

ENGLISH ONLY

**STATISTICAL COMMISSION and
UN ECONOMIC COMMISSION FOR
EUROPE (UNECE)**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**CONFERENCE OF EUROPEAN
STATISTICIANS**

Joint UNECE-Eurostat Work Session on Population Censuses
Organized in cooperation with UNFPA
(Geneva, 23-25 November 2004)

Session 2– Invited paper

EVALUATION OF CENSUS QUALITY AND COVERAGE

Submitted by Bureau of the Census, United States of America *

1. THE U.S. DECENNIAL CENSUS EVALUATION PROGRAM

1.1 Summary of the Research, Evaluation, and Experimentation (REX) Program: 1940-1990

1. The United States Census Bureau has a long history of evaluating our own programs. Research and evaluation of the U.S. decennial census officially began with the 1940 Census. There was research and testing both prior to and during the census. After the 1940 Census, there was a post-censal program to evaluate some of the questions and there were attempts to estimate underenumeration based on preliminary sample data, however, they were considered too approximate to report.

2. For the 1950 Census, the Census Bureau initiated a Research, Evaluation, and Experimentation (REX) program that looked back at the preceding census and forward to the next one. One purpose was to look at the accuracy of the census. We identified sources and measures

* Paper prepared by Preston J. Waite.

of error in order to make data users aware of any limitations. Another major purpose was to evaluate coverage, both over- and under- enumeration.

3. The 1960 Census included content and coverage improvement evaluations and a post-enumeration survey to measure coverage. The 1960 REX program results led to the expansion of self-enumeration and improved processing procedures for the next census.

4. There were three major studies of content error in the 1970 Census as well as procedural evaluations and tests and an experiment to determine whether it was feasible to further expand the mail census.

5. The 1980 Census REX program consisted of over 40 separate projects including coverage improvement evaluations, content evaluations, processing evaluations, a number of experiments and other studies, and a post-enumeration survey.

6. The 1990 Census REX program was similar to the 1980 program in both content and scope, focusing on content, coverage, and procedures and processing.

1.2 Overview of the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program

7. The Census 2000 Testing, Experimentation, and Evaluation (TXE) Program was the most ambitious formal test and evaluation program ever undertaken by the Census Bureau. The success of the Census 2000 program can be measured from the findings of 87 evaluations (organized into 18 categories), 20 reports for 6 experiments and studies, 15 topic reports, and 4 synthesis reports. The research findings provide measures of effectiveness of the Census 2000 design, operations, systems, and processes, as well as measures of coverage and data quality. The experiments, especially, provide information on the value of new or different methodologies. Together, the evaluations and experiments include recommendations that are currently being incorporated into the testing program for the 2010 Census and are informing the planning and development of the American Community Survey and the Master Address File Updating System.

8. Evaluation reports assessed the effectiveness of Census 2000 operations, systems, and processes. Evaluations of some of the newer programs or systems included the national paid advertising campaign, the automated data capture system which included Optical Mark and Optical Character Recognition, address list development operations, field followup operations to improve coverage, and a new multiple category option for race reporting.

9. Experiments were conducted during Census 2000 because the decennial census environment provided the best conditions to learn about new methodologies. The effects of the following were measured by experiments during Census 2000:

- alternative questionnaire designs, including tests of coverage questions, presentation of the race question, and navigation through the form;
- administrative records as a data collection tool and as a means to assess imputation methods;
- the use of administrative records on respondents' behavior and attitudes, including requesting a respondent's Social Security Number; and

- alternative electronic response mode options (with and without incentives).

10. Topic Reports summarized and integrated relevant findings from experiments, evaluations, operational assessments, and other research on related census subjects, and made recommendations for the 2010 Census. Fifteen topic reports were prepared on the following subjects: Address List Development, Automation of Census Processes, Content and Data Quality, Coverage Improvement, Coverage Measurement, Data Capture, Data Collection, Data Processing, Ethnographic Studies, Partnership and Marketing, Privacy, Puerto Rico, Race and Ethnicity, Response Rates and Behavior Analysis, and Special Places and Group Quarters.

1.3 Methods Used

11. Some of what was included in the Census 2000 TXE program were more like operational profiles than evaluations. These profiles documented number of cases (such as operational workloads), rates (such as mail response rates), data from production files (such as number of cases failing coverage edits), quality assurance information (such as number or percent of cases that failed quality assurance checks), and information collected from enumerator or other staff debriefings.

12. Other evaluations analyzed, interpreted, and synthesized the effectiveness of census components and their impact on data quality or census coverage. Examples of this type of evaluation include analyses of imputation rates and item nonresponse rates, the impact of the data capture system on data quality, coverage gains from coverage questions on the enumerator-completed questionnaires, the impact of address list development operations on the Master Address File, and potential housing unit and person coverage gain from the coverage improvement followup operation.

13. To evaluate components of the Partnership and Marketing Program, we conducted sample surveys. We conducted targeted reinterview surveys to evaluate responses to the race question. We also conducted surveys to determine customer satisfaction with telephone questionnaire assistance, web-based questionnaire assistance, and Internet data collection.

14. Some evaluations included focus group interviews to learn how or why respondents behave in a certain way. For example, we conducted a series of focus groups across Puerto Rico to learn more about how people in Puerto Rico viewed the questions on race and ethnicity.

15. We used ethnographic and social network methods to study the effects of mobility on census coverage.

16. As noted earlier, experiments were conducted to learn about new methodologies that might be included in future censuses. The census environment is the best way to determine how new methodologies might work in a future census. One of the experiments conducted in Census 2000 explored two methods for conducting an administrative records census.

1.4 Highlights of some of the evaluation results include:

17. Census 2000 enumerated over 281 million people in close to 116 million housing units and almost 8 million people in about 192 thousand group quarters. Group quarters include places such as colleges, correctional facilities, and nursing homes.

18. About 80 percent of addresses were in mailout/mailback census areas. Almost 19 percent of addresses were in areas where census enumerators distributed questionnaires for respondents to fill out and mail back. The other 1 percent of addresses were in areas using specially targeted procedures.
19. The mail response rate was approximately the same as in 1990, halting the decline in response rates from recent censuses.
20. It appears that the paid advertising contract, along with our partnership program with local governments, community-based organizations, religious groups, and businesses, was effective in stimulating awareness of the census for all targeted populations.
21. Setting competitive enumerator pay rates and overhiring early in the process (a system we called frontloading) led to dramatic improvement in timely completion of nonresponse followup.
22. The Housing Unit Unduplication Operation identified and removed more than 3.6 million persons and 1.4 million housing units. While this was an ad hoc operation for Census 2000, we will likely need a comprehensive unduplication program for the 2010 Census.
23. The automated data capture system using Optical Mark Recognition (OMR) and Optical Character Recognition (OCR) was very successful in capturing respondent data. Typical error rates for OMR were less than 1.5 percent and for OCR were about 1.0 percent. The evaluations also pointed out areas for improving the data capture system for 2010.
24. Item nonresponse rates were very low for the hundred percent items that are asked of the entire population. The rates ranged from about 1 percent for the gender item to about 4 percent for the housing tenure item.
25. Quality assurance of Census 2000 field operations was an important element in preventing significant errors and in preventing the clustering of significant errors.

2. COVERAGE MEASUREMENT

2.1 Coverage Measurement in the 1950-1990 Censuses

26. The Census Bureau conducted its first coverage measurement survey, a post-enumeration survey, after the 1950 Census. An evaluation of the 1940 Census demonstrated the need for an assessment of census coverage. A match of draft registration records to the 1940 Census evaluated the coverage of adult males of an age eligible for the draft. Surprisingly, there were more males registered for the draft than enumerated in the census. The 1950 post-enumeration survey reinterviewed 22,000 households and estimated a net undercount of 2.1 million people. This differed greatly from the demographic analysis estimate of between 5 and 5.5 million people.
27. The Census Bureau also conducted a post-enumeration survey after the 1960, 1980, 1990, and 2000 Censuses and in test censuses in the 1970s, 1980s, and 1990s in preparation for censuses. The design of the post-enumeration survey has evolved over the years. The implementation and estimation methodology in the 1950 and 1960 post-enumeration surveys had some basic differences with the designs used in 1980, 1990, and 2000.

28. The 1950 and 1960 post-enumeration surveys did not use dual system estimation to estimate census coverage error. Rather the idea was to form a much higher quality census interview with highly trained interviewers. The results of the post-enumeration survey interview were assumed to be the truth. However, the results were disappointing. Neither the 1950 nor 1960 post-enumeration surveys measured as large an undercount as found by demographic analysis.

29. The Census Bureau did not evaluate the 1970 Census using a post-enumeration survey. Analysis of the methods used in the 1950 and 1960 Censuses revealed that the technique of demographic analysis provided more accurate estimates of census population coverage than estimates obtained from re-enumeration or record check studies. Therefore, a decision was reached to use demographic analysis as the method for coverage evaluation in the 1970 Census. Other coverage checks were used, but they were essentially to support the demographic analysis or to provide additional information on certain components of undercoverage or operational problems, rather than to estimate total coverage.

30. Improvements in the methodology came from development in other countries where dual system estimation was applied. Dual system estimation required assuming the post-enumeration survey was only a second enumeration, not that it was a perfect enumeration. The 1980, 1990, and 2000 implementations used dual system estimation.

31. In the 1980 Census, population coverage was estimated through the post-enumeration survey, administrative record matching, and demographic analysis. The post-enumeration survey used three separate surveys as its principle sources: the April 1980 Current Population Survey sample of households, the August 1980 Current Population Survey sample of households, and a sample of households selected from the census itself (the E-sample). The two Current Population Survey samples, supplemented by a sample of people in group quarters, formed the P-sample.

32. For more than a decade prior to 1990, there had been a political and legal controversy about whether the post-enumeration survey and demographic analysis methods should be used to adjust census results to correct for the undercount. In 1989, litigation led to an agreement that the Census Bureau would evaluate the 1990 Census to decide whether to adjust the census results. The 1990 post-enumeration survey consisted of a sample of nearly 172,000 housing units in 5,290 block clusters. There was oversampling in areas with American Indian reservations and tribal trust lands and in areas with significant Black, Hispanic, or Asian populations. High-level department and bureau officials were split in advising the Secretary of Commerce whether to adjust the census numbers. In July 1991, the Secretary announced that the 1990 Census would not be statistically adjusted.

33. Litigation and controversy continued throughout the 1990s. In January 1999, the U.S. Supreme Court ruled that the law covering the census, Title 13, did not permit sampling for apportionment of the House of Representatives.

2.2 The Census 2000 Accuracy and Coverage Evaluation (A.C.E.) Survey

34. The Census 2000 A.C.E. followed the basic methodology of post-enumeration surveys using dual system estimation. The A.C.E. evaluated the coverage of Census 2000 only for individuals living in housing units. The A.C.E. did not assess the coverage of individuals living in group quarters.

35. The E- and P-samples for the A.C.E. used the same sample of block clusters. All the census enumerations geocoded to the sample block clusters were in the E-sample. An independent listing of housing units that did not rely on any of the Census 2000 addresses was constructed in the sample block clusters. All people living in the housing units in the sample block clusters were in the P-sample.

36. The A.C.E. design had several basic steps. First, interviewers created a list of all the housing units in the sample block clusters. Next, P-sample interviews were conducted at all the housing units. With the data collected, cases were matched to the census on a case-by-case basis. When there was uncertainty about whether a P-sample person was enumerated or an E-sample enumeration should have been in the sample blocks, a followup interview was conducted. With the additional information, a final matching operation was conducted.

37. The Census Bureau conducted the A.C.E. expecting it could be used to adjust the Census 2000 results for all non-apportionment purposes if it improved the census data. The original March 2001 A.C.E. estimates became available in time to correct the Census 2000 redistricting files. However, on March 1, 2001, the Census Bureau released the "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy" which reported that we were unable to conclude, based on the information available at the time, that the adjusted Census 2000 data were more accurate for redistricting. Accordingly, we recommended that the unadjusted census data be released as the Census Bureau's official redistricting data.

38. We conducted further evaluations over the next six months to examine the difference and determine if Census 2000 data, other than redistricting data, should be corrected. The demographic analysis estimates were also investigated further, resulting in revised demographic analysis estimates.

39. In October 2001, the Census Bureau released the "Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses." After assessing considerable new evidence, we recommended that unadjusted Census 2000 data also be used for non-redistricting purposes. The effect of this new evidence is that the A.C.E. overstated the net undercount by at least 3 million persons. The cause of this error was that the A.C.E. failed to measure a significant number of census erroneous enumerations, many of which were duplicates. This level of error in the A.C.E. measurement of net coverage is such that the A.C.E. results could not be used in their current form. This finding of substantial error, in conjunction with remaining uncertainties, necessitated that revisions, based on additional review and analysis, be made to the A.C.E. estimates before any potential uses of these data could be considered. The Census Bureau released the remaining Census 2000 data products, post-censal estimates, and survey controls using unadjusted data. It was, however, reasonable to expect that further research and analysis might lead to revised A.C.E. estimates that could be used to improve future post-censal estimates.

40. Coupled with the revisions to the demographic analysis estimates, the inconsistency with demographic analysis was explained by the failure of the A.C.E. to measure a large number of census erroneous enumerations. The earlier concerns in A.C.E. with balancing, contamination, and missing data were also resolved. The level of other errors was believed to be small by comparison and therefore was not a major factor in the second decision.

3. PLANS FOR THE 2010 CENSUS RESEARCH PROGRAM

3.1 Basic Differences from the Census 2000 TXE Program

41. The proposed plan for the 2010 Census program embodies several basic differences from 2000:

- The proposed 2010 Census program will make a distinction between true evaluations and “accounting type” assessments. In Census 2000, several formal evaluations were actually operational assessments that provided metrics with no analysis.
- The Census Bureau executive staff plans to decide the high-level scope and focus of research programs *before* participating divisions develop research proposals. This will ensure that the selected evaluations and experiments address the most relevant topics and program funding is optimally expended.
- Planning and design milestones for 2010 Census research components are slated to occur earlier than in the Census 2000 cycle.

3.2 Specific Areas of Interest for the 2010 Census Research Program

42. As part of the Census 2000 TXE Program, we constructed the Master Trace Sample database to help answer 2010 Census research questions that go beyond those addressed by the Census 2000 evaluations and experiments. Several recommendations for future databases came out of the process of creating the Census 2000 Master Trace Sample database:

- The database should be expanded to include data on Group Quarters.
- The database should be expanded to include coverage measurement data associated with persons. The A.C.E. final Census Day housing unit status is the only Census 2000 A.C.E. data included in the prototype.
- We might also consider building a ‘trace’ database specific to each major demographic or economic survey, as well as for the Economic Census.

43. We want to measure:

- the effectiveness of procedures for unduplication,
- geocoding accuracy,
- the effectiveness of race reporting,
- the effect of residence rule strategies, and
- the impact of expanded use of automated tools by our field enumerators.

44. We intend to continue research on the use of administrative records to supplement development of the address list and to obtain data for nonresponse items.

45. And, of course, we expect to continue improving upon our efforts to measure census coverage, including net error as well as components of under- and over-enumeration.

REFERENCES

Abramson, Florence H. (2004), *Census 2000 Testing, Experimentation, and Evaluation Program, Census 2000 Testing, Experimentation, and Evaluation Summary Report*, Washington, D.C., DRAFT.

Childers, Danny R. and Rita Petroni (2004), *Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No.4, TR-4, Coverage Measurement From the Perspective of March 2001 Accuracy and Coverage Evaluation*, Washington, D.C., 2004.

Clark, Jon R. and Darlene Moul (2004), *Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No.10, TR-10, Coverage Improvement in Census 2000 Enumeration*, Washington, D.C., 2004.

Hill, Joan M. and Jason D. Machowski (2003), *Census 2000 Evaluation B.6: Master Trace Sample*, September 2003.

Mulry, Mary H. (2004), *Summary of Accuracy and Coverage Evaluation for Census 2000*, Washington, D.C., DRAFT.

U.S. Census Bureau (2002), *Census 2000 Testing, Experimentation, and Evaluation Program*, May 2002.

U.S. Census Bureau (2004), *Proposal for Planning and Designing the 2010 Census Assessment, Testing, Evaluation, and Experimentation Program*, Internal Document, DRAFT.

U.S. Department of Commerce (1976), U.S. Census Bureau, *1970 Census of Population and Housing – Procedural History*, Washington, DC, 1976.

U.S. Department of Commerce (1989), U.S. Census Bureau, *1980 Census of Population and Housing – History. Part E*, Washington, DC, 1989.

U.S. Department of Commerce (1995), U.S. Census Bureau, *1990 Census of Population and Housing History, Part C*, Washington, DC, October 1995.

U.S. Department of Commerce (1996), U.S. Census Bureau, *1990 Census of Population and Housing History, Part D*, Washington, DC, March 1996.

Vitrano, Frank A., Robin A. Pennington, and James B. Treat (2004), *Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No.8, TR-8, Address List Development in Census 2000*, Washington, D.C., 2004.
