and software engineering standards that, at a minimum, address requirements specification, design and development, testing, documentation, and quality assurance.

The bureau cited its response for the third and the fourth recommendations.

We are asking the bureau to do more than improve requirements definition— the bureau needs to improve all phases of its system and software engineering processes.

6. Integrate the PFU quality assurance process as part of an automated PFU operation to ensure efficiency.

The bureau agreed with this recommendation for the 2010 census.

APPENDIX A
Terms and Acronyms

A.C.E.	Accuracy and Coverage Evaluation
DSCMO	Decennial Systems and Contracts Management Office
DSSD	Decennial Statistical Studies Division
Dual system estimation	The estimation methodology used for the A.C.E. that uses a geographic sample of block clusters to find persons missed by the census and other errors in the census
E-sample	Includes persons enumerated by the census in housing units from the A.C.E. block clusters
Group quarters	Special places where occupants share certain residential facilities, such as a cafeteria, and include dormitories, nursing homes, orphanages, prisons, and halfway houses
HCUF	Hundred-percent Census Unedited File, a computer file containing the results of the decennial census
Housing unit	A house, an apartment, a group of rooms, or a single room whose occupants live separately from other persons and have direct access from the outside of the building or through a common or public hall
Local census office	520 temporary Census Bureau offices established for Census 2000 data collection purposes
NPC	National Processing Center, Jeffersonville, Indiana
OIG	Office of Inspector General
P-sample	Person records collected during A.C.E. independently of the census
PFU	Person follow-up, an A.C.E. operation resolving unmatched E-sample persons
Special places	A place containing one or more group quarters where people live or stay, such as, a college or university, nursing home, hospital, prison, hotel, migrant and seasonal farm worker camp, or military installation or ship
SRD	Statistical Research Division

UNITED STATES DEPARTMENT OF COMMERC
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001
OFFICE OF THE DIRECTOR

APPENDIX B. Bureau Response

December 11, 2001

MEMORANDUM FOR Edward L. Blansitt
Deputy Inspector General

Through: Kathleen B. Cooper KBC
Under Secretary for Economic Affairs

From: William G. Barron, Jr.
Acting Director

Subject: *Selected Operations for Census 2000 Accuracy and Coverage Evaluation Need Improvements Before 2010*
Draft Inspection Report No. IG-14226

This is in response to your memorandum transmitting the above-referenced draft inspection report, which made the following recommendations:

Recommendation #1:

Evaluate whether residents of retirement homes were systematically omitted from the Accuracy and Coverage Evaluation (A.C.E.) independent sample and, if so, what impact this had on estimates of the elderly population.

Action Taken or Planned:

The U.S. Census Bureau has completed all A.C.E. evaluations that required field work. Additional evaluation work to satisfy the recommendation would be difficult, because the data are not characterized in a manner that would allow us to identify retirement homes. In addition, there would be approximately two years time between the census interview and the evaluation interview. Factors such as a change in residence from a retirement home to a "regular" home during this two-year period would raise questions about the validity of such an evaluation. However, the Census Bureau is considering including group quarters in the coverage measurement survey for Census 2010, which may help reduce any bias as a result of the retirement home situations identified in the report.

Recommendation #2:

Ensure the definitions for housing unit, special place, and group quarters, as well as any other definitions as appropriate, are consistent for both the census and the A.C.E.

Action Taken or Planned:

We agree with the recommendation.

Recommendation #3:

Document and clearly communicate requirements when planning systems that have to meet requirements of both the census and the A.C.E.

Action Taken or Planned:

We realize this was an issue. However, after the Supreme Court's decision on the use of sampling for producing the apportionment counts, the Census Bureau made every effort to document and clearly communicate requirements within a very tight schedule. We plan to make improvements in this area for the 2010 census.

Recommendation #4:

Ensure that computer matching requirements are fully developed and documented.

Action Taken or Planned:

The automated matching requirements were fully developed and understood by those who conducted this operation. We agree that, in general, we need to do a better job of documenting requirements.

Recommendation #5:

Ensure that computer systems are developed and modified in accordance with rigorous, documented system and software engineering standards that, at a minimum, address requirements specification, design and development, testing, documentation, and quality assurance.

Action Taken or Planned:

Our response to this recommendation is covered in our responses to Recommendations #3 and #4.

Recommendation #6:

Integrate the PFU quality assurance process as part of an automated PFU operation to ensure efficiency.

Action Taken or Planned:

We agree that a Person Follow-up (PFU) quality assurance process as a part of the automated PFU operation is important and should be carefully studied for the 2010 census. In order to minimize risk, we did not implement this integration for Census 2000, because we did not conduct PFU quality assurance on automated PFU in the Census 2000 Dress Rehearsal.

cc: US/EA