Research and Plans for Coverage Measurement in the 2010 Census: Interim Assessment

Panel on Coverage Evaluation and Correlation Bias in the 2010 Census, Robert Bell and Michael L. Cohen, Editors, National Research Council

ISBN: 0-309-10876-4, 118 pages, 8 1/2 x 11, (2007)

RESEARCH AND PLANS FOR COVERAGE MEASUREMENT IN THE 2010 CENSUS: INTERIM ASSESSMENT

Panel on Coverage Evaluation and Correlation Bias in the 2010 Census

Robert Bell and Michael L. Cohen, Editors

Committee on National Statistics
Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The project that is the subject of this report was supported by contract no. YA1323-04-CN-0006 between the National Academy of Sciences and the U.S. Census Bureau. Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (Number SBR-0112521). Any opinion, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number -13: 978-0-309-10875-1
-10: 0-309-10875-6

Additional copies of this report are available from The National Academies Press, 500 Fifth Street, NW, Washington, DC 20001 or (202) 334-3096; Internet, http://www.nap.edu

Printed in the United States of America

Suggested citation: National Research Council (2007). *Research and Plans for Coverage Measurement in the 2010 Census: Interim Assessment.* Panel on Coverage Evaluation and Correlation Bias in the 2010 Census, Robert Bell and Michael L. Cohen (Eds.). Committee on National Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PANEL ON COVERAGE EVALUATION AND CORRELATION BIAS IN THE 2010 CENSUS

ROBERT BELL *(Chair)*, AT&T Research Laboratories, Florham Park, NJ
LAWRENCE D. BROWN, Department of Statistics, The Wharton School, University of Pennsylvania
RODERICK J. LITTLE, Departments of Biostatistics and Statistics, University of Michigan
XIAO-LI MENG, Department of Statistics, Harvard University
JEFFREY S. PASSEL, Pew Hispanic Center, Washington, DC
DONALD YLVISAKER, Department of Statistics, University of California, Los Angeles (emeritus)
ALAN M. ZASLAVSKY, Department of Health Care Policy, Harvard Medical School

MICHAEL L. COHEN, *Study Director*
DANIEL L. CORK, *Senior Program Officer*
AGNES E. GASKIN, *Senior Program Assistant*
BARBARA A. BAILAR, *Consultant*
MEYER ZITTER, *Consultant*

COMMITTEE ON NATIONAL STATISTICS
2006-2007

Acknowledgments

In order to develop a sufficient understanding of the Census Bureau's coverage measurement plans for 2010, the Panel on Coverage Evaluation and Correlation Bias in the 2010 Census met often with Census Bureau staff. The panel met three times in plenary session (August 2-3, 2004; May 2-3, 2005; and December 6-7, 2005) and four times in small groups (December 14, 2004; January 24, 2005; July 21, 2005; and February 2, 2006). The latter sessions were devoted to more focused issues. In each of these meetings, Census Bureau staff made presentations that described the results of current research efforts and directions for further work. In addition, the Census Bureau made available to the panel staff (as sworn Census Bureau agents) data from the 2000 Accuracy and Coverage Evaluation for further investigation of the logistic regression models that the Census Bureau may use in 2010 to replace poststratification in the estimation of net coverage error.

The panel wishes to thank the many people who contributed to our work. The initial idea for the study came from Hermann Habermann, then deputy director of the Census Bureau. Other Census Bureau personnel were also instrumental in providing assistance. The contracting officer for this study was Philip Gbur, whose efforts should serve as a model of how best to provide for smooth communications between a National Research Council (NRC) panel and its sponsor. Donna Kostanich and her staff gave excellent summary presentations on the status of their various research efforts. Also, along with Philip Gbur, Donna Kostanich established a friendly, collegial environment between her staff and the panel.

We thank the following coverage measurement staff for their presentations: Tamara Adams, Paul Livermore Auer, William Bell, Pete Davis, James Farber, Gregg Diffendal, Rick Griffin, Tom Mule, Mary Mulry, Sally Obenski, Doug Olson, Robin Pennington, Preston J. Waite, and David Whitford.

The Census Bureau also provided on-site access to the A.C.E. Research Database. Huilin Li (University of Maryland) carried out many difficult computations on this database at the direction of the panel and staff, and we thank her for her patience and expertise.

As consultant to the panel, Barbara Bailar provided important insights on the history of coverage measurement and its implications for 2010. Also, Roger Tourangeau, a member of a sister NRC panel on residence rules in the decennial census, assisted the panel in learning about probes for alternative residences on both the coverage follow-up interview and the census coverage measurement interview.

The panel is indebted to Michael Cohen, who served as primary study director to the panel and who drafted much of the report in response to the panel's direction. Christine McShane provided expert technical editing of the draft report. Finally, Christine Chen, Lance Hunter, and Agnes Gaskin provided excellent administrative support for the panel.

It has been a pleasure overseeing this very talented and collaborative panel. In particular, panel member Alan Zaslavsky stands out for his detailed rewriting and editing of a preliminary draft of this report.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the NRC. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their participation in the review of this report: Eugene Ericksen, Department of Sociology, Temple University; David McMillen, External Affairs Liaison's Office, National Archives and Records Administration, Washington, DC; Samuel H. Preston, Population Studies Center, University of Pennsylvania; Keith Rust, Westat, Inc., Rockville, MD; Bruce D. Spencer, Department of Statistics, Northwestern University; and Martin T. Wells, Department of Social Statistics, Cornell University.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by Henry Riecken, Professor of Behavioral Sciences, Emeritus, University of Pennsylvania, and Ingram Olkin, Department of Statistics, Stanford University. Appointed by the NRC, they were responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report, however, rests entirely with the authoring panel and the institution.

Contents

Executive Summary

For the past three decades, coverage measurement of the decennial census has employed a postenumeration survey to provide estimates of net coverage error for subnational and demographic domains based on dual-systems estimation. Coverage measurement has three potential purposes: (1) to inform census data users about the quality of the counts for various uses, (2) to inform how census processes might be modified to improve the quality of the next census, and (3) to modify or adjust the census counts for official purposes.

In the 2000 census, the coverage measurement program was referred to as Accuracy and Coverage Evaluation (A.C.E.). A.C.E. was expressly designed primarily with the third purpose—adjustment of the census counts—in mind. However, a 1999 Supreme Court decision forbade the use of sampling, and therefore A.C.E. for generating census counts used for apportionment of the U.S. House of Representatives. Ultimately, inconsistencies in the A.C.E. results led to Census Bureau decisions not to use these data to adjust the 2000 census counts for redistricting or other official purposes.

For the 2010 census, the use of a coverage measurement program to adjust apportionment counts is still precluded by the 1999 Supreme Court decision. The use of a coverage measurement program as a basis for adjusting the census counts for legislative redistricting is seen by the Census Bureau as problematic, given the lack of time for a comprehensive evaluation of the quality of the adjusted counts. Given these limitations, as well as other considerations, the Census Bureau has decided to change the primary objective of coverage measurement in the 2010 census to that of providing information to improve the next census. This is consistent with Recommendation 6.1 of the report *The 2000 Census: Counting Under Adversity* (National Research Council, 2004b) and is supported here.

Although these three basic purposes of coverage measurement are related, they place different demands on a coverage measurement program. The focus of coverage measurement for adjustment is to estimate net coverage error; for census process improvement, estimates of net coverage error are insufficient, since they may hide offsetting errors arising from problems with census processes. For example, an erroneous enumeration may or may not be a duplicate of another enumeration; for net error measurement it is not crucial to know if it is a duplicate, but this question is important for improving census processes.

Therefore, the focus of coverage measurement in the 2010 census will be on exploring the four basic coverage errors: omissions, erroneous enumerations, duplications, and enumerations in the wrong place. In addition, the overall census design for 2010 is considerably different from that of 2000, the primary differences being that (1) the census long form will be eliminated, (2) the field enumerators will use hand-held computing devices for nonresponse follow-up, (3) the Master Address File/TIGER system will be updated and improved, (4) there will be a major effort to identify duplicate counts in the census and remove them from the final tabulations—this effort includes the collection of data on alternate residences and a national data search for duplicates, and (5) the coverage follow-up interview will be expanded to try to identify and rectify possible omissions from the census and enumerations in the wrong place. Despite these changes to the coverage measurement goals and the census itself, the Census Bureau plans to rely again on a postenumeration survey to collect data for coverage measurement and on dual-systems estimation to estimate net coverage error. Simultaneously adjusting to the new goals for coverage measurement and to a new census design raises a number of complex problems. The Census Bureau has requested the assistance of the National Academies to review and critique their test and research efforts to plan the coverage measurement program in 2010.

This interim report of the Panel on Coverage Evaluation and Correlation Bias in the 2010 Census describes and reviews the research activities carried out to date by the Census Bureau in developing the coverage measurement program for 2010. The panel will provide more direction in its final report on several of the technical challenges facing the Census Bureau raised by these research activities in working toward 2010. Chapter 4 provides a list of the topics the panel hopes to address. Those of particular importance are (a) the data to save in 2010 to support the various coverage measurement models, (b) random effects modeling for small area estimation, (c) treatment of nondata-defined cases in logistic regression, (d) allowable covariates in the logistic regression models for correct enumeration status and for match status, (e) sample design for the postenumeration survey in 2010, (f) improvements in demographic analysis in 2010, (g) the products to use to inform about census component coverage error, and (h) very generally, how to best operate a feedback loop for census improvement. In addition, the panel proposes an analytic framework that may suggest additional research activities, which may also be expanded in the final report.

THE 2010 COVERAGE MEASUREMENT RESEARCH PROGRAM

The Census Bureau has initiated a number of important projects in response to the need to redesign coverage measurement and related activities. These research activities include:

1. Design of the coverage measurement program for the 2006 census test to collect information on various operational parameters to accommodate the changing goals of coverage measurement and to assess the potential for

contamination if the coverage follow-up interview overlaps in time with the postenumeration survey interviews in 2010.

2. Development of a framework for defining types of census coverage error and the assumptions needed for their estimation.
3. The matching of postenumeration cases with census enumerations that have minimal information (that had been previously judged as having insufficient information for matching).
4. Refining methods for identifying and removing duplicate enumerations in real time.
5. The use of logistic regression for net error modeling, replacing the use of poststratification and synthetic estimation.
6. The modification of the A.C.E. sample design for the postenumeration survey used in 2010.

In addition, the Census Bureau is making impressive progress in the creation of merged, unduplicated lists (referred to collectively as E-StARS) from various administrative records of both residential addresses and persons, which could have important implications for both the census and coverage measurement in 2010.
The panel is impressed with the various research programs, which provide important information for use in coverage measurement in 2010. In this report, the panel offers advice on this research program in the following areas:

- Use of cross-validation for assessing alternative logistic regression models for estimating match probability and correct enumeration probability.
- Use of survey weights in the development and analysis of logistic regression models.
- Appropriate selection of covariates, in the logistic regression models for match and correct enumeration probability. Also, the use of random effects to incorporate small-area variation in these models.
- Sample design for the postenumeration survey to be used in coverage measurement in 2010.
- Use of administrative records for assisting with coverage measurement in 2010.

There is a substantial research literature on why people are missed in the census, as well as a more limited literature on why people are duplicated and erroneously enumerated. Furthermore, there remains substantial information from A.C.E. in 2000 on why census coverage errors were made on various households and people. Building on this base, the goal should be to develop statistical models that incorporate what is currently known about the sources of census coverage error and that help create a feedback loop from census coverage errors to deficient census processes. Further development of such statistical models after the 2010 census will benefit from the availability of linked data on (a) person, household, and area characteristics; (b) the specific census processes used to enumerate a person; and (c) whether the person was missed, erroneously enumerated, enumerated in the wrong place, duplicated, or correctly enumerated.

RECOMMENDATIONS

The panel offers four recommendations concerning coverage measurement plans for 2010. They are as follows:

Recommendation 1: The Census Bureau should evaluate, for use in the 2010 Census Coverage Measurement Program, a broader range of models, most importantly logistic regression models, for net coverage error that include variables in addition to those used to define the A.C.E. poststratification. These should include a wider range of predictors (e.g., geographic, contextual, family and housing variables and census operational variables), alternative model forms (e.g., classification trees), and the use of random effects to model small-area variation.

Recommendation 2: The Census Bureau should choose one or more of the proposed uses of administrative records (e.g., tax record data or state unemployment compensation data) for coverage improvement, nonresponse follow-up, or coverage measurement and comprehensively test those applications during the 2008 census dress rehearsal. If a process using administrative records improves on processes used in 2000, that process should be implemented in the 2010 census.

Recommendation 3: The Census Bureau should collect data in the 2010 census to support development of a database that links person, household, and housing unit characteristics, census processes, and the presence or absence of census component coverage error. This database should also represent coverage errors, including erroneous enumerations, enumerations in the wrong place, duplications, and omissions. The use of this database would better identify the sources of high rates of census component coverage error.

Recommendation 4: Given the number of important research activities currently under way, the needed design of the coverage measurement programs in the dress rehearsal and in the 2010 census, and the additional research suggested by the panel, the Census Bureau should provide the coverage measurement group with sufficient resources to carry out its current research program, its planning activities regarding the dress rehearsal and the 2010 census, and the activities listed in this report.

1

Introduction

The decennial census is an enormously complex endeavor. It requires counting residents in all types of living situations, from the densest urban setting to rural Alaska, in linguistically isolated areas and in gated communities, largely with a very temporary workforce that must be trained in only a few days. Given these circumstances, it is not surprising that the census counts are imperfect. Furthermore, even if an optimized census process could be developed for one census year, the dynamic nature of the United States population could make this process inefficient for the next census. It is therefore very important both to assess the quality of the census count and to learn as much as possible about what did and did not work well to inform future process improvements.

The Census Bureau has a 50-year history of carrying out careful assessments of the quality of its censuses. In particular, it has devoted substantial resources to the measurement of net coverage in the decennial census—that is, estimates of the difference between the census count and the "true" count, for various geographic and demographic groups.

Since 1978, panels of the National Research Council's (NRC) Committee on National Statistics have advised the Census Bureau on the assessment of census coverage. In particular, the Panel to Review the 2000 Census fully reviewed the operations, statistical methods, and results of the Accuracy and Coverage Evaluation (A.C.E.) program that the Census Bureau used to evaluate the coverage of the 2000 census. The description and evaluation of A.C.E. in that panel's final report (National Research Council, 2004b) is comprehensive and is referred to often in this report.

The work of the Panel on Correlation Bias and Coverage Measurement in the 2010 Census is the latest effort of the NRC to assist the Census Bureau as it plans the 2010 census. The panel is studying how to adapt census coverage measurement[1] to assess coverage better and to guide improvements in census processes.

---

[1]"Coverage measurement" and "coverage evaluation" are sometimes used synonymously. However, coverage measurement explicitly denotes a quantitative exercise, and coverage evaluation has more to do with the broad purposes of the activity, which is to assess, through a variety of operations, the completeness of the coverage through use of various quantitative and qualitative tools.

This report is the first installment in the panel's work. Some of the topics taken up in this interim report will be examined in more depth in the panel's final report.

COVERAGE MEASUREMENT IN THE DECENNIAL CENSUS

Uses of Coverage Measurement

Broadly speaking, coverage measurement potentially serves three primary uses: (1) assessment of coverage accuracy, (2) guidance for improvement of census processes, and (3) adjustment of reported counts.

Assessment of Coverage Accuracy

Census counts are used for many purposes vital to the nation, including the apportionment of seats in the U.S. House of Representatives (the constitutional mandate of the census); federal, state, and local redistricting; fund allocation to state and local jurisdictions; public planning; and learning about the population. Consequently, it is important for the nation to monitor the quality of population coverage overall and that of demographic or other groups. The purpose of the Census Bureau's coverage measurement programs for the 1950, 1960, and 1970 censuses was primarily to inform users as to the quality of census coverage.

Census Process Improvement

In addition to providing information to users about census quality, coverage measurement programs were also used to identify components of census processes that, if improved, could potentially reduce net coverage problems in the next census. For example, the relatively high undercoverage rate of black men ages 20 to 54 in the 1970 census motivated the implementation of several coverage improvement programs in the 1980 census. One such program, the nonhousehold sources program, which looked for names on certain administrative lists that did not match to census records, aimed to reduce *differential* coverage—that is, the difference between net coverage for a specific demographic group compared with that for the nation as a whole. However, the information that these coverage measurement programs provided was not very specific for identifying which components of the census process needed modification to address the measured undercoverage.

Adjustment

Starting with the 1980 census, an additional use was proposed for coverage evaluation programs, which was to use the information to adjust the census for undercoverage.[2] That is, the alternative counts produced through coverage measurement were to be used for some or all of the purposes to which census counts are applied, especially reapportionment and redistricting. This possibility was a result of technical improvements in coverage measurement methodology (discussed below) that allowed for estimation of net coverage error at the level of states. It was also due to the increased importance given to the uses of census data, such as to support general revenue sharing and redistricting. This use of coverage measurement information was also proposed prior to the 1990 and the 2000 censuses. However, adjustment has proved to be controversial, and to date adjusted counts have not been used for any official purposes, with the exception that the population controls for the Current Population Survey (and possibly some other surveys) were adjusted for census undercoverage during the 1990s.

Approaches to Coverage Measurement

The approach currently employed as the primary method for coverage measurement was introduced for the 1980 census, when it provided the first subnational geographic information on census undercoverage. A *postenumeration survey* (PES) is used to collect data on households from a random sample of census block clusters, referred to as the P-sample (arrived at independently of the census Master Address File or MAF). The responses from the survey are matched to the census enumerations to assess, for each of hundreds of population groups throughout the United States called *poststrata*, the rate at which individuals in the P-sample match to the census. The match rate for a poststratum serves as an estimate for the proportion of the true population captured in the census. In addition, the census enumerations in the P-sample blocks (referred to as the E-sample) are checked to estimate, again by poststratum, the proportion of census enumerations that are correct. The match rate and the correct enumeration rate are then used to estimate the rate of net coverage in each of the poststrata.

The specifics of the methodology are more complicated than this outline suggests. There are correlation bias and unmodeled local heterogeneous effects. (This report makes frequent mention of correlation bias, which is a bias in estimating the number of people missed by both the postenumeration survey and the census. This bias results from a departure from either the assumption of homogeneity of the enumeration propensities in the census and in the postenumeration survey, or from the assumption of independence of the two enumeration processes.) There is also unit and item nonresponse in both the census and the postenumeration survey due to a lack of full cooperation and because people move during census-taking, and there is also misresponse for a variety of reasons. Treatment of these issues greatly complicates the estimation of net undercoverage. Innovative approaches to deal with these complications have had varying degrees of success.

The estimation method that depends on the matching of two independent attempts to count a population is referred to as *dual-systems estimation* (DSE), and it has been the

---

[2]There were also post hoc proposals to adjust the 1970 census using synthetic estimates based on demographic analysis for intercensal purposes (e.g., Trussell, 1981).

cornerstone of the Census Bureau's coverage measurement efforts since 1980. The postenumeration survey and the DSE method that were used in the 2000 census are referred to as A.C.E.. An excellent description of A.C.E. methodology is given in U.S. Census Bureau (2003).

Even before the introduction of DSE, the Census Bureau began using demographic analysis to estimate net coverage error for demographic groups classified by age, sex, and black or nonblack.[3] Demographic analysis constructs an estimate of the population count at the census date for comparison with the census count using demographic accounting relationships. The data sources for demographic analysis include vital statistics for births and deaths, administrative data on immigration and the elderly, as well as analytic estimates developed from previous censuses and various surveys.

Demographic analysis can provide a useful alternative to DSE in measuring the national net coverage for the indicated demographic groups and providing measures of differential undercount for some demographic groups (age groups, sexes, and some race groups). However, because of limitations in the accuracy and precision of measures of internal migration, demographic analysis cannot be used to provide subnational estimates of undercount. Furthermore, demographic analysis cannot provide reliable estimates of net undercoverage for Hispanic populations, due to limitations in vital records and immigration data.

A.C.E., ADJUSTED CENSUS COUNTS, AND THE 2000 CENSUS

Detailed Investigations of A.C.E.

The understanding of Census Bureau officials, leading up to the 2000 census, was that if A.C.E. could provide reliable estimates of net coverage error for poststrata and if the estimated net error differed appreciably by poststrata, then adjusted population counts from A.C.E. would be used for redistricting and for other official purposes. The use of adjusted census counts for apportionment had already been precluded by the Supreme Court decision of 1999, which prohibited the use of sampling methods to produce counts for that purpose (*Department of Commerce v. United States House of Representatives,* 525 U.S. 316). As it turned out, however, several problems—including discrepancies between the initial A.C.E. counts and estimates from demographic analysis, concerns associated with balancing error (discussed below), the uncertain impact of a substantial number of late additions to the census, and the validity of whole-household imputations—led to a recommendation by the Census Bureau and the secretary of commerce's decision on March 6, 2001, not to use A.C.E. counts for redistricting.

[3]The first study of decennial census undercoverage assessed by demographic analysis may have been that of Price (1947) for the 1940 census. Other early applications are Coale (1955) for the 1950 census, and Siegel and Zelnik (1966) for the 1960 census.

The Census Bureau subsequently carried out research to determine the sources and magnitudes of error in the 2000 census, A.C.E., and demographic analysis. It also collected additional information relevant to specific concerns regarding A.C.E. The Census Bureau evaluated the extent of person duplication, and, on a sample basis, collected additional information to identify the correct residence for E-sample enumerations and to determine the correct match status for P-sample cases. This research included (1) the Evaluation Follow-Up Study, which involved reinterviewing a subsample of 70,000 people in E-sample housing units in 20 percent of the A.C.E. block clusters to determine the correct residents on Census Day (with additional clerical review of 17,500 people who were unresolved) and (2) the Person Duplication Study, which involved nationwide computer matching of E-sample records to census enumerations using name and date of birth. This nationwide search therefore permitted the first determination of the extent of remote duplication in the census, that is, duplication in which both housing units do not reside in the postenumeration survey block cluster. The Census Bureau also examined the implementation of the targeted extended search for matches to P-sample cases (extended search means searching outside of the relevant P-sample block cluster for a match for situations in which there is a likely error in identifying the correct block), the match rate and correct enumeration rate for people who moved during the data collection for A.C.E., and the impact of census imputations.

As a result of these very detailed investigations, the Census Bureau judged that A.C.E. counts substantially underestimated the rate of census duplication and hence tended to overestimate the true population size. (Mule, 2003, estimates a total of 9.8 million duplicates in the 2000 census.) The Census Bureau subsequently released revised A.C.E. estimated counts on October 17, 2001, which are referred to as A.C.E. Revision I counts. However, the Census Bureau recommended that these adjusted counts *not* be used for the allocation of federal funds or other official purposes.

Between October 2001 and March 2003, the Census Bureau undertook further review of all the data collected in the census and the A.C.E., as well as the subsequent matching and checking for enumeration status in the A.C.E. It also increased the A.C.E. estimates of the number of black males—which contained some apparent discrepancies that may have been related to correlation bias—based on matching sex ratios from demographic analysis.

The result of this effort is referred to as A.C.E. Revision II, along with a more extensive assessment of the error in the demographic analysis. On the basis of this work, on March 12, 2003, the Census Bureau announced that the A.C. E. Revision II counts, the final effort at coverage measurement in 2000, would not be used to produce intercensal population estimates.

In its final report (National Research Council, 2004b), the Panel to Review the 2000 Census generally agreed with the decisions made at each stage of this three-stage process, namely not to use the A.C.E. counts—either the original, Revision I, or Revision II—for purposes of redistricting, fund allocation, or other official purposes or for purposes of intercensal estimation. However, the NRC panel was not in complete agreement with the supporting arguments of the Census Bureau. The specific arguments made by that panel, a much more detailed description of A.C.E. and the various evaluation studies, and the material on which this abbreviated history is based can be found in *The 2000 Census: Counting Under Adversity* (National Research Council, 2004b: Chapters 5-6).

Extensive Documentation of the A.C.E. Process

As a by-product of this intensive effort to understand whether adjusted counts were preferable to unadjusted counts for various purposes, the Census Bureau produced comprehensive documentation and evaluation of the A.C.E. processes. A considerable amount of material is available to those interested in more information.

Evaluations supporting the March 2001 decision can be found at http://www.census.gov/dmd/www/EscapRep.html, evaluations supporting the October 2001 decision can be found at http://www.census.gov/dmd/www/EscapRep2.html, and evaluations supporting the March 2003 decision can be found out at http://www.census.gov/dmd/www/Ace2.html. Collectively, these reports document the A.C.E. procedures in detail, examining what was learned about the quality of A.C.E. and A.C.E. Revisions I and II through the additional information collected.

PLANNING FOR 2010

Shift in the Purpose of Coverage Measurement

The 2000 census demonstrated the great time and effort required to carefully collect data from the postenumeration survey, follow up the nonmatching cases, compute adjusted counts, and assess their quality in comparison to the census counts. On the basis of that experience, the Census Bureau concluded that there would not be the time needed to perform coverage measurement for adjustment of the counts used for redistricting by the mandated date of April 1, 2011, one year after census day (see, e.g., National Research Council, 2004b: 267).

Although that decision did not rule out the possibility of adjustment for other purposes (e.g., intercensal estimates, fund allocation), it dramatically shifted the focus of the coverage measurement program for 2010. Even so, this shift does not reduce the importance of conducting a high-quality coverage measurement program. Evaluating the accuracy of coverage will remain at least as important in 2010 as it has been in previous censuses (perhaps more so, given such innovations as plans to delete duplicates in real time). In addition, the panel thinks that increased attention should be paid to the use of coverage measurement to inform efforts to improve census processes for the future.

This shift does have implications for the desired output of coverage measurement. Census adjustment requires accurate estimates of *net* coverage at various levels of geography and for other population divisions, but given those estimates, information about components of error—the numbers of omissions, erroneous enumerations, duplicates, and enumerations in the wrong place—is basically irrelevant. If DSE concludes that there was a 1 percent net undercoverage for some group, the adjustment is the same whether that net undercoverage resulted from 3 percent omissions less 2 percent erroneous enumerations or from 8 percent omissions less 7 percent erroneous enumerations. For that reason, past implementations of DSE have not been designed to separately estimate the numbers of omissions and erroneous enumerations.

In contrast, any evaluation of census quality should take into account information about both net coverage and components of error. The two scenarios mentioned in the last paragraph would lead to very different conclusions about the overall quality of the basic census processes, as well as confidence about any conclusions from coverage measurement. Likewise, information about specific components of error is critical to the use of coverage measurement to inform efforts to improve census processes. In this case, it is important not only to know the frequency of specific types of errors, but also to identify those cases accurately in the coverage measurement sample, so that errors can be linked to specific census processes for subsequent analysis.

As a consequence of this change in the primary purpose of coverage measurement, that is, in support of census process improvement, the Census Bureau is putting much greater emphasis in 2010 on measuring the components of coverage error. However, the 2010 census coverage measurement (CCM) program will again rely on a postenumeration survey as the primary data collection in support of census coverage evaluation. Given the new focus on estimation of rates of census component coverage error and on developing a feedback loop in support of census improvement, while keeping in mind that all three goals of coverage measurement above remain important, three questions are raised that this panel has considered and will address more completely in its final report:

1. How well can the goal of census component error measurement be met using an approach that was initially developed to measure net coverage?
2. What modifications to the A.C.E. sample design would provide a CCM sample design that is more effective for this new purpose?
3. How well can components of census coverage error be linked to the associated census processes?

There is also the real possibility that decennial census management information data, which may not be routinely saved, may be useful in providing additional information on the functioning of some specific historically problematic census component processes. We also point out that an advantage of a postenumeration survey is its omnibus nature, providing information for any unanticipated problems in the census. It is also not yet clear what analyses will be most useful in diagnosing census deficiencies, given the data that are available.

In summary, this change in the focus of coverage measurement has implications for CCM sample design, data collection more broadly, the production of coverage measurement statistics and databases, and subsequent data analysis. These questions are given some consideration in this report, and we will provide additional advice in our final report.

Problems with A.C.E. in 2000

In addressing these and other questions, the panel has taken into consideration the previous findings of both the Census Bureau and the Panel to Review the 2000 Census regarding the limitations of the A.C.E. design for addressing both the previous goal of estimating net coverage and the new goal of measuring components of census coverage error.

- First and foremost, inadequate information collected as part of the census and the PES allowed too many mistakes in the A.C.E. final determination of Census Day residence. This problem was demonstrated most vividly for duplicates, but only learned well after the PES operation had been completed. Consequently, even when duplicates were identified, there was generally no basis for selecting one location as the place of the correct enumeration. In addition, there was some evidence that A.C.E. underestimated the number of omissions. (For details, see U.S. Census Bureau, 2003.)
- Demographic analysis provided evidence of correlation bias for at least black men. However, it is unclear whether the correlation bias "correction" applied to counts for black men was successful. Furthermore, due to the lack of data on ethnicity in vital statistics, this approach was not available for nonblack Hispanics, a group that might be expected to have similar levels of correlation bias. (For details, see Bell, 2001, and Haines, 2002.)
- The approach taken to estimate net census coverage error relied on balancing erroneous enumerations against omissions in cases in which there was insufficient information to match E-sample and P-sample cases. Consequently, A.C.E. was not effective at estimating components of census error. (For details, see Adams and Liu, 2001.)
- The poststratification in A.C.E. (which tries to partition the U.S. population into relatively homogeneous subgroups to reduce correlation bias) was constrained to use a very limited number of variables. Because the approach cross-classified many of the factors, each additional factor greatly increased the number of poststrata and correspondingly reduced the sample size per poststratum.[4] (For details, see U.S. Census Bureau, 2003.)

The remainder of this report is primarily concerned with the Census Bureau's plans to address problems and the panel's assessment of those plans.

[4]In fact, collapsing of poststrata was needed because many of the cross-classified cells had such small sample sizes.

How the Plans for the 2010 Census Differ from the 2000 Census

Any consideration of changes to coverage measurement plans for 2010 should account for how plans for the 2010 census differ from those for the 2000 census. We list six major differences here; more details are provided in Chapter 3.

1. With the American Community Survey, a continuous implementation of the census long form, now in full operation, it is anticipated that the 2010 census will use the short form only.
2. Current plans are for field staff to use handheld computing devices during nonresponse follow-up for data collection, data transmission, real-time editing and error correction, and navigation to assignments.
3. There are currently efforts to improve both the Census Bureau's MAF and its geographic referencing system, TIGER (Topologically Integrated Geographic Encoding and Referencing).
4. The Census Bureau plans to add selected coverage improvement questions to the short form asking whether there are alternative households in which someone may have been enumerated and whether there were any other people who sometimes live in the household.[5]
5. Use of the coverage follow-up (CFU) interview will be greatly expanded compared with 2000. Additional households that are planned to be followed up in 2010 include households with a possible duplicate enumeration, other addresses at which at least one resident sometimes lives, and those with other people who sometimes live in the household. This additional data collection close to the time of the CCM interviews may pose a contamination threat (i.e., the CCM interview may affect the census in the CCM blocks, making the CCM blocks unrepresentative), so the Census Bureau has asked the panel to examine a number of ways of addressing this possible problem. This is addressed in Chapter 3.
6. Using information from the main census returns and the CFUs, the Census Bureau plans to delete from the census households persons identified as duplicates counted in the wrong place.

INITIATIVES FOR IMPROVEMENTS IN THE 2010 COVERAGE MEASUREMENT PROGRAM

In response to the change in the objectives for coverage measurement in 2010, the various limitations of A.C.E. to address those objectives, and the changes currently planned for the 2010 census in relation to the 2000 census, the coverage measurement staff of the Census Bureau, and decennial staff in general, has undertaken several important initiatives likely to improve the coverage measurement program for 2010. These include

[5]The addition of these questions has been cognitively tested, and a report on this from the Census Bureau is expected soon.

- **Estimating (and adjusting for) correlation bias**. The Census Bureau has made only limited progress to date to directly address this difficult problem. However, the additional data collection mentioned above, other potential improvements in coverage measurement, and the use of logistic regression modeling (discussed below) provide some hope for reducing the size of correlation bias. Whether the Census Bureau will implement the correction based on sex ratios used in 2000 is unclear.
- **Estimating components of census coverage error**. The Census Bureau has produced an excellent report on the definition of census coverage component error and its measurement, "Framework for Coverage Error Components" (U.S. Bureau of the Census, 2005). This report greatly clarifies what component errors are to be measured and the assumptions underlying their measurement. In addition, the Census Bureau has several research initiatives to improve measurement of the rates of census component coverage errors. These include reducing matching error through collection of additional information on people's residence and through attempts to match people with very limited information (discussed below). In addition, the Census Bureau views the estimation of remaining matching error as a large missing data problem, and it may apply multiple imputation techniques to provide better estimates. However, little work on this has been initiated to date. Finally, the Census Bureau would like to incorporate estimates of census omissions that take into account correlation bias when it estimates census component coverage error. This is a particularly challenging problem.
- **Improving net coverage estimation**. The Census Bureau has been developing an alternative approach to net coverage measurement by replacing measurement at the poststratum level of average match rate and average correct enumeration rate with logistic regression models of both match and correct enumeration probabilities at the level of the individual person. This approach accommodates more predictive factors than poststratification, and it allows use of continuous predictors. Also, this alternative to poststratification accommodates greater heterogeneity in match rates and correct enumeration rates. Using this approach may provide improvements through (1) reduction of bias through more flexible variable selection, (2) more options for handling missing data, and (3) reduction of unmodeled local heterogeneous effects. (For details, see Griffin, 2005a.)
- **Designing the CCM sample of block clusters**. Although the design of the CCM data collection will, in many respects, approximate the design of the A.C.E., the CCM might differ from the A.C.E. in terms of the design for sampling block clusters, to better support the new objectives of CCM in 2010.
- **Finally, measuring residency status more reliably in the census**. Through revisions of the census questionnaire and the CFU interview, the Census Bureau will be collecting more information on possible alternative households in which someone may have been enumerated, as well as more information on possible duplicate enumerations. The anticipated result is more reliable matching, more reliable assessment of correct enumeration status, and more reliable assessment of duplicate status. This additional data collection puts a large demand on census

field staff, and it may have implications for the timing of various census operations. It is therefore unclear how much of this additional data collection will be feasible in 2010.

PANEL CHARGE AND APPROACH

To recapitulate: the Census Bureau is planning, in 2010, to return to a previous, although substantially expanded objective for CCM, which is to assess the amount of census component coverage error, both to inform users as to the quality of the census counts, but more importantly to support examination of ways of improving census-taking for the next census. To provide more targeted information for the latter purpose of improving census-taking over time, the ultimate hope is to attribute the various types of census errors to particular census processes and, as a result, to concentrate efforts for improvement on the parts of the census that are most in need.

At the Census Bureau's request, the National Academies established the Panel on Coverage Evaluation and Correlation Bias in the 2010 Census to examine coverage measurement plans for 2010. The panel's charge reads as follows:

> This project involves a study of four issues concerning census coverage estimation with the goal of developing improved methods for use in evaluating coverage of the 2010 census. A panel of experts will conduct the study under the auspices of the Committee on National Statistics of the Division of Behavioral and Social Sciences and Education. The panel is charged to review Census Bureau work on these topics and recommend directions for research. The panel's work may require development of statistical models to extend the DSE approach, and may also include suggestions for the use of auxiliary data sources such as administrative records. DSE, as applied to the 1990 and 2000 censuses, had several benefits as well as limitations as a means for estimating net census coverage. Some of the limitations were

1. The approach was designed for estimating net census coverage errors and did not provide accurate estimates of gross coverage errors, i.e., of gross census omissions separate from gross census erroneous enumerations. In the DSE approach applied in the 1990 and 2000 censuses, certain census enumerations classified as erroneous were balanced against certain coverage survey cases classified as nonmatches (census omissions) for the purpose of estimating net census coverage. Some of these paired census enumerations and coverage survey cases did not necessarily reflect gross errors.
2. The application of DSE in A.C.E. Revision II during the 2000 census accounted for duplicates found in the census in a simplistic way due to lack of information as to which member of a duplicate pair was a correct enumeration and which was an erroneous enumeration. This led to

estimation error, as did the simplistic treatment of A.C.E. cases (P-sample) that matched to census enumerations outside the search area.

3. The post-stratification approach used to apply the DSE had certain limitations. First, the number of factors that could be included in the post-stratification was limited because the approach cross-classified the factors, so that each factor added to the post-stratification greatly split the sample. (Collapsing of post-strata was needed because many of the cross-classified cells had small sample sizes.) Second, the synthetic error that arose from the synthetic application of the post-stratum coverage correction factors to produce estimates for subnational areas and population subgroups was not reflected in their corresponding variance estimates.
4. Comparisons of aggregate tabulations of DSEs with estimates from demographic analysis (DA), in both 1990 and 2000, suggested underestimation by DSE of persons missed by both the census and the coverage survey (correlation bias). In the 2000 A.C.E. Revision II, sex ratios from DA were used to determine factors to correct adult male estimates for correlation bias, assuming no correlation bias for children and adult females. This approach appeared effective for adult blacks, but there were concerns about the appropriateness of its assumptions for other race/origin groups (particularly Hispanics). Also, DA totals for young children (0-9) exceeded the corresponding aggregated DSEs from A.C.E. Revision II by a sufficient amount to suggest possible correlation bias in estimates for young children. The Census Bureau is interested in improving the DSE methodology to address the above issues to the extent possible, to develop improved methods for estimating coverage of the 2010 census both in regard to net errors and gross errors.

We interpret the charge to the panel as follows: to evaluate the Census Bureau's plans and to provide suggestions and recommendations for changes and additions to those plans, in determining how coverage measurement and related activities might be used to measure the components of census coverage error and thereby assess the role of the various census component processes in contributing to coverage error.

The original charge to the panel had three areas of focus: (1) the treatment of duplicates, (2) the use of alternative approaches to poststratification, especially model-based alternatives, and their impact on the ability to model local heterogeneous effects, and (3) the use of demographic analysis to correct for correlation bias. It was understood from the outset that the panel's work might involve assistance in the development of statistical models to modify or extend the dual-systems approach, and it might also include suggestions for the use of auxiliary data sources, such as administrative records, apart from their use in demographic analysis. While these areas are still of interest to the Census Bureau and to the panel, since the panel has started its work, the needs of the Census Bureau in the area of coverage evaluation have broadened. As a result, the panel has also been asked to review and examine additional issues related to coverage evaluation not explicitly mentioned in the original charge.

Specifically, the panel has been asked to (a) examine the Census Bureau's draft document providing a framework for the definition of component errors and estimation of their rates of occurrence, (b) examine the possibility of estimating the match status of cases previously categorized as having insufficient information for matching, in order to reduce the number of cases identified as erroneous enumerations due solely to item nonresponse, and (c) assess various alternatives that reduce or avoid the contamination likely to result from the similarity and simultaneity of the census CFU and the PES interviews in 2010.

The Census Bureau has also asked for the panel's views on a number of other issues, including the CCM postenumeration survey design and the form of the census CFU interview and the CCM initial and follow-up interviews. In addition, the Census Bureau is interested in having the panel look at other issues listed above as limitations for A.C.E. in addressing the new goals for coverage measurement in 2010, suggesting alternatives that could be implemented in time for 2010. Finally, part of this review is to evaluate the broad research priorities of coverage evaluation at the Census Bureau, leading up to the 2010 CCM, and to provide advice as to whether the priorities should be altered in light of the broader goals described above.

The general data collection and matching operations of the 2010 CCM are taken as fixed. That is, we take as given that the CCM program will include a sizable postenumeration survey that will be matched to the census to assess match status for a sample of census block clusters. Given this, the panel is examining the alterable aspects of the data collection for the 2010 coverage measurement program, including sample design, to see if improvements can be recommended. The panel will *not* address the broader issue of what type of coverage measurement program, that is, what alternative to CCM, would best support improvement of census-taking over time.

Furthermore, all the data retained from the 2010 census—not only the postenumeration survey and matching results, but also data collected by the various management information and quality assurance systems that monitor census processes—could affect the coverage measurement models that could be developed. Therefore, the panel will also advise on what data should be retained from the 2008 census test and the 2010 census.

The panel also asserts that many of the design questions for the 2010 census and its coverage measurement program must be further informed through greater use of the data collected in 2000. We also consider how the Census Bureau can further exploit those data to improve the CCM design.

The possibility remains that there will be a sizable differential undercount in 2010. One such scenario would arise if the 2010 census design is very effective in deleting duplicates in real time, but no more effective than the 2000 census in reducing census omissions. The result could then be a substantial differential undercount that one would like to reduce through the use of modified counts. We view a substantial differential undercount as an unlikely contingency, but what would be done in that event is deserving of greater consideration by the Census Bureau.

Finally, the Census Bureau's current program for research on coverage measurement is not as comprehensive as might be desired. The panel has therefore slightly expanded our scope in this report by suggesting additional activities that would support component census coverage error measurement. By doing this, we hope to encourage the Census Bureau to allocate greater resources to this effort in the years remaining prior to 2010.

ORGANIZATION OF THE REPORT

Following this introduction, this report consists of three chapters. Chapter 2 defines the components of census error, describes how census errors are measured through the use of DSE and demographic analysis, and then outlines the three purposes of census coverage measurement: the measurement of census quality, census process improvement, and potential census adjustment. Chapter 3 describes and assesses the Census Bureau's current research program on coverage evaluation. It begins by listing the limitations of the 2000 A.C.E. for measuring component census errors and describing differences between the 2010 and 2000 census plans as well as plans for the coverage evaluation program in the 2006 test census. Next it describes the major topics of the current coverage evaluation research program, including measuring components of census error, models for net coverage error, contamination due to the extension of the CFU interview, the sample design for the CCM postenumeration survey, and use of the E-StARS administrative records system in coverage measurement. Chapter 4 describes the value of integrating census process data, and person, household, and area characteristics data, with census component coverage error data. It further argues that 2000 A.C.E. data can still be used to inform the design of the coverage measurement program in 2010. Finally, the issue of user requirements for documentation and tabulation of census coverage errors in 2010 is raised.

2

Back To Basics:
What Are Census Errors and How Can They Be Measured?

As a result of the issues that the Census Bureau has raised regarding census coverage measurement, it is useful to go over some concepts needed for the remainder of this report. First, we address the various types of census error that can occur, defining omissions, duplications, erroneous enumerations, and errors of geography and demographic characteristics, and consider the adequacy of these terms to categorize types of error.

We next describe and assess the two basic approaches to error measurement currently used by the Census Bureau, dual-systems estimation (DSE) and demographic analysis. We then discuss the use of coverage measurement for assessment of census quality, to support census process improvement, and for adjustment of census counts.

TYPES OF CENSUS ERRORS

Coverage errors in census enumerations are of two types: inclusion of people in the enumeration who should not have been included, and omission of people who should have been included. People mistakenly included in the census comprise two types. First, erroneous enumerations are those who should not have been included in the census because they were not residents of the United States on Census Day, such as babies born after Census Day, people who died before Census Day, temporary visitors, and fabricated people. Second, there are duplicates of correct enumerations, representing people who appear more than once in the list of census enumerations. Duplicates can be repeat enumerations of the same individual at the same address, either as a result of the multiple opportunities for being enumerated in the census, or from an address being represented in more than one way on the Census Bureau's Master Address File (MAF). Duplicates can also result from the inclusion of an individual at two different residences, possibly both of which are part-time residences. (We do not consider whole-person imputations or whole-household imputations, used either when an enumeration has less than two characteristics or when the number of persons living at a housing unit is estimated, to be a source of either duplications or erroneous enumerations, but rather to be a means for producing counts that are as accurate as possible when aggregated to various levels of geography.)

People who were not included in the list of census enumerations but should have been are census omissions. Omissions can result from a missed address on the MAF, a missed housing unit in a multiunit residence in which other residences were enumerated, a missed individual in a household with other enumerated people, and people with no residence.

In addition to omissions, erroneous enumerations, and duplications, enumerations in the wrong location can also affect the accuracy of census counts. A count in the wrong location can result from (1) a misunderstanding of the census residence rules and the resulting reporting of someone in the wrong residence—for example, having an enumerator assign someone to the wrong choice from among several part-time residences, and (2) placing an address in the wrong census geographic location (called a geocoding error).

Furthermore, demographic errors, which occur when a person's demographic characteristics are incorrectly reported or assigned and which can also result from an improper imputation of an individual's demographic characteristics, can add error to census counts. For example, if someone's age is misreported on the census form, this adds one tally in error to the count for one age group and subtracts one tally in error for another. However, this does not impact census counts that are not disaggregated by age group.

Erroneous enumerations and omissions contribute to errors in census counts for any geographic aggregate that includes the addresses of the persons involved with those errors. Whether or not errors in geographic or demographic characteristics result in errors in census counts depends on the level of demographic and geographic aggregation for which the census counts are used. The more detailed the geographic and demographic domain of interest, the greater the chance that errors in geographic and demographic detail will affect the quality of the associated counts. For example, placing a person in the wrong census tract but in the right county is not an error for census applications except when one uses census counts below the county level. However, placing someone in the wrong state affects most uses of census counts. Similarly, attributing someone to the wrong age group does not affect overall population counts at any level of geographic aggregation, but it will result in an error for counts by age group.

As touched on above, errors in census counts can result from missing information and the resulting use of imputation for item and unit nonresponse. For example, missing information on the total number of residents in a housing unit can result in imputation of this number, which can add to errors in census counts of the total population for areas containing that housing unit. As described in National Research Council (2004b: 128, Box 4.2) the 2000 census used item imputation, whole-person imputation, and four types of whole-household imputation to complete responses with missing information. The procedure used depended on which persons in a household were and were not census data-defined. (A person's enumeration was data-defined if there were at least two basic data items reported, including name as an item). Item imputation was used when all members of a household were data-defined, but some basic items were not reported. Whole person imputation was used when at least one member of a household was data-defined, but at least one other member was not. (Therefore, any enumeration that is not data-defined results in a whole-person imputation.) For the members of a household who were not data-defined, all basic information was imputed, using characteristics of other household members. Finally, four types of whole-household imputation were used, depending on whether the number of residents was known, the number of residents was not known but the occupancy status was known, the occupancy status was not known but it was known that the address was a housing unit, and finally it was not known whether the address was a housing unit.

As discussed in Chapter 3, the result of an individual imputed enumeration should not be considered to be correct or incorrect, but rather, one should assess an imputation algorithm based on its contributions to the bias and variance of estimated counts for various geographic areas and demographic groups. Therefore, whole-person and whole-household imputations can increase the errors in census counts for any demographic and geographic domains containing the people in question. Furthermore, imputation of characteristics can impact the quality of the counts for the associated demographic groups.

Two approaches have been taken to date to assess the overall (coverage) quality of census counts. One view is that census quality should be measured, separately by domain, by estimating the percentage net coverage error for each domain, for example, for each state. A second view is that census quality for a given domain should be measured instead by the percentage of census error—by census error we mean the totality of omissions, erroneous enumerations, duplications, and errors in the wrong location, with all errors receiving the same weight. This statistic is often referred to as the rate of gross census error.

As explained in Chapter 1, the Census Bureau is moving away from the view that the primary measure of census quality should be net error, because net error ignores errors of omission and erroneous enumeration and duplication that balance out for some levels of aggregation. On one hand, such errors that cancel at some level might contribute to error in measures at a lower level at which they do not cancel. On the other hand, the rate of gross census error is also deficient as a summary measure, in that many enumerations in the wrong location will affect only the more detailed aggregates. This argues for separate treatment of errors in the wrong location. Furthermore, since component coverage errors have partially distinct causes, it is important to separate the summaries of these various components so that their magnitudes can be assessed individually, rather than trying to place them into a single error measure. These last two points argue for separate measures of the four components of census coverage error: duplicates, erroneous enumerations, omissions, and enumerations in the wrong location. In addition, for errors in the wrong location, rather than a percentage error measure, which would be appropriate for omissions, erroneous enumerations, and duplications, a summary assessment would require a representation of the distribution of the size of the geographic errors to assess which applications of the counts are likely to be affected by various magnitudes of errors.

Measures of component census error consistent with the above considerations will provide useful information in support of a feedback loop for identifying alternative census processes that are preferred to current ones. However, this does not mean that the Census Bureau should not also continue to provide estimates of net coverage error. Such measures still have importance since (1) they can be compared with previously published estimates for historical comparisons of census quality, (2) as discussed in Chapter 3, net error measures are needed for estimating census omissions, and (3) users find net error measures useful for evaluating the utility of estimates for some applications.

HOW CENSUS ERRORS ARE MEASURED

In this section we provide some additional detail concerning the two main approaches to coverage measurement that were outlined in Chapter 1: DSE and demographic analysis.

Dual-Systems Estimation

A detailed description of DSE, which has been used as the primary methodology for coverage measurement for the last three censuses, can be found in National Research Council (2004b: 159-163 and Chapter 6) and in U.S. Census Bureau (2003). A history of DSE can be found in National Research Council (1985: Chapter 4) and in Cohen (2000). We provide a brief outline here.

A *postenumeration survey* (PES) is conducted, following the census data collection in any given housing unit, although possibly partially overlapping in terms of the overall schedule. This is a survey of the residents in a sample of census block clusters, who are referred to collectively as the P-sample. The addresses in those blocks are listed independently of (that is, not using any information from) the Census Bureau's MAF, which is the address list used to take the decennial census. Then the housing units at those addresses are interviewed to establish who was resident on Census Day. Additional information is also collected to support matching to the census and to assign the persons to poststrata, which are defined by demographic characteristics, as well as household and area characteristics. For example, mailback rates and whether someone is an owner or renter, along with demographic and other characteristics, are used to define poststrata. The characteristics used to define poststrata are those that have been associated with the propensity to be missed in past censuses. Given the heterogeneous coverage properties across the poststrata, the coverage measurement described here is carried out separately by poststrata.

The P-sample enumerations are then matched to the census enumerations to determine who in the P-sample was also counted in the census. Persons who failed to match to the census are reinterviewed, to determine the reason for the failure to match, and to make any needed corrections due to discovered errors in the data collection or matching.

The *census enumerations* in the P-sample block clusters are referred to as the E-sample. The E-sample is used to determine the percentage of the census enumerations resident in the P-sample block clusters that are correct. This is accomplished by visiting the E-sample people who fail to match P-sample records to determine whether each individual was enumerated in the census, or whether they were enumerated in error.

The independence of the P-sample enumerations and the census enumerations is crucial to support the estimation of census undercoverage for the following reason. The fundamental relationship underlying this approach to the estimation of census net undercoverage is that, poststratum by poststratum, the following approximate equation should obtain:

$$\frac{M}{P} \cong \frac{C}{DSE}$$

where:

- M stands for the estimate of the number of P-sample persons who match with an E-sample person,
- P stands for the estimate of the number of all valid P-sample persons,
- C stands for the number of census enumerations, and
- DSE stands for the dual-systems estimate of the total number of residents, that is, the estimated true count.

This approximate equation should hold because, ignoring some complications, the first ratio (M / P) is an estimate of the percentage of census enumerations in the subpopulation of P-sample enumerations, that is, an estimate of the census "capture" rate within the P-sample population, and the second ratio (C / DSE) is an estimate of the percentage of census enumerations in the full population (all within some poststratum). If the P-sample selection and field measurement processes are independent of the census processes, and if the *operational* independence of the census and the PES also engenders *statistical* independence, then the fact of enumeration in the P-sample should provide no information as to whether a person was or was not enumerated in the census. Therefore, the subpopulation of P-sample enumerations should have the same underlying probability, conditional on poststratum, of being enumerated in the census as the full population. Given that, *and temporarily ignoring erroneous enumerations, duplications, and whole-person imputations in the census*, these two ratios should be approximately equal (except for sampling and other variation). The above relationship can be reexpressed as $DSE = C\left(\frac{P}{M}\right)$, that is, the estimate for the total population size is the product of the number of census enumerations times the number of P-sample enumerations, divided by the number of matches. The calculation of estimates within poststrata is motivated by the additional assumption that both the census and the PES enumeration propensities are uncorrelated, which is supported by the homogeneity of coverage properties within poststrata. Failure of this assumption results in correlation bias.[1]

This derivation ignores the key role of the E-sample, which provides a needed correction to the above, given that a percentage of census enumerations are either duplicates, erroneous (including in the wrong location), or whole-person imputations and therefore not able to be matched to the P-sample. To address this, *C*, the census count, in the above formula is replaced by $(C - II)\left(\frac{CE}{E}\right)$, where *II* represents the number of people lacking sufficient information for matching, *CE* represents an estimate of the number of E-sample persons correctly enumerated in the census, and *E* represents an estimate of the number of E-sample enumerations in the P-sample block clusters. (Note that *E* is a sample weighted quantity, whereas *C* is not). The number of people lacking sufficient information for matching, *II*, is subtracted from the census count since their match or correct enumeration status cannot be determined. The assumption is that their net coverage error is the same as that for the remaining census enumerations. The number of matchable persons, $C - II$, is multiplied by $\frac{CE}{E}$ to estimate the percentage of matchable persons that are correct enumerations, that is, we multiply the matchable count by the percentage of correct census enumerations. The resulting DSE formula is

$$DSE = (C - II)\left(\frac{CE}{E}\right)\left(\frac{P}{M}\right).$$

This derivation still ignores several additional nontrivial complications, including the treatment of other forms of missing data, the treatment of data from movers, and the precise area of search for matches of census enumerations outside the P-sample blocks. Other complications arise in Revisions I and II of Accuracy and Coverage Evaluation (A.C.E.) due, among other causes, to incorporation of information from the Matching Error Study, the Evaluation Follow-up Study, and the Person Duplication Study.

Problems with Dual-Systems Estimation

Both the decennial census data collection and the data collection for the PES inevitably involve errors and unit and item nonresponse. As a result, matching errors are made, quite likely more in the direction of false nonmatches than false matches. Furthermore, while the use of poststrata is intended to partition the population into subgroups that have relatively homogeneous propensities to be enumerated in the census (in order to reduce correlation bias), the poststrata are still not completely homogeneous.

---

[1]For a more detailed exposition of the errors in dual-systems estimation, see Alho and Spencer (2005: Chapter 10).

Missing data complicate the application of DSE in the following ways. While erroneous enumerations in the E-sample that have sufficient information for matching are typically identified as erroneous (though there are cases for which correct enumeration status has to be imputed), as discussed above, erroneous enumerations that have insufficient information for matching (or are nondata-defined—see the discussion of what are called KEs in Chapter 3) are removed from the computation and are assumed to behave like the remaining E-sample enumerations in their poststratum through an implicit reweighting adjustment. This assumption can be examined using studies like the Evaluation Follow-Up Study in 2000. For duplicates, data-defined enumerations with name and date of birth in the E-sample within the P-sample block clusters are typically discovered, but until 2010, those duplicates outside the P-sample blocks were categorized as erroneous enumerations. These cases will now be identified as duplicates in 2010, assuming the national search for duplicates is implemented.

P-sample persons with sufficient information for matching that are missed in the census are typically correctly identified as census omissions. However, cases with insufficient information for matching are accounted for by giving additional weight to those cases with sufficient information that are similar on available characteristics thought to be predictive of match status. The validity of these weighting models can also be examined using such studies as the Evaluation Follow-Up Study in 2000. In previous censuses, when the corresponding census enumeration was located outside the P-sample block, a number of P-sample persons that were not omissions were designated as such, resulting in overestimation of the number of omissions. However, this should be addressed in 2010 with the implementation of the national search for matches. Finally, the number of P-sample persons missed both in the P-sample and the census is estimated assuming both independence of the two enumerations and homogeneity of enumeration propensities (the absence of which engenders correlation bias). However, since no data are collected for this group, it is unclear how well this group is estimated (although merged administrative records might be used for this purpose). The general expectation is that this group of census omissions is underestimated.

In addition, A.C.E. and its predecessors in 1980 and 1990 were not designed to distinguish among different types of census errors. An important limitation in this regard arises from the restriction of searches for E-sample matches to P-sample enumerations to be either in the P-sample block cluster or sometimes slightly outside in a (targeted) extended search area. Given this, a failure to match a P-sample enumeration to the census could result from any of several types of error, including (a) a person's name and date of birth were captured with substantial error (possibly by the optical character recognition used in scanning the census form), (b) a housing unit was erroneously geocoded a few blocks outside the search area, or (c) the census enumeration was mistakenly located at a 3-month winter residence rather than at the 9-month residence during the remainder of the year. These situations are all represented as an omission in the census of the associated P-sample enumeration, along with an erroneous enumeration in the census either (a) at the correct residence, (b) a few blocks away, or (c) possibly hundreds of miles away. For some applications of census counts, these errors will cancel each other, and for others they will not. For example, in counting the population for states, geocoding errors of short distances are unlikely to matter.

The local restriction of the search for matches in the 1980, 1990, and 2000 censuses was intended to overstate the number of census omissions and the number of erroneous enumerations by the same amount. If one is interested only in net error, the intention is that these errors would balance out, resulting in a zero net effect. (There is, however, an increase in the variance of the estimate of net coverage error due to the need for these random amounts to balance out for various domains.) Furthermore, various deficiencies in the operation of the field processes will cause the balance to be inexact even in expectation, leading to "balancing error." However, for the new objective of assessment of census component coverage errors, the (necessarily) restricted search area results in a substantial increase in the estimated rates of omission and erroneous enumeration, much of which is due to counting someone in the wrong location, which may not be an error for many applications of census data.

Finally, errors in geography or demographics can also result in the placement of individuals in the wrong poststratum, which can also bias the estimation of net coverage error.

Demographic Analysis

The Census Bureau has made substantial use of demographic analysis for several censuses, going back to 1940 (see for example Price, 1947; Coale, 1955; Coale and Zelnik, 1963; Coale and Rives, 1972). We present a short overview here; for a more detailed treatment relevant to the 2000 application, see Robinson (2001).

Demographic analysis makes use of the following "balancing equation" to estimate the population in an age group from historical data sources:

$$P = B - D + I - E,$$

where

P = the population in the age group at the census date;
B = births (or the population at a previous census date);
D = deaths to the group occurring from the initial date to the census date;
I = immigration to the group occurring from the initial date to the census date; and
E = emigration from the group occurring from the initial date to the census date.

Given their high quality, Medicare enrollment data are now used to estimate the population over 65 without resorting to this accounting equation.[2]

[2]The above equation is used in "reverse time" to backdate the Medicare-based population estimates to earlier censuses. For example, the population age 55 and over in 1990 can be estimated by "reviving" the Medicare-based estimate for ages 65 and over in 2000 by adding deaths occurring in 1990-2000, subtracting 1990-2000 immigration, and adding estimated 1990-2000 emigration.

Problems with Demographic Analysis

The logic of demographic analysis requires that the population estimate constructed from the basic demographic accounting relationship be comparable with the population measured by the census. As a result of demographic changes in the U.S. population over the past generation, many of the assumptions made by demographic analysis have become more problematic than they were for the censuses of 1940 through 1980. Specifically, immigration and emigration have become much more important sources of population growth. Also, as a result of immigration, intermarriage, and larger societal trends, the current definition and measurement of racial and ethnic groups have become less consistent with historical definitions in the data used to construct the demographic estimates.

While historical data on the numbers of births and deaths are of relatively high quality, data on international migration are more problematic. Estimates of the number of emigrants are subject to considerable variability; in addition, undocumented immigration has become as important numerically as legal immigration, but the available measures are not very exact. Demographic analysis has generally been restricted to national estimates of age, sex, and race groups, since the available measures of subnational migration are not sufficiently reliable to support production of estimates at the state or lower levels of geographic aggregation. Furthermore, because ethnicity has been captured in vital records on a national scale only since the 1980s, demographic analysis has not been used to estimate net undercoverage for Hispanics. Finally, with the introduction of multiple-race responses in the 2000 census, it has become necessary to map the census race categories into historical single-race categories (or vice versa) with the attendant introduction of additional variability into the demographic analysis estimates. This introduction of multiple-race responses is part of the growing complexity of racial classification, which is likely to increase discrepancies between birth certificate reporting and self-reporting of race by adults.

Having pointed out some of the deficiencies of demographic analysis, it is important to emphasize its continuing value in coverage measurement. Demographic analysis places the census results within the well-defined, consistent, and essentially tautological framework of demographic change. The realities of the balancing equation shown above place severe limits on certain results from other studies. Thus, for example, with the passage of one year, all living people get exactly one year older; or, for every boy baby born, there will be approximately one girl baby born. If the results from other coverage measurement studies give results outside the bounds implied by such demographic realities, the departures need to be explained. Some explanations may be demographic—for example, higher levels of immigration or emigration than included in the demographic estimate. But they may also point to statistical or measurement issues—for example, the persistent correlation bias that affects DSE measures of adult black men.

The current demographic analysis program at the Census Bureau also links the measures from the current census with past censuses back to 1940. Each matching study stands alone as a measure of the particular census, but the demographic analysis program grounds its current results in the historical data series, so that it is possible to assess one census relative to others.[3] This linkage can place limits not only on the demographic analysis measures but also on the plausibility of results from other studies.

Due to the relatively deterministic nature of estimates of net coverage error from demographic analysis, estimates of its error or uncertainty are difficult to justify. However, there have been a few attempts to provide error estimates, in particular Robinson, et al. (1993).

It is important to recognize that some results from demographic analysis are more robust than the overall results, so that they may be incorporated into a comprehensive coverage measurement program. While immigration has become an increasingly important component of population change, it has very little impact on the youngest age groups. Thus, it is essential that DSE results for children be consistent with demographic analysis results. Many population ratios, including sex ratios, are much less sensitive to assumptions about problematic components (such as undocumented immigration) than the measured population size. As a result, it may be possible to incorporate demographic analysis results into overall measures of coverage. In 2000, demographic analysis proved to be very useful in the coverage measurement program, notwithstanding the noted deficiencies. Demographic analysis provided an early indication that the initial estimates of the total U.S. population from A.C.E. may have been too high. Demographic analysis may yet provide input to correct for correlation bias in DSE. (See Bell, 1993, for more discussion of this.)

USES OF COVERAGE MEASUREMENT

Coverage measurement serves multiple, not fully complementary purposes. These are, not necessarily in order of importance: (1) assessment of coverage, (2) process improvement, and potentially, (3) adjustment.

Assessment of Coverage Quality

Careful, thorough assessment of the quality of a decennial census is extremely important. Census counts serve a variety of important purposes for the nation, including apportionment, legislative redistricting, fund allocation, governmental planning, and many private uses, such as for business planning. It is important for users of census data to know how accurate the counts are to determine how well they can support various applications. Given that the census could never count every resident exactly once and in the correct location, users need to be able to assess the extent to which the census falls short, the extent to which the accuracy of census coverage differs by location or by

[3]Again, the realities of the balancing equation provide this linkage. Thus, if current research suggests that the demographic analysis coverage measures for a specific age group need to be adjusted (because, for example, the Medicare data show more or fewer enrollees than expected, or the births for a historical time period appear to be too high, or immigration during a decade had to have been higher or lower), the adjustment affects the size of the age group not only in the current census but also in past ones as well.

demographic group, and the extent to which progress has been made in comparison to the previous census.

The total population count of the United States is probably the most visible output of a census, so one obvious measure of coverage accuracy for the census is the error in the count for the entire United States over all demographic groups. However, almost all uses of the census depend on population counts at various levels of geographic and demographic detail (notably racial/ethnic). Uses of the counts, such as for redistricting and local planning, depend on the accuracy of the population counts at detailed levels of geography and for some demographic detail as well. Furthermore, many uses of census data (e.g., apportionment, fund allocation) depend on the counts only in the form of proportional shares of the population. Given this, rates of net undercoverage by various geographic or demographic domains and the impact on population counts, population shares, or both matter a great deal to many users of census data. A key issue has been the differential net undercount of blacks and Hispanics, which has persisted over several decades (see, e.g., Ericksen et al., 1991).

Process Improvement

While it is important to assess census coverage, it would also be extremely helpful to use that assessment to improve the quality of subsequent censuses. Consequently, a valuable use of coverage measurement is to help to identify sources of census coverage errors and to suggest alternative processes to reduce the frequency of those errors. Although drawing a link between census coverage errors and deficient census processes is a challenging task, the Census Bureau thinks that substantial progress can be made in this direction, since its objective going into the 2010 census is to use, to the extent feasible, the 2010 coverage measurement programs to help indicate the sources of common errors in the census counts. This information can then be used to allocate resources toward developing alternative census designs and processes that will provide counts with higher quality in 2020. It is conceivable that use of such a feedback loop could provide sufficient savings in census costs, in addition to improvement in census quality, to more than fund the census coverage measurement program. The panel fully supports this modification of the objectives of coverage measurement in 2010.

Consider, for example, the finding from demographic analysis of the 2000 census that there was a substantial undercount of young children relative to older children. Specifically, the net undercount rates (i.e., $(DSE - C)/DSE$, where DSE indicates the adjusted count, and C indicates the corresponding census count), by demographic group in 2000 based on the revised demographic analysis estimates (March 2001) were as follows (National Research Council, 2004b: Chapter 6):

Demographic Groups	Age Groups	
	0-10	10-27
Black male	3.26	–1.88
Black female	3.60	–1.20
Nonblack male	2.18	–2.01
Nonblack female	2.59	–1.55

One hypothesis is that this undercoverage was at least in part due to the imputation of age for those left off the census form in households exceeding six members (given that the 2000 census forms collected characteristics data for at most six household members). For households that reported more than six members, characteristics data for the additional members either were collected by phone interview (for households that provided a telephone number) or were imputed on the basis of characteristics of other household members and the responses for other households. The hypothesis is that these imputations systematically underrepresented young children since they were underrepresented in the pool of "donor" households.[4]

While demographic analysis can measure the undercoverage of this group, it cannot shed further light on the validity of this hypothesis. However, A.C.E. data are useful in this regard, because characteristics data were collected for most residents counted in the PES, and those data allow an assessment of the extent to which imputations in large households distorted the age distribution. Support of this hypothesis would imply a need to improve either collection or imputation of data (or both) for members of large households in 2010.

While coverage measurement results should be used in support of census improvement whenever possible, coverage measurement will not always uniquely determine a deficient census process. On one hand, if people, ages 18-21 have a duplication rate that is extremely high, one might surmise that it is at least partially due to the inclusion of college students in their parents' households as well as at their household at college. Here the process in need of modification is clear. On the other hand, a housing unit might be placed in the wrong location for many reasons, including an incorrect address in the MAF, a geocoding error using the TIGER geographic database, or an incorrect address entered by the respondent on a Be Counted form.[5] The extent to which coverage measurement programs can specifically discriminate between different sources of census errors depends on the situation.

[4]We note that even if it were determined that increasing this limit from six to seven would reduce the rate of omission of young children in large households, other considerations involving the rate of nonresponse and the quality of the collected information would have to be evaluated before making such a change.
[5]Be Counted forms were provided in public areas in 2000 for people to fill out if they believed that they had been missed in the census.

Specific reasons for errors are more likely to be determined if the Census Bureau saves as much contextual information as possible from the 2010 census. It will need assessments of which individuals and households were enumerated in error, along with various characteristics of persons and households, and of the census processes that gave rise to each enumeration. In Chapter 4 we present some initial ideas on what data might be useful to save, and we plan to provide more specific guidance on what data to save in 2010 for this purpose in our final report. One possibility is to design a comprehensive master trace sample database (see National Research Council, 2004a: Chapter 8).

Census Adjustment

As we have pointed out, the 1999 Supreme Court decision (*Department of Commerce v. United States House of Representatives,* 525 U.S. 316) precluded the use of adjustment based on a sample survey for congressional apportionment, and time constraints strongly argue against the feasibility of using adjusted counts (based on a PES) for redistricting (see National Research Council, 2004a: p. 267). Furthermore, the current approach to adjustment estimation has a number of remaining complications that continue to present a challenge to the production of high-quality estimated counts, including the treatment of movers, matching errors, the treatment of missing data, and the heterogeneity remaining after poststratification of match and correct enumeration status (resulting in correlation bias).

In addition, the use of adjustment is also complicated by the multitude of numbers needed, since one needs adjusted counts at relatively low levels of demographic and geographic aggregation. A decision whether to use adjusted counts for any purpose must therefore rest on an assessment of the relative accuracy of the adjusted counts compared with the census counts at the relevant level of geographic and/or demographic aggregation. The Census Bureau's decision not to adjust the redistricting data, due for release by April 1, 2011, was based on the difficulty of making this assessment within the required time frame.

3

ASSESSMENT OF THE CENSUS BUREAU'S CURRENT RESEARCH PROGRAM FOR COVERAGE EVALUATION IN 2010

The Census Bureau is currently engaged in a number of important research initiatives that they expect will improve their coverage evaluation program in the 2010 census. Part of this research effort has been focused on the design of the coverage evaluation programs for the 2006 census test and for the 2008 dress rehearsal, the program for the dress rehearsal representing the last major opportunity to test plans for coverage evaluation prior to the 2010 census. In particular, the Census Bureau has devoted considerable energies to researching new methods that would be effective in the measurement of components of census coverage error in the 2010 census.

In this chapter, we describe and assess both the census test design in 2006 and the other major activities of the coverage evaluation research program. We introduce this by comparing the plans for the 2010 and 2000 censuses and then describing the limitations of Accuracy and Coverage Evaluation (A.C.E.) in measuring census component coverage error. Following the Census Bureau's terminology, we refer to the 2010 coverage evaluation program as census coverage measurement, or CCM.

HOW THE 2010 CENSUS DIFFERS FROM THE 2000 CENSUS

The 2010 census has an innovative design, resulting in a census that differs from its predecessor as much as any since the incorporation of mailout-mailback data collection in 1970. Furthermore, the design for the 2010 census is dramatically different from the 2000 census in ways that will appreciably affect the 2010 coverage evaluation program. In this section we outline how the 2010 census will differ from the 2000 census and how those changes are likely to affect CCM.

The primary differences between the 2000 and 2010 census designs, as currently planned, are

1. **A short-form only census**. The Census Bureau has now fielded the American Community Survey (ACS), which is a continuous version of the decennial census long form. Therefore, under current plans there will be no long form in the 2010 census. This reduces respondent burden and will facilitate several aspects of data collection in the census, including data capture, data editing and imputation for nonresponse, the work of follow-up enumerators, and the management of foreign language forms and foreign language assistance. As a result, this change is likely to improve data quality.
2. **Use of handheld computing devices for nonresponse follow-up**. The enumerators that follow up nonrespondent households will now use a handheld computing device to (1) administer the census questionnaire (computer-assisted personal interviewing), (2) edit the responses in real time, (3) collect, save, and transmit the data to census processing centers, (4) help locate residences through the use of computer-generated maps (and possibly geographic coordinates), and (5) possibly help organize enumerator routes.
3. **Improved MAF/TIGER system**. The Master Address File (MAF) has been identified as being deficient. For example, see National Research Council (2004b: Finding 4.4). There are currently efforts to improve, for 2010, both the Census Bureau's MAF and its geographic database, the TIGER (Topologically Integrated Geographic Encoding and Referencing) system. The MAF provides a list of household addresses, and TIGER is used to associate each address on the MAF with a physical location. The MAF/TIGER Enhancement Program includes (1) the realignment of every street and boundary in the TIGER database; (2) development of a new MAF/TIGER processing environment and the integration of the two previously separate resources into a common technical platform; (3) expansion of geographic partnership programs with state, local and tribal governments, other federal agencies, the U.S. Postal Service, and the private sector; (4) implementation of a program to use ACS enumerators to generate address updates, primarily in rural areas; and (5) use of periodic evaluation activities to provide quality metrics to guide corrective actions (Hawley, 2004). One motivation for this initiative was the recognition by the Census Bureau that many census errors and inefficiencies in 2000 resulted from errors in the Master Address File and in the information on the physical location of addresses.
4. **Coverage follow-up interview**. The Census Bureau is greatly expanding the percentage of housing units that will be administered a coverage follow-up (CFU) interview in 2010, in comparison to those in 2000 who were administered either the Coverage Edit Follow-up (CEFU) or the Coverage Improvement Follow-up (CIFU) interviews. CEFU was used to determine the correct count and characteristics for people in households with more than six residents (since the census form had space for information for only six persons), and the correct count for households with count discrepancies (e.g., differences between the number of separate people listed on the questionnaire and the indicated total number of residents). CIFU was used to determine whether addresses that were initially judged as being vacant were in fact vacant. The expansion of CFU over CEFU and CIFU was motivated by the recognition, partially provided by A.C.E., that confusion with residence rules made an important contribution to census coverage

error. The CFU interview will be greatly expanded in 2010 to include not only those three situations, but also the following: (a) households with a possible duplicate enumeration identified by a computer match of the census returns to themselves, (b) other addresses at which at least one resident sometimes lived (to avoid enumerations in the wrong location), and (c) other people who sometimes lived in the household (to avoid undercoverage). The latter two situations will be detected through the addition of two "coverage probe" questions to the census form. However, due to resource and time constraints, the Census Bureau may be able to administer the CFU to only a subset of the qualifying households in 2010. The Census Bureau thinks that it may be able to follow up only 5 to 10 percent of the nation's addresses for this purpose, but some preliminary estimates suggest that a larger percentage may satisfy one or more of these contingencies.[1] In that case, the Census Bureau may have to prioritize by selecting a subset of the qualifying households that were more likely to provide information that would result in a less erroneous count. Implementation of this operation will depend on information collected in the 2006 test census and the 2008 dress rehearsal.

5. **Removal of duplicate enumerations in real time**. As implied in (4) above, the CFU interview will be used to follow up suspected duplicate enumerations that are identified through use of a national computer search for duplicate enumerations, with the objective of determining which address of a pair of duplicates is the correct residence and consequently removing the erroneous duplicate enumeration from the census.

This new census design has some benefits for the coverage measurement program in 2010. Focusing on the collection of short-form data and the use of handheld computing devices might improve the quality of the information collected, thereby improving the quality of the matching of the postenumeration survey (PES) to the census. Having an improved and more complete MAF should reduce the extent of whole-household undercoverage. Finally, the national search for and field verification of duplicate enumerations should reduce the number of duplicates in the census, which may facilitate the estimation of component errors in the census and may also simplify the application of the net coverage error models used in dual-systems estimation (DSE). So the changes to the 2010 census design are also likely to improve the quality of the coverage measurement information provided in 2010.

It is important to emphasize that some of the changes to the 2010 census design were motivated by the results of the 2000 A.C.E. program. Specifically, the large number of erroneous enumerations, especially duplicates, motivated the expansion of the CFU interview, as well as the implementation of the national search for duplicates. Also, although not directly a finding from A.C.E., the recognition that the 2000 census Master Address File had a large number of duplicates and was otherwise of uncertain quality motivated some of the improvements of the MAF/TIGER system.

[1]We think that reasonable estimates may already be possible given data from 2000 and the later census tests. For example, the 2004 census test indicates that categories (b) and (c) may sum to 11 percent or so.

LIMITATIONS OF A.C.E. IN MEASURING COMPONENT COVERAGE ERROR

A.C.E. in the 2000 census was planned from the outset as a method for adjusting census counts for *net* coverage error. Hence, A.C.E. focused on estimating net census coverage error rather than summaries of census component errors. For example, the limited geographic search for matches used in A.C.E. relied on the balancing of some erroneous enumerations and omissions that were actually valid E-sample enumerations but in the wrong location. Such errors could result, for example, from a geocoding error (placing an address in the wrong census geography) or enumeration of someone at a second home. Because such erroneous enumerations and omissions were expected to balance each other, on average, they were expected to have little impact on the measurement of net coverage error. Therefore A.C.E. did not allocate the additional resources that would have been required to distinguish these situations from entirely erroneous enumerations or omissions. Similarly, A.C.E. did not always distinguish between an erroneous enumeration and counting a duplicate enumeration at the wrong location.

The following are limitations of A.C.E. in 2000 for measuring census component coverage error:

- Inadequate information collected as part of the census and the PES allowed too many mistakes in the A.C.E. final determination of Census Day residence. In 2000, comprehensive information was not collected from a household either in the census or in the A.C.E. interview regarding other residences that residents of a household often used or on other individuals who occasionally stayed at the household in question. This limited the Census Bureau's ability to correctly assess residency status for many individuals. The Census Bureau intends to include more probes to assess residence status in the 2010 census questionnaire, in the census follow-up interview, and on the 2010 CCM questionnaires. Also, in 2010, the duplicate search will be done nationwide, not only for the PES population. In addition, the Census Bureau plans on incorporating a real-time field verification of duplicate enumerations in 2010. (For details on issues in determining correct residence, see U.S. Census Bureau, 2003.)
- Also, nonresponse in the E- and P-samples complicated matching of the P-sample to the E-sample (for coverage measurement) and of the E-sample to the census (to identify duplicates). It also complicated estimation because it interfered with assigning a person to the correct poststratum (under the 2000 design) or creates missing values for predictor variables (as discussed below, under the proposed use of logistic regression in 2010). (For details, see Mulry, 2002.)
- Furthermore, the missing data treatments used for individuals with extensive nonresponse failed to fully utilize the available data. Procedures are now being examined that make greater use of the available data, especially on household composition, to determine the match status of these individuals in 2010.

- Also, the methodology used for individuals who moved between Census Day and the day of the postenumeration interview (known as PES-C) resulted in a large percentage of proxy enumerations, which in turn resulted in matching error.[2] (PES-C was implemented in 2000 due to the early plans, later cancelled, to use sampling for nonresponse follow-up in the 2000 census.) The Census Bureau will probably return in 2010 to the use of PES-B (similar to the 1990 methodology), which relies completely on information from the inmover.
- The A.C.E. Revision II estimates modified undercoverage estimates for adult black men using sex ratios from demographic analysis (ratios of the number of women to the number of men for a demographic group) to correct for correlation bias (for details, see Bell, 2001; Haines, 2002). This method assumes that the estimated adult sex ratios from demographic analysis are more accurate and precise than those from the A.C.E. For nonblack Hispanics, estimation of adult sex ratios requires a long historical series of Hispanic births and deaths and, more importantly, highly accurate data on the magnitude and sex composition of immigration (both legal and undocumented). The historical birth and death data for Hispanics are available only since the 1980s, and the available measures of immigration are too imprecise for this application. Consequently, this use of demographic analysis to modify A.C.E. estimates was not directly applicable to nonblack Hispanic males in 2000.[3]
- The approach taken to estimate net census coverage error relied on balancing erroneous enumerations against omissions in cases in which there was insufficient information for matching E- and P-sample cases. Consequently, A.C.E. was not effective at estimating components of census error.
- Poststratification is used to reduce correlation bias (see description in Chapter 2), since it partitions the U.S. population into relatively homogeneous groups. The number of factors that could be included in the poststratification used in A.C.E. was limited because the approach fully cross-classified many of the defining factors, with the result that each additional factor greatly reduced the sample size per poststratum. (For details of the 2000 poststrata, see U.S. Census Bureau, 2003.) The 2010 plan uses logistic regression modeling to reflect the influence of many factors on coverage rates without having to define a large number of poststrata.
- Also, small-area variations in census coverage error that are not corrected by application of the poststratum adjustment factors to produce estimates for subnational domains (referred to as synthetic estimation) were not reflected in the variance estimates of adjusted census counts. The Census Bureau is examining the use of random effects in their adjustment models to account for the residual

---

[2]PES-C collected information about whether a PES outmover household matched to the census through use of information about the outmover household (often using proxy information), but resulting matches were applied to the size of the inmover household rather than the size of the outmover household because the information on the number of inmovers was considered to be of greater reliability.

[3]In support of this argument, it is useful to note that a majority of working-age (18-64) Hispanics are foreign-born—about 55 percent, whereas only less than 5 percent of whites and slightly more than 5 percent of blacks are.

variation in small-area coverage rates beyond what is modeled through synthetic estimation.

In addition to these issues, other features of A.C.E., including aspects of data collection and sample design, made the 2000 A.C.E. less informative than it might have been in measuring census component coverage errors. As stated above, this was only to be expected given the focus of A.C.E. on producing adjusted census counts, well justified by the desire to remedy long-standing patterns of differential undercoverage of minorities in the census. However, with the new priority of measuring census component coverage error, a number of design and data collection decisions, within the general framework of PES data collection, remain open to modification. Furthermore, as we argue below, estimation of net census error also remains important for assessment of census component coverage error, specifically census omissions.

PLANS FOR COVERAGE EVALUATION IN THE 2006 CENSUS TEST

The goals for the 2006 test census relevant to coverage evaluation were as follows (U.S. Census Bureau, 2004):

- To examine how the Census Bureau can improve the determination of Census Day residence in the CCM process through modification of the census questionnaire, the initial PES questionnaire, and the PES follow-up interview. This may be the most important problem facing coverage evaluation and the greatest opportunity for improvement, because the A.C.E. underestimated erroneous enumerations by 4.7 million people in 2000, and overestimated the P-sample population by 2.2 million, much of which was probably due to errors in enumerating people in their proper census location (see National Research Council, 2004b: 218 and 253, for details). A request for information on alternative addresses or additional part-time residents was not included in the 2000 census, which limited attempts to ascertain correct Census Day residence.
- To test procedures for determining more accurately the location of a person's Census Day residence outside the P-sample blocks for P-sample inmovers and for people with multiple residences.
- To determine how the more extensive matching for duplicates and people with multiple residences (following up on information collected in the CFU interview) can be implemented with the anticipated resource and time constraints in 2010.
- To identify additional data to be collected on census processes in support of the measurement and analysis of census component coverage error.
- To measure possible contamination of the CFU interview by the (possibly simultaneous) collection of coverage measurement information and to assess the implications for CCM data collection and estimation.

The coverage evaluation program of the 2006 test census began with a PES, after the census data collection was complete, in which computer-assisted personal interviews were administered to an independent sample of approximately 5,000 housing units (drawn from the same address list as the census) to determine Census Day residence. This was followed by an automated and then clerical match to the census enumerations, with field follow-up of those with unresolved match status. A person follow-up interview was conducted simultaneously with the PES to collect additional data to resolve residence status for various situations.

Once matching was completed, the CCM program used DSE based on the usual E- and P-samples, except that the addresses for the P-sample were identical to those of the test census. This exception prevented the measurement of whole-household omissions in the test census. Movers between the time of the census and the collection of postenumeration survey (CCM) data utilized PES-B methodology, which counts the number of people resident in the CCM blocks at the time of the postenumeration survey rather than the number of people resident on Census Day. Information on the other locations at which a person might also have been counted was collected in a follow-up interview for households that indicated other residences on the census questionnaire. This was to assist in the assessment of correct residence and to better define omissions, erroneous enumerations, and duplications.

The CCM person interviewing used a laptop for the initial interview, and unresolved matches were followed up with personal visits. For census returns that provided a phone number, the CCM interviews were carried out by telephone, as in 2000. CCM personal visits did not begin until nonresponse follow-up was concluded. However, CCM interviewing was simultaneous with the CFU follow-up interview. There was an automated and computer-assisted clerical search for P-sample matches and duplicate census enumerations at the Census Day residence location, as well as at other locations where the person may have been counted. There was also an automated search across all census enumerations in the test site both for P-sample matches and for duplicate census enumerations. There was an attempted match to census enumerations that had a missing or deficient name or were otherwise difficult to match due to limited information to better estimate components of census coverage error (see discussion of KEs below.) No weighting or imputation was carried out for missing data, and coverage estimates were not produced. Finally, the Census Bureau will explore various estimation methodologies to generate estimates of components of census coverage error and net coverage error, conditional on the limitations of the census test, to examine whether sufficient and consistent data are being collected.

Unlike the case for decennial coverage measurement programs, no attempt was made to collect data to assess whole-household undercoverage. Also, no attempt was made to assess the undercoverage of individuals living in group quarters. (CCM is also planned to exclude group quarters, about 2.7 percent of the U.S. population, from coverage measurement in 2010. This is unfortunate, given the difficulty in counting the institutional population.) Data needed to estimate coverage error (both net coverage error and components of coverage error) for persons living in housing units will be assembled by census operation to support the linkage of census component coverage error with specific census operations.

MAJOR ACTIVITIES OF THE CENSUS COVERAGE MEASUREMENT RESEARCH PROGRAM

The CCM research program involves several activities grouped into the following categories: (1) research on measuring components of census error, which includes development of a framework for coverage measurement, matching of cases with minimal information, and identification of census duplicates in real time; (2) research on models for net error, including alternatives to poststratification and synthetic estimation; and (3) research on contamination due to the CFU interview. We also examined preliminary ideas of the Census Bureau regarding the design of the CCM postenumeration survey and the current application of E-StARS to coverage measurement; E-StARS is the Census Bureau research program examining possible applications of merged, unduplicated lists of administrative records.

All of these research efforts support the objective in 2010 of measuring census component coverage errors. Matching cases with minimal information reduces the need to rely on imputation of match status and therefore more clearly determines whether those cases are errors and, if so, what type. The identification of duplicates clearly facilitates their estimation and reduces the estimated number of erroneous enumerations. Improved estimation of net error improves the estimation of the number of omissions. Finally, contamination of the CFU by the CCM interview could result in an unrepresentative census in the P-sample block groups and therefore bias the estimates produced by DSE. We now describe and comment on each of these areas of research in turn.

RESEARCH ON MEASURING COMPONENTS OF CENSUS ERROR

The Census Bureau's Framework Paper

In considering the measurement of erroneous enumerations, omissions, duplications, and enumerations in the wrong place, it became apparent that the definitions of these coverage errors needed clarification (see National Research Council, 2004b: 252). The Census Bureau therefore decided to develop a framework of precise definitions of census errors, as well as what assumptions supported their estimation, to better guide development of its coverage measurement plan for 2010. The resulting draft document "Framework for Coverage Error Components" (U.S. Census Bureau, 2005) is an excellent attempt to provide this foundation.

This document defines erroneous enumerations as (1) duplicate enumerations, (2) people born after Census Day, (3) people who died before Census Day, and (4) people who are not residents of a housing unit in the United States. Omissions are people who should have been enumerated in the census but were not. Contrary to this, in A.C.E., which focused on net error, persons had to be enumerated in a housing unit within the search area of the residence (generally the relevant E-sample block cluster) to be considered correctly enumerated. In this new framework, the starting position is that persons must only be enumerated in a housing unit somewhere in the United States to be considered to be a correct enumeration. This definition of a correct enumeration used in the framework document is not Census Bureau policy; it is instead a useful starting point in developing a comprehensive and clear understanding of the measurement of census coverage error, with the expectation that the geographic dimension will be addressed in later expansions of the framework.

The varying amount of information available for census enumerations complicates the classification of census errors. Data-defined enumerations are those with at least two recorded characteristics; others are non-data-defined enumerations. Among the former, some enumerations have sufficient information for matching and follow-up (complete name and two additional characteristics), and others have insufficient information. The non-data-defined and insufficient information cases could be either correct or erroneous enumerations, since the data are often insufficient to make any further determination.

Finally, information to determine enumeration status is collected from the PES. The Census Bureau refers to the list of people that would be enumerated if the P-sample were applied nationally as the notional P-census. Thus, conceptually every potential enumeration falls into one of four cells: (1) those in both the P-census and the census, (2) those in the P-census but not in the census (census omissions), (3) those in the census but not in the P-census (erroneous enumerations and P-census omissions), and (4) those missed by both the P-census and the census.

Potential E-sample cases include correct enumerations and erroneous enumerations but not non-data-defined people or census omissions. The A.C.E. definition of E-sample erroneous enumerations also includes (a) correct enumerations in the wrong location and (b) enumerations with insufficient information for matching. Measurement of census component coverage errors requires separate estimates of the number of enumerations that are in the wrong location and the number of enumerations with insufficient information that are actually erroneous.

To assess the number of omissions, A.C.E. used the P-sample nonmatches, which under the new definitions could be omissions, people enumerated in the wrong location, or nondata-defined people. The challenge here in moving toward a focus on error components is to determine how many of those people were actually missed in the census.

To provide high-quality estimates of census component coverage errors in 2010, the Census Bureau needs to make progress on two fronts. First, it must reduce the inflated estimate of erroneous enumerations. Enumerations with insufficient information need to be examined further, enumerations in the wrong place need to be identified as such, and the remaining unresolved cases need to be treated as nonrandomly missing data. Second, a better method is needed to estimate the number of people missed by both the P-sample and the census. The current approach assumes independence of correct enumeration status and match status within poststrata, and failure of that assumption results in correlation bias.

Since net error is defined as omissions minus erroneous enumerations, one can estimate omissions by summing reliable estimates of net error and the number of erroneous enumerations. Since net error can be estimated by DSE minus the census, omissions may be estimated by taking the dual-systems estimate minus the census plus the number of erroneous enumerations. However, this estimation strategy needs to be improved through additional data collection to help distinguish enumerations in the wrong location and to better handle cases with insufficient information, as well as through better estimation of the number of people missed by both the PES and the census.

The framework document also addresses how to estimate these various error components and what assumptions they are based on. Additional information will be collected in 2010 regarding other residences at which someone might have been counted to determine more accurately whether a nonmatched P-sample enumeration is actually an omission and which of a set of duplicates is the correct enumeration. Furthermore, there will be greater efforts made to match cases with "insufficient" information. Finally, missing data models will be developed to treat cases that are not data-defined.

The panel supports the general approach described in this draft framework, which is consistent with recommendations in National Research Council (2004b). This is an important first step toward developing a feedback loop linking the measurement of census component coverage error to deficiencies in specific components of census processes.

The panel has some concerns about the proposed treatment of imputations in the draft framework. A focus on the correctness of an imputation as an enumeration is misplaced, as are concerns about the correctness of imputations of characteristics. Imputations are simply the means to an end, which is improved census estimation, and it is the quality of the estimates collectively that should be assessed. For example, if a characteristic of a known person is imputed, the question of whether that is the person's correct value is of no interest. The critical question concerns whether census estimates that involve that characteristic are collectively improved by the imputation, which will tend to be the case if the imputation model is sensible. The same principle applies to whole-person imputations. (This approach is compatible with a focus on components of error, since the measures used are for aggregates rather than individuals.)

Finally, different errors may be important for different uses of the census numbers, so the framework should be sufficiently flexible to allow for aggregating component errors in more than one way. For example, for estimation of broad demographic distributions (to predict future Medicare enrollment), an error in age might be important, but misplacing a person geographically would be of little consequence. Conversely, for redistricting purposes, a person's exact age is unimportant but geographical accuracy is critical. The panel hopes to examine this more in our final report.

Matching Cases with Minimal Information

In the 2000 census, for an enumeration to have sufficient information for matching and follow-up, it needed to include that person's complete name and two other nonimputed characteristics. In A.C.E. in 2000, there were 4.8 million (sample survey weighted) data-defined enumerations with insufficient information for matching and follow-up, meaning that they contained two characteristics. These cases were coded as "KE" cases in A.C.E. processing, and we retain that terminology. A.C.E. estimation treated KEs as erroneous enumerations, and they were removed from the census enumerations prior to dual-systems computations. (If KEs are similar in all important respects to census enumerations with sufficient information for matching, removal from dual-systems computations increases the variance of the resulting estimates, but it does not greatly affect the estimates themselves.) Removal of KEs helped to avoid counting a person twice because matches for these cases are difficult to ascertain. Also, it was difficult to follow up these E-sample cases to determine their match status if they initially were not matched to the P-sample, because of the lack of information with which to identify the person to interview.

However, some unknown and possibly large fraction of these cases were correct enumerations. Therefore, removing these cases from the matching inflated the estimate of erroneous enumerations, and it also inflated the estimate of the number of census omissions by about the same amount, since roughly the same number that are correct enumerations would have matched to P-sample enumerations. Given that the emphasis in 2000 was on the estimation of net census error, this inflation of the estimates of the rates of erroneous enumeration and omission was of only minor concern. However, with the new focus in 2010 on estimates of components of census error, there is a greater need to find alternative methods for treating KE enumerations. One possibility that the Census Bureau is currently exploring is whether many of these cases can be matched to the P-sample data using information on other household members.

To examine this, the Census Bureau carried out an analysis using 2000 census data on 13,360 unweighted data-defined census records that were found to have insufficient information for matching, to determine whether some of them could be reliably matched. (For details, see Auer, 2004, and Shoemaker, 2005.) This clerical operation used name, date of birth, household composition, address, and other characteristics to match these cases to the P-sample. For the 2000 A.C.E. data, 44 percent of the KE cases examined were determined to match to a person who lived at the same address on Census Day and was not otherwise counted, with either "high confidence" or "medium confidence" (which are reasonable and objectively defined categories of credibility). For the 2000 census, this would have reclassified more than 2 million census enumerations from erroneous to correct enumerations, as well as a like number from P-sample omissions to matches, thereby greatly reducing the estimated number of census component coverage errors. For the remaining unresolved cases, the current view on the part of the Census Bureau is to treat them as missing data in the estimation of rates of component census error. However, this issue has yet to be studied.

The treatment of KEs can be viewed as another component of "error" in the same way that a person incorrectly geocoded is an error—that is, that it is a problem for processing but not a part of what we would call an omission or an erroneous enumeration. The use of the term "erroneous enumeration" for these cases in the past is inappropriate. Cases with insufficient information should be treated as having unknown or uncertain enumeration or match status. The term "erroneous" should be reserved for incorrect enumerations. The terminology used therefore needs to distinguish between types of error and the uncertainty associated with these types of error for particular cases.

The panel strongly supports this research. In considering further development of the idea, it would be useful to try to find out more about any characteristics associated with KEs in order to find out how to reduce their occurrence in the first place. (E-StARS might be useful for this purpose.) Furthermore, the clerical operation used to determine the status of KEs was resource intensive, and it would be useful to try to automate some of the matching to reduce the size of this clerical operation in 2010. Ultimately, the Census Bureau should rethink the definition of cases with insufficient information for matching more generally.

DISCOVERY OF CENSUS DUPLICATES AND P-SAMPLE MATCHES TO CENSUS RECORDS OUTSIDE THE SEARCH AREA

Duplication in the census can result from a number of different circumstances. Some possibilities include housing unit duplication in the Master Address File, counting college students both at home and away at school, counting children in joint custody at both parents' homes, counting movers both at their current home and at their previous home, counting people with vacation homes both at their usual home and at their vacation home, counting friends and relatives at a home at which they are staying temporarily, counting people at both residences who have one residence to commute to and from work and a separate residence on weekends, and counting people in nursing homes and prisons and at a residence of their immediate family members.

The Census Bureau implemented a computer and clerical operation to identify and remove duplicate housing units during the middle of the 2000 census due to an indication that a large number of duplicate housing units were included in the census (see National Research Council, 2004b, for details). In this operation, potential duplicate housing units were identified through the use of both person and housing unit matching. Criteria were developed that attempted to distinguish between actual housing unit duplications and form misdeliveries (typical of multiunit structures), which were retained in the census, since there is often no error as a result.

The Census Bureau subsequently carried out research to determine how many duplicate persons remained in the census (see, e.g., Mule, 2001, and Fenstermaker, 2002.) This was done by computer matching the E-sample to the entire nation's enumerations, wherein a "match" was determined by agreement of birth date and year and first and last names. (This could not have been accomplished in any previous census, since name and date of birth had not been captured electronically prior to the 2000 census.) In addition to the computer search, Fay (2002, 2003, 2004) developed a series of increasingly refined statistical models that estimated the percentage of matches discovered in this way that were coincidental and therefore not real duplicates. This research indicated that there were 5.8 million duplicates in the census after the duplicate housing search, which was compared with the 1.9 million duplicates found by A.C.E. (National Research Council, 2004b: Chapter 6). This research also provided characteristics of people who were more likely to be duplicated, including minority children, college students, minority young adults, people duplicated between housing units and group quarters (especially correctional institutions and nursing homes), and minority men ages 25 to 64.

Given the success of this research effort, the Census Bureau is now planning to implement the identification of duplicates and the corresponding correct enumerations nationally and in real time in 2010. Similarly, P-sample enumerations will be matched to census records outside the P-sample search area (the surrounding block clusters).

Implementing these operations on the scale needed and under severe time constraints will present a number of challenges. It is sometimes difficult to distinguish between duplication and form misdelivery. Also, in some cases, it can be difficult to determine a person's correct residence, for example, for children in a shared custody situation. Furthermore, the goal is to resolve the entire household roster rather than just determining which enumeration is correct. The Census Bureau needs to estimate the resources that will be needed to support this effort. Estimating various timing and resource issues through a census test will be difficult, since census tests involve field work for only a few counties, and there is typically no field validation of cases outside the test census area.

RESEARCH ON MODELS FOR NET COVERAGE ERROR

Even with a primary goal of estimating census component coverage error, there are still two important reasons to continue research on net coverage error models. First, as mentioned previously, models for net coverage error are needed to estimate the number of census omissions. Second, census data users find information on net coverage error useful. Reliable estimates of net coverage error indicate the degree to which demographic groups are differentially undercovered in the census, and they indicate the degree to which geographic jurisdictions are differentially undercovered. This helps users decide whether census counts should be used for various purposes.

A.C.E. Research Database

The Census Bureau's research on net coverage error has been greatly facilitated by the development of an A.C.E. research database. Briefly stated, this database contains the data collected through A.C.E. to support estimation of net coverage error in 2000, and it is also weighted to represent the additional information collected from the national duplicates search and the evaluation follow-up survey, so that the net coverage error estimates produced are nearly identical to those from A.C.E. Revision II.

The panel has made modest use of this database, and it commends the Census Bureau for supporting its development. It would be beneficial if this database could be made available to researchers outside the Census Bureau after addressing confidentiality issues. One possibility would be to make a confidentiality-protected version of this database available at the Census Research Data Centers.

Logistic Regression for Estimating Net Coverage Error

The Census Bureau is examining the use of logistic regression modeling to estimate net census error, replacing the use of poststrata and synthetic estimation. The motivation is to be able to utilize many more predictors, including continuous predictors, in fitting the probability of match and correct enumeration status.

Poststratification is mentioned in the earliest literature advocating DSE (Sekar and Deming, 1949), and it has been used in the census since the 1980 postenumeration program (PEP) to reduce correlation bias. This is accomplished by estimating the adjusted counts $(C - II)\left(\frac{CE}{E}\right)\left(\frac{P}{M}\right)$ separately within poststrata. (Recall that C stands for the number of census enumerations, II represents the number of people lacking sufficient information for matching, CE represents the number of E-sample persons correctly enumerated in the census, E represents the number of E-sample enumerations, P stands for the estimate of the number of all valid P-sample persons, and M stands for the estimate of the number of P-sample persons who match with an E-sample person. Note that here, CE is defined consistent with the definition of a correct enumeration in A.C.E., that is, an enumeration that is located within the search area and is therefore *not* the definition of a correct enumeration found in the framework document.)

A perfect poststratification would partition the true population and the E-sample population so that the underlying enumeration propensities for individuals within a poststratum are identical. However, this is unattainable and therefore the practical goal is to partition the sample cases so that individuals are more alike within a poststratum than individuals are from different poststrata. If this is accomplished, correlation bias should be reduced. Poststratification also supports the use of synthetic estimation, which is used to carry down adjustments to very low geographic levels. Synthetic estimation makes use of coverage correction factors, $\left(\frac{C - II}{C}\right)\left(\frac{CE}{E}\right)\left(\frac{P}{M}\right)$. These factors are applied to any subpopulation within the poststratum by multiplying this factor by the relevant subpopulation's census count to produce the adjusted count for that subpopulation. Estimates of the uncertainty about synthetic estimates for small areas combine estimates of the variance from the estimation of the undercount rates in each poststratum and residual variation due to heterogeneity of small areas within the same poststratum. The first component can be estimated by standard methods; approaches to the second are more difficult and are discussed in a later section.

While poststratification has the advantages of reducing correlation bias and supporting synthetic estimation, a major disadvantage is that it allows only a relatively small number of factors to be included in the poststratification scheme (and in the resulting synthetic estimation). This is because the Census Bureau includes the full cross-classification of most of the factors to define the poststratification, and, as a result, the poststrata quickly become very sparsely populated, despite the large sample size of the PES. Use of many poststrata thus improves homogeneity within poststrata at the price of estimates with high sampling variances. Furthermore, because the formation of independent poststratum estimates does not recognize that poststrata with similar characteristics are likely to have similar rates for matching and correct enumeration, separately treating those poststrata fails to pool data when it would be beneficial to do so.

An alternative to poststratification is logistic regression of the binary match/nonmatch and correct enumeration/not correct enumeration variables on the available predictive factors. This approach allows the inclusion of more factors in the model, since it does not require factors to be treated as categorical, and it allows high-order interactions to be included or omitted as desired. Poststratification is the special case of logistic regression in which the predictors are indicator variables for membership in the categories defining the poststrata and all interactions are included in the model (see Box 3-1). In theory, small-area estimates under logistic regression could improve on those provided through synthetic estimation by using more predictors to predict the probabilities of match and correct enumeration status, and hence reducing correlation bias.

Logistic regression was suggested for use in the general area of DSE by Huggins (1989) and Alho (1990) and specifically applied to census undercoverage in Alho et al. (1993). However, these studies did not consider unresolved cases and made use of the data only from the P-sample blocks, rather than the full census. Haberman et al. (1998) introduced some additional features that addressed the above limitations. They proposed two separate logistic regressions to model match status (using P-sample data) and correct enumeration status (using the E-sample data). To represent cases with unresolved match status (with a completely analogous discussion of correct enumeration status), two records are constructed, one "matched" and the other "unmatched," and weights are used to represent the "probability" that a given record matched to the census, given the available characteristics. (Match and correct enumeration probabilities for unresolved cases could be provided by a computer matcher developed by the Census Bureau.) Survey weights are also attached to all the records to reflect the complex sample design.

BOX 3-1

Logistic Regression as a Generalization of Poststratification

To see that logistic regression is a generalization of poststratification, consider the following generic logistic regression model (used for either modeling percentage matched or percentage correct enumeration): $log\left(\frac{p_i}{1-p_i}\right)=\sum_{k=1}^{K}\beta_k X_{ki}$, where p_i is the probability that individual *i* in the PES matches a census enumeration or is a correct enumeration, X_{ki} is the value for individual *i* of the k^{th} explanatory variable, and β_k is the associated regression coefficient for X_{ki}. If we assume that each of the X_{ki}'s is *a* variable that equals 1 when the i^{th} individual is in the k^{th} poststratum, and 0 otherwise, then $\hat{\beta}_k$ is chosen so that the observed matching rate in the k^{th} poststratum is equal to $\frac{\exp(\hat{\beta}_k)}{1+\exp(\hat{\beta}_k)}$.

The Census Bureau's approach is a generalization of this, whereby it uses several categorical variables at various levels (e.g., sex and age), and combinations of levels of the categorical variables play the role of the above variables for the individual poststrata. A simple example would be where $\hat{\beta}_i, \hat{\beta}_j, \hat{\beta}_{ij}$ are selected so that the observed matching percent in the i,j^{th} poststratum (relative to the i^{th} and j^{th} levels of the two classification variables), was equal to: $\frac{\exp(\hat{\beta}_i+\hat{\beta}_j+\hat{\beta}_{ij})}{1+\exp(\hat{\beta}_i+\hat{\beta}_j+\hat{\beta}_{ij})}$, where $\hat{\beta}_{ij}$ indicates the regression coefficient for the relevant interaction term.

This approach is the Census Bureau's leading candidate for net coverage error modeling in 2010.

Haberman et al. (1998) fit two separate logistic regression models. The first one, using P-sample data, models the probability of matching a P-sample case to the E-sample, and the second one, using E-sample data, models the probability that a census enumeration is correct. Relating this to DSE, in the expression $(C-II)\left(\frac{CE}{E}\right)\left(\frac{P}{M}\right)$, the probability that a census enumeration is correct is the second factor, and the probability of a match is the inverse of the third factor. Therefore, the two logistic regression models estimate two of three main factors in DSE, the remaining factor representing the number of matchable enumerations in the census. The regression coefficients of these two logistic regression models are fit by maximizing the weighted log likelihood, which is a measure of the goodness of fit of the logistic regression model to the data.

Using logistic regression, synthetic estimation can now be replaced by the following methodology. Letting $\hat{p}_{cei}$ represent the estimate from logistic regression of the correct enumeration probability for person *i*, and letting $\hat{p}_{mi}$ represent the estimate from logistic regression of the match probability for person *i*, the estimated number of people in a small area is the sum of the ratio $\frac{\hat{p}_{cei}}{\hat{p}_{mi}}$ over the individuals *i* in that area (ignoring the treatment of cases with insufficient information for matching).[4] A grouped jackknife procedure is used to obtain the standard errors of the small-area estimates. If the explanatory variables are limited to those collected in the census, not characteristics or process variables from A.C.E. or CCM, small-area estimates can be computed directly using the method just described. However, this approach sacrifices the additional predictive power of covariates collected for cases in the P-sample. Techniques suggested by Eli Marks may be used to accommodate use of P-sample variables at the subnational level. For details, see Marks et al. (1974).[5]

A few complicating issues are raised when using these methods. One is how to incorporate the survey weights in the model-building and model-fitting processes. The CCM PES sampling weights need to be incorporated not only in the estimation of the logistic regression coefficients, but also in the decision as to which predictors to include in the logistic regression models and which model form to use, as well as in estimating the variance of the resulting estimates. The question of how to treat the complex sample design in these types of models has a substantial research literature. The approach taken by the Census Bureau is to weight the fractional cases by the inverse of the sampling weights. An alternative approach, which may produce more efficient estimates, is to

[4]In this discussion, we are ignoring missing data in covariates, which introduce some complexities into the above development.

[5]The Census Bureau has examined competing estimators that all have empirical deficiencies in comparison to the above estimate. As mentioned, the estimate for the population of a domain is $\sum_i \frac{\hat{p}_{cei}}{\hat{p}_{mi}}$ for all individuals *i* in a domain. A competing estimator that the Census Bureau has mentioned is $\sum_j w_j \frac{\hat{p}_{cej}}{\hat{p}_{mj}}$, which is now a sum over the individuals *j* in the PES blocks and in the relevant domain. Another competing estimator replaces the correct enumeration probability $\hat{p}_{cej}$ in these two alternatives by an indicator function for those individuals in the domain that had correct enumeration status, reducing the modeling to only the logistic regression model of match status. The problem with these two alternatives is that they are too sensitive to sampling variation. The Census Bureau has also considered variants of these two alternatives by reweighting the data elements so that the data-defined persons from the E-sample are ratio-adjusted to the census counts within poststrata. A further problem with some of these approaches is that small-area estimates do not necessarily sum to the estimates for larger areas.

include the variables that make up the sampling weights as predictors in the model (e.g., Little, 2003). Comparisons of these two approaches would be of interest.[6]

A second complication is the treatment of missing data. Specifically, it is not clear how to treat cases with insufficient information for matching in the estimation of the logistic regression coefficients. Regarding the small-area estimation that results from the use of the logistic regression model, it is also not clear how to treat non-data-defined cases. We hope to provide more guidance on these issues in our final report.

Over the past two years, the Census Bureau has examined the use of logistic regression models to estimate net census error, focusing attention to date on the performance of six sets of explanatory variables for both the P-sample matches and the E-sample correct enumerations. These sets all use explanatory variables that are indicator variables of various combinations of the levels of the following six factors: race/origin (7 groups), age/sex (7 groups), tenure (owner, nonowner), metropolitan statistical area/type of enumeration area (MSA/TEA) (4 groups), region (4 groups), and mail return rate (high or low, with boundaries dependent on race/origin domain). The six sets are

1. The 416 indicator variables for the poststrata used in the March 2001 poststratification.
2. The 150 main effects and first-order interactions of the variables used to define the March 2001 poststratification.
3. The 23 main effects of the variables used to define the March 2001 poststratification.
4. The 98 main effects and all interactions from the variables for just three of the six factors from the March 2001 poststratification, that is, race/origin, age/sex, and tenure (omitting MSA/TEA, region, and mail return rate). The acronym ROAST is used to distinguish this reduced set of factors from the full set used in the 2001 poststratification.
5. The 62 main effects and first order interactions from ROAST.
6. The 14 main effects from ROAST.[7]

These six models were fit to data from the A.C.E. research database (for further details, see Griffin, 2005a).

[6]Another complication that we will not discuss further here is that the adjustments made on the A.C.E. research file have resulted in the dependent variables occasionally lying outside of the interval (0,1).

[7]The number of interactions does not correspond to the situation of fully-crossed effects since the poststratification used in 2000 did not fully cross the six variables. For example, the poststrata of American Indians or Alaskan Natives living on a reservation is only crossed by age/sex, and tenure, but not by MSA/TEA, region, or mail return rate, and this extends to the ROAST models.

Model Comparisons

The Census Bureau compared the five logistic regression models (models 2-6) with the 2001 poststratification (model 1) for *predicting* A.C.E. data. Equivalence would seem sufficient in this comparison, since logistic regression can make use of many variables in addition to those used to define the poststrata.

When comparing nested models, that is, models that are identical except that some of the parameters in one of the models have been constrained to be constant (usually zero), the distinction between fitting and prediction is typically represented by adding a penalty for the number of additional parameters to a goodness-of-fit measure. As in linear regression, the additional parameters guarantee that the model with the larger set of predictors will fit the data at least as well, if not better, than the more parsimonious model, but this advantage may be offset by the increased variances of the fitted values, due to estimating more parameters. The combination of the goodness-of-fit statistic and the penalty for additional parameters reflects this trade-off. Haberman et al. (1998) suggested using a logarithmic penalty function, with a jackknife bias estimate to adjust for the use of unnecessary predictors (overfitting). The Census Bureau studied the use of goodness-of-fit tests based on the Satterthwaite approximation. Measures such as Mallows' C_p , and the information criteria AIC, and BIC provide useful penalties for comparing regression models in a predictive situation.

The theory for comparing nonnested models is less straightforward, but such comparisons will also be needed. One such nonnested alternative separates the modeling of undercoverage into two models, one for the probability that an entire household will be missed, and another for the probability that an individual in a partially enumerated household will be missed (for details, see Griffin, 2005b). The panel and the Census Bureau agreed that cross-validation would be a suitable technique for comparing rival nonnested models. In cross-validation, the sample is split so that the model can be fitted to one part and the accuracy of predictions evaluated on the other part; the accuracy of prediction is thus not overstated due to fitting and evaluating the model on the same data. A standard approach is to split the data into several equal-sized pieces and remove each piece in turn from the fitting data set. The performance of each fitted model is assessed using some loss function in predicting the values for the set-aside fraction, and the loss function is averaged over all of the replications so defined. The Census Bureau implemented cross-validation using 100 equal-sized groups, and the loss function used was the logarithmic penalty function from Haberman et al. (1998). Finally, the average over all 100 groups was weighted using the A.C.E. survey weights.

The results of the Census Bureau's cross-validation comparison of the five alternative logistic regression models to the 2000 A.C.E. poststratification (Griffin, 2005a) are given in Table 3-1.The column labeled Correct Enumeration provides the cross-validation statistic for each of the six models in estimating the correct enumeration rate, and the column labeled Match provides the cross-validation statistic for each of the six models in estimating the match rate. The orderings observed and to a substantial degree even the average weighted log likelihoods (not shown here) did not change when

TABLE 3-1 Cross-Validation of Six Preliminary Logistic Regression Models: Average Weighted Log Penalty Function

Model	Number of Parameters	Correct Enumeration	Match
1. Poststratification	416	.2351	.2603
2. Main effects and two-way interactions	150	.2349	.2598
3. Main effects	23	.2354	.2598
4. ROAST	98	.2355	.2617
5. ROAST main effects and 2-way interactions	62	.2355	.2618
6. ROAST Main effects	14	.2360	.2619

NOTE: The logarithmic penalty function that was used in the cross-validation for the correct enumeration rate modeling was:

$$-\frac{1}{W_E}\left\{\sum_{i\in E-sample} w_{ei}\left[p_{cei}\log(\hat{p}_{cei})+(1-p_{cei})\log(1-\hat{p}_{cei})\right]\right\}$$

, where W_E is the weighted total for the E-sample, w_{cei} is the sampling weight for the j^{th} E-sample individual, p_{cei} is the correct enumeration rate, and $\hat{p}_{cei}$ is the predicted correct enumeration rate from the model. An analogous function was used for match rate modeling. Given the negative sign in this expression, smaller values of this statistic imply a better fit to the data.

the number of cross-validation replications changed from 100 to 25 or 20 (also not shown here). On one hand, the similarity of fit across the models suggests that many of the interactions in the poststratification model are relatively small. On the other hand, these findings support the view that even the most effective of the five alternative models, model 2 (main effects and two-way interactions of the poststratification variables), offers only minor benefits over the full poststratification. However, these models are limited to use of the variables in the 2000 poststratification and do not assess the potential of other predictors or model forms.

The panel further examined the use of cross-validation to assess the impact of the use of survey weights on the performance of the model. To examine this, using the logistic regression model with only the main effects from the poststratification (model 3), we formed 100 groups for the cross-validation. (This was done in two ways to examine the degree to which the block clusters were homogeneous. In one computation, we randomly selected P- and E-sample persons into 100 groups for cross-validation without regard for block cluster membership; in the second computation, we randomly selected P- and E-sample persons into 100 groups maintaining the block cluster structure of the A.C.E. sample design.) Using the cross-validation, we compared the performance of the logistic regression model unweighted by the survey weights with the performance of the logistic regression model weighted by the survey weights, with performance assessed using weighted log likelihood penalty function. The results are given in Table 3-2.

TABLE 3-2 Examination of Variance of Cross-Validation and Impact of Survey Weighting

		Average Log-Likelihood Penalty Function over 100 cross-validation replications	
		E-sample	P-sample
Unweighted			
	Random selection	.278181	.327804
	Maintain clusters	.278484	.328091
Weighted			
	Random selection	.235700	.261477
	Maintain clusters	.235982	.261930

NOTE: See note to Table 3-1 for details on the average log-likelihood penalty function.

The results suggest that use of the survey weights in computing the logistic regression coefficients substantially improves the performance in comparison to unweighted fitting, as assessed by the weighted criterion. This raises the possibility that inclusion of the survey design variables as predictors may provide some benefits.

Any predictors used in a logistic regression model must be available from census data to allow estimation of net census error nationally (at least in the form currently preferred by the Census Bureau). This restricts the available predictors to functions of the six factors used in the A.C.E. poststratification, a few additional variables from the short form in 2010, any variables collected during census processing, and contextual variables collected at aggregate geographic levels (say, from the American Community Survey or E-StARS). Recently, Schindler (2006) examined many of these other possible variables to see whether they provided substantial additional benefits as additional factors in producing post-strata (but not in a logistic regression approach). He considered the following variables: (1) geographic—census region, state, urban-rural, and mode of census data collection (mailout-mailback, list-enumerate, or list-leave); (2) contextual variables at the tract level—mail return rate, and percentage minority; (3) family and housing variables—marital status, relation to the head of household, and structure code (single unit or multiunit); and (4) census operational variables—indicator of mail or enumerator return, date of return, and proxy status. Schindler (2006) did not discover any variables that provided substantial benefit over and above that of the 416 indicator variables from the poststratification used in A.C.E.

This analysis, while extremely important, should not be considered conclusive, in particular in the context of a logistic regression model. For example, in the related problem of examining large numbers of subsets of a collection of possible predictors for use in regression-type models, it is difficult to know whether one has missed an effective combination of variables. This is difficult work, and it may be that further examination of potential predictors (or transformations of the predictors or interaction terms of the predictors) may still prove useful. (To assess the novel contributions of sets of covariates that are correlated, principal component analysis might provide useful insights.)

The panel strongly advocates further work in developing these logistic regression models, given their promise. We add two additional possibilities for broadening the approaches under consideration that the Census Bureau may wish to examine. The Census Bureau may look into the applicability of some of these methodologies, even if only to aid in the search for predictors for their logistic regression models.

First, it is not necessary that one logistic regression model be used nationwide. Different regression coefficients and even different predictors could be used for different geographic and/or demographic domains.

Second, logistic regression is only one of many statistical models that predict a dichotomous dependent variable. Of late, methods such as classification trees have been shown to have some applicability. One way to consider this research problem, broadened to encompass not only net coverage error modeling through use of logistic regression, but also census component coverage error measurement, is that these problems are essentially discriminant analysis problems. With respect to net coverage error measurement, the Census Bureau needs to identify variables that are predictive of match rate or erroneous enumeration rate. With respect to component coverage error, the Census Bureau needs to identify variables that are predictive of the rate of census omissions, erroneous enumerations, duplications, and being enumerated in the wrong location (for various definitions of wrong location). Taking the example of match rate, there are two types of individuals in the PES, those who match and those who do not, along with a number of predictors that might be helpful in the discrimination. Identification of these predictors might utilize logistic regression, but there might be advantages to the use of other, more flexible techniques, such as classification trees, recursive partitioning, support vector machines, and modeling with flexible link functions. For instance, classification trees develop a tree structure of decision rules, indicating a prediction of matched or unmatched, that identify subsets of the joint range of values defined by the possible predictors for which the percentage of matches or nonmatches is substantially different than for the overall data. Such an approach does not rely on the linearity used in logistic regression modeling and is therefore more flexible. Even if such an approach was not used in a production capacity in 2010, new information about what types of people or addresses are missed in the census might be discovered through use of these techniques.

What Are Legitimate Predictors in the Logistic Regression Model?

An issue concerning allowable logistic regression predictors is related to an issue that was raised when the poststratification design used for A.C.E. Revision II was modified from that used in the original A.C.E. The Census Bureau decided to include new factors in the poststratification that were available only for the E-sample due to their predictive power, resulting in different poststrata for the E-sample and the P-sample. The Census Bureau thought that these new factors would provide preferable poststrata for estimation of the probability of correct enumeration status. The new factors were (1) a variable indicating whether the E-sample enumeration was or was not a proxy enumeration and (2) a variable indicating the type of census return (early mail return, late mail return, early nonmail return, and late nonmail return). While the addition of these variables to the E-sample poststrata may have improved the partitioning of the E-sample into more homogeneous groups to reduce correlation bias and to improve synthetic (small-area) estimation, there was a concern that the difference in poststrata for the E-sample and the P-sample might cause a substantial number of failures in the balancing assumption. For example, a proxy enumeration often results in an interview with insufficient information. Insufficient information enumerations were treated in 2000 as if they were erroneous enumerations, and the P-sample enumerations that would have matched to those cases were treated as census omissions. Since the E-sample cases were proxy enumerations and therefore placed in a poststratum that did not exist for the corresponding P-sample cases for A.C.E. Revision II, these errors would be unlikely to balance.

It is clear that a similar issue arises with the use of logistic regression models of both the match rate and the correct enumeration rate, but it is substantially more difficult to assess. That is, P-sample information can be used to model matching rate, and census information can be used to model correct enumeration rate. If the variables for these two logistic regression models are different, coverage rate estimates for some combinations of these variables might be biased, although it is not known whether this would cause bias for the domains (defined by geography, age, race/ethnicity, etc.) that are of interest for census estimation. It is therefore important to determine when the benefit of improved predictive power outweighs the loss from balancing problems. The problem may be reduced in the 2010 census given the likely reduction in the number of "erroneous" enumerations; given the collection of data on census residence, the removal of duplicates in real time, other data improvement processes; and given the improved matching of KE cases.

In related research, Mulry et al. (2005) examined the following anomalous results in A.C.E. More than 5 percent of incorporated places[8] in 2000 had an estimated net *overcount* of greater than 5 percent, and 0.5 percent had a net overcount of greater than 10 percent. This result runs counter to findings from the 1980 and 1990 coverage measurement programs of the potential degree of net overcoverage due to true erroneous enumerations and duplications. In contrast, only 0.1 percent of places had an estimated net undercount of greater then 5 percent, and nationally, the degree of overcoverage and undercoverage were of essentially the same magnitude in the 2000 census. There is a concern that the lack of balance of designated erroneous enumerations and designated omissions mentioned above may be due to the use of proxy status and the type of census return as poststratification variables for the E-sample but not for P-sample computations.

[8]See http://www.census.gov/dmd/www/ACEREVII_PLACES.txt for a list.

To examine this further, Mulry et al. (2005) demonstrated that by using proxy status in the E-sample poststratification, there were 91 places with a net overcount of greater than 10 percent, but if it is assumed that there was no error for proxy enumerations, this changes to only 16 places with net overcounts greater than 10 percent. Furthermore, if the assumption is made that there are no errors for proxy enumerations and also that there are no errors for late nonmail returns, the result is that there are only four places with a net overcount of greater than 5 percent. Given this and given that 27 percent of proxy enumerations had insufficient information for matching and follow-up, it is clear that proxy enumerations could be involved with substantial balancing error. The Census Bureau concluded that proxy enumerations contributed to these anomalous findings, but the judgment was that this was not the only cause.

Related research carried out by Spencer (2005) examined the quality of synthetic estimates for block clusters based on A.C.E. Revision II estimates, either using 938 E-sample poststrata and 648 P-sample poststrata, or using the same 648 poststrata for the E- and P-samples. His findings, in which the standard of comparison was either (a) the direct dual-systems estimate or (b) the census count plus people found in the P-sample who were omitted in the census for each block cluster, suggested that coarser but consistent poststrata may have provided more accurate estimates of net coverage error than finer poststratifications based on different E- and P-sample stratifications. However, for large blocks with proxy rates greater than 10 percent, the finer and inconsistent poststrata performed better.

A concern raised by Alho (1994) is whether a problem is caused by the use of census operational variables as predictors in these models. Since proxy enumeration and type of census return are both operational variables, it is possible that they should not be included as predictors in these models. Furthermore, as argued in Griffin (2000), it is conceivable that errors in responses, such as for household composition, could result in persons either being assigned to the wrong poststrata or being given incorrect covariates for use in logistic regression models, resulting in the failure of errors to balance. The panel hopes to address this general topic of which variables should and should not be included in the logistic regression models in its final report.

Modeling Geography Via Random Effects

In addition to the systematic effects of the variables described above, match rates and correct enumeration rates may also vary across the local census offices used to manage the workload in the census. The local office identifiers are on the A.C.E. research database, but they were not included in the six logistic regression models described above or the study by Schindler (2006). The reason census office indicator variables might be predictive of match and correct enumeration rates is because factors that are particular to small areas could affect ease of enumeration. For example, local economic conditions and the expertise and capabilities of local census office administrators could vary. Because of the large number of local census offices (over 500) and the limited amount of data for each, these effects are more naturally represented as random effects. By including these random effects in the logistic regression models, the Census Bureau could estimate the effects of individual offices on match and correct enumeration rates and obtain valid estimates of the contribution of variability across offices to uncertainty about coverage rates in each area.

Malec and Maples (2005) explored this approach by incorporating random effects into a synthetic estimation model and then measuring the variance component of the random effects for local census offices. The ultimate objective of this approach is a small-area estimation methodology that would provide a compromise between synthetic estimation and a separate design-based estimator for each local office area.

Because of the A.C.E.'s complex design (weighted cases within samples of block clusters), many of the empirical correct enumeration rates and match rates used in Malec and Maple's model are more variable than the nominal sample sizes would indicate. To account for the extra variability, Malec and Maples (2005) used a pseudo-likelihood approach with effective sample sizes estimated via the bootstrap.

In this approach, both logistic regression models (for match rate and correct enumeration rate) have the following generic form: $log\left(\frac{p_{i,k}}{1-p_{i,k}}\right)=\beta_i+\mu_k+\alpha_{i,k}$ where β_i is the fixed effect for i^{th} poststratum membership, μ_k is a random effect for the k^{th} local census office, and α_{ik} is model error. Furthermore, $\mu_k \sim N(\underline{0},\Sigma)$, and $\alpha_{ik} \sim N(\underline{0},\gamma^2_{ce(i)})$, where *ce(i)* is an index representing the collapsing of the poststrata into 11 or 8 cells, depending on whether the model is applied to the E- sample or the P-sample. Malec and Maples (2005) were able to estimate the large number of parameters in these models using Bayesian simulation.

This research suggests that inclusion of small-area effects could substantially improve coverage estimates. The work is still preliminary, and there remain outstanding questions concerning how best to treat the complex sample design, how many random effects can be included and at what level of aggregation, what is the best way to estimate the model parameters, and how should model fit be assessed. The panel was impressed with this high-caliber research addressing an important issue in coverage modeling and strongly advocates further work in this area.

PANEL COMMENTS ON THE RESEARCH ON LOGISTIC REGRESSION MODELING

The immediate objective of the Census Bureau's research on logistic regression is to determine whether it is preferable to poststratification for estimation of net coverage error in 2010 for small domains. Part of this assessment must include whether the model can be relied on in a production environment. The panel supports this research, anticipating that it is likely to identify an approach that will be preferable to poststratification. This research is consistent with arguments in National Research Council (2004b) supporting the use of model-based alternatives to poststratification.

However, the panel would like the Census Bureau to explore a wider range of options in determining what model form and predictors work best in a predictive environment. Also, the Census Bureau's comparison of the logistic regression approach to poststratification when the logistic regression predictors are restricted to those used in the poststratification ignores a primary benefit of logistic regression of accommodating a larger number of predictors. This suggests that a more appropriate comparison is between the 2000 poststratification and logistic regression models with additional variables determined to provide additional predictive power.

Furthermore, there have been a variety of studies, outlined in Chapter 4, especially ethnographic work, that provide information as to why certain housing units are missed in the census and why people with various characteristics are missed in otherwise enumerated housing units. This information is moderately consistent with the variables currently included in the logistic regression models being examined by the Census Bureau, but the linkage between the research findings and the predictors in these models is not as direct as one would like. We think that the logistic regression models need to represent what is known about the sources of census coverage error, to the extent that this information is represented on the short form and in available contextual information. There also seems to be an unnecessary rush to pursue a model that can be used in a production environment, while there is still time to operate in a more exploratory manner. The panel therefore thinks that the Census Bureau has been too cautious in its examination of potential sets of predictors. The six models that have garnered the majority of attention to date are too similar to learn enough about what model forms and collections of predictors will work well. The Census Bureau should therefore expand the important research carried out by Schindler (2006) and apply it to the logistic regression models, attempting to identify unanticipated correlations between match rate or correct enumeration rate and the available predictors, using cross-validation to evaluate the resulting logistic regression models.

With respect to model form, the Census Bureau has also carried out some preliminary work on a very different use of two logistic regression models to model census net coverage error (see Griffin, 2005b). The first logistic regression models the probability that a housing unit will be missed in the Census Bureau's Master Address File. The second logistic regression model, conditional on the first, estimates the probability that, given that the housing unit is included in the Master Address File, an individual with certain characteristics will be missed. A number of details remain unclear with this approach, including how to handle erroneous enumerations and duplications. However, the panel strongly endorses further work on this and other modeling ideas that, even if not used in a production environment, will add to the Census Bureau's understanding of census coverage error.

Finally, the switch from use of poststrata to logistic regression modeling has important implications for census data users in communicating summaries of net coverage error. First, logistic regression modeling is likely to be more statistically efficient in its use of data than poststratification and, if so, may support estimates at lower levels of geographic and demographic aggregation. Therefore, the Census Bureau should examine what the reliability will be for estimates of various levels of aggregation and consider releasing estimates at a more detailed level than the A.C.E. poststrata, should the estimates support that. Second, for ease of comparison, while there is likely to be no poststrata in 2010 due to the use of logistic regression modeling, the Census Bureau should consider release of estimates of net coverage error for the 2010 census for comparable aggregates to support the comparison of net coverage error from census to census.

Recommendation 1: The Census Bureau should evaluate, for use in the 2010 census coverage measurement program, a broader range of models, most importantly logistic regression models, for net coverage error that includes variables in addition to those used to define the A.C.E. poststratification. These should include a wider range of predictors (e.g., geographic, contextual, family, and housing variables and census operational variables), alternative model forms (e.g., classification trees), and the use of random effects to model small-area variation.

The panel hopes to provide more guidance on the issue of which covariates to include in these logistic regression models in its final report. In the mean time, the Census Bureau should continue to investigate the full range of predictors while the panel and the Census Bureau continue to consider for which applications models with various predictors are appropriate.

RESEARCH ON CONTAMINATION DUE TO THE COVERAGE FOLLOW-UP INTERVIEW

Previous PESs initiated their data collection after the conclusion of the census data collection, with the minor exception of telephone A.C.E. interviewing in 2000. In 2000, this meant starting the A.C.E. nontelephone field interviewing after the conclusion of the nonresponse follow-up and the CEFU and CIFU interviewing. This was done for two reasons: (1) to avoid the possibility that the A.C.E. interview might impact the still incomplete census operations, thereby causing the PES blocks to be unrepresentative of the census, and (2) so that the evaluation that A.C.E. provided was of the complete census. However, the wait to begin the A.C.E. interviews increased the number of movers in the period between the census and the A.C.E., which reduced the quality of the data collected for A.C.E.

Any impact of the PES (CCM) interview (or other PES operations) on census processes in the PES blocks is a type of contamination. One way in which contamination might operate is if the census follow-up interview were affected by confusion with the already completed CCM interview. One possible impact is a refusal to participate in the census follow-up interview, but one can also posit other more subtle impacts on the census follow-up interview from CCM operations.

The impact of contamination on the entire census is essentially negligible, since the PES blocks represent a very small percentage of the country (less than 1 percent in

2000). However, given that contamination could result in a census in the CCM blocks that was not representative of the census in the remainder of the country, it might lead to substantial bias in the dual-systems estimates of net undercoverage.

In previous censuses, waiting for the various follow-up interviews and other coverage improvement programs to be completed prior to the collection of PES data was of less concern, since these were, generally speaking, relatively limited operations that could be concluded fairly expeditiously. However, in 2010, as noted above, the CFU interview could potentially involve a large fraction of the households in the United States and take a substantial time to complete. This would push back the CCM interviewing to September or later, resulting in a substantial increase in the number of movers and generally reducing the quality of the data collected in the CCM. Furthermore, there is a substantial similarity to the CFU and the CCM questionnaires, which might increase the possibility of contamination.

This issue has two aspects. The first is assessing the degree to which having the CCM interview precede the CFU interview affects the responses collected in the CFU interview. Attempts to measure contamination in the 1990 and 2000 censuses found no appreciable contamination (see, e.g., Bench, 2002), but, as argued above, the threat of contamination in the 2010 CFU seems more serious. If this contamination is ignorable, then the Census Bureau could let the interviews coexist in the field in 2010, in which case, one would also like to assess the impact of the CFU interview on the quality of the CCM interview.

There were two attempts to measure contamination in the 2006 census test. The first attempt compared interview rates and rates of erroneous enumerations and omissions in the two populations defined by the order of the interviews. This analysis was stratified by the various situations that result in a CFU interview, listed above. However, the measurement of contamination was indirect, and the modest sample size reduced the statistical power of the analysis. In addition, there was a matched pair design in which a second sample using the same sample design as the CCM was selected, using a block geographically proximate to the CCM sampled block. Then the population estimates for the two samples were compared. Again, the small sample size for this study was a concern.

Although it is too late for the 2006 test, the panel was interested in more direct observation of the impact of several proximate interviews used to determine the residents of a household. It is possible for a household to have a form mailed to it, with nonresponse resulting in several attempts to follow up by a field enumerator. If one of the various situations generating a CFU interview occurs, there will be attempts to carry out a CFU interview, and if then selected into the CCM sample, the household will be interviewed again, and finally, if there is a difficulty in matching to the E-sample, the household could be field interviewed a fourth time. To better understand the impact of several interviews occurring close in time and with similar content on respondents, the Census Bureau could carry out a limited test of this during 2007 or during the 2008 dress rehearsal. The panel is concerned that after the second interview, the chances either of a refusal or the collection of poor-quality data could increase.

The second aspect of the contamination issue is what to do in 2010 if appreciable contamination is either observed or cannot be ruled out. One might address this problem in several ways (see Kostanich and Whitford, 2005, for a discussion of some of these approaches).

1. **Combine the CFU and the CCM interviews into one multipurpose interview.** The panel has some sympathy for this position, given the similarity of the interviews. However, the CCM interview must be an independent count of a housing unit to satisfy the assumptions underlying DSE, whereas the CFU interview is dependent on information received in the initial census questionnaire. It is therefore difficult to combine these interviewing instruments.
2. **Have the CFU interview occur either before or after the CCM interview, but apply the CCM coverage measurement program to the census before the application of the CFU interview.** This is referred to as evaluating a truncated census, since the definition of the census for purposes of coverage evaluation is the census that existed prior to the taking of the CFU interview. Any enumerations added by carrying out CFU interviews after the CCM interviews were completed could be treated as "late additions" were treated in 2000, that is, removed from the census for purposes of coverage measurement. A problem with this approach is that if the CFU adds an appreciable number of people, or corrects the enumerations of an appreciable number of people, one is evaluating a truncated census that is substantially different from the actual census. Also, if these additions or corrections are considerably different in coverage error characteristics in comparison with the remainder of the population, that would add a bias to the dual-systems estimates. One could include the CFU interviews that occurred prior to the CCM interviews in the truncated census, in which case the net coverage error models could condition on whether a CFU interview was carried out prior to the CCM interview, which would remove any bias if the P-sample inclusion probabilities depended on the occurrence of the CFU interview (but not on its outcome—for details, see Bell, 2005). Information on what the CFU interview added from outside the CCM blocks also could be used in these models. There are some operational complexities to this idea, including the need to duplicate the formation of relatively large processing files. Finally, as mentioned previously, one is not evaluating the complete census, and therefore to assess components of census coverage error resulting from the application of the CFU, one would need to carry out a separate evaluation study outside the CCM blocks, which is a serious disadvantage.
3. **Do not use the CFU in the CCM blocks.** This avoids any contamination, but then the CCM evaluates an incomplete census, with essentially the same problems listed in (2).

4. **Let the CFU and CCM interviews occur in whatever order they do, and treat contamination as a constant effect times an indicator variable for which of the two interviews comes first for households that have both CFU and CCM interviews.** The difficulty with this approach is that it is not clear what the impact will be on whichever interview comes second, so it is not clear that contamination can be effectively modeled through use of a constant effect. For example, contamination might be a function of various characteristics of the household and therefore be subject to various interactions.
5. **Delay the CCM interviews until the CFU interviews are complete.** This does solve the contamination problem. However, coverage evaluation interviews that occurred in August 1980 were less useful than those in April due to the large number of movers that occurred during the four-month period. Therefore, this could have a substantial, negative impact on the quality of the CCM data that are collected in 2010, depending on how long one has to wait.

The panel has not yet come to a consensus on this question. The panel was interested in further examination of the implications of a truncated census (option 2) or combining the two instruments (option 1). The Census Bureau believes that the best approach is to delay the CCM interviews until after all CFU interviews are completed (option 5). The basis for this decision was that in this way the Census Bureau will not plan to have a substandard census in any area (which would certainly be true of option 3), and combining the interviews might harm both interviews in option 1. Furthermore, option 4 is unknown and difficult to test prior to the 2010 census. (For more details on the Census Bureau's views on contamination, see Kostanich and Whitford, 2005.) However, the panel did not find the argument about the difficulty of duplicating census processing files for option 2 compelling, given the current availability of inexpensive computer memory.

The panel does have concerns about not starting the CCM interviews until September 2010, given the increased number of movers that this would create between Census Day and the CCM interview. It is hoped that by expediting certain operations, an August start for the CCM might still be possible. For this reason, it is important to collect good data in 2006 and 2008 on the impact of delays of various length on the number of movers. In this and several other respects, the results from the 2006 census test will inform the Census Bureau's position on this issue.

SAMPLE DESIGN FOR THE CCM POSTENUMERATION SURVEY

An important question concerning the CCM program is what modifications should be made to the design of A.C.E. in looking toward the CCM in 2010, given the change in objectives in coverage measurement between the 2000 and the 2010 censuses. That is, to what extent can the new goal of process improvement be incorporated into the design of the CCM PES?

The proposed design for the CCM PES in 2010 is as follows (for details, see Fenstermaker, 2005). The Census Bureau is assuming that the CCM PES will draw a sample of 300,000 housing units, with primary sampling units comprising block clusters. The panel is in support of Recommendation 6.1 of National Research Council (2004b), which supports a PES survey that would produce net coverage estimates of the same precision as that of the 2000 A.C.E. These block clusters are meant to contain around 30 housing units, and the plan is to subsample them in the event that they contain substantially more. These block clusters will be stratified into three categories: (1) medium and large clusters, with some subsampling within large block clusters, (2) American Indian Reservation block clusters, and (3) small block clusters, which will utilize a two-phase design to sample block clusters under a certain size but to retain all small block clusters greater than that. In allocating the sample of 300,000 housing units to states, the general approach will be to sample proportional to the total population of each state. However, each state's sample will contain a minimum of 60 block clusters, and Hawaii will be allocated 150 block clusters. In addition, there will be a separate American Indian Reservation sample drawn proportionally to the 2000 census count of American Indian and Alaskan Native populations living on American Indian reservations.

The rationale behind the state allocations for the 2010 CCM PES is that this is intended to be a general purpose sample, so any oversampling in comparison to proportional allocation needs to be strongly justified. In addition, the Census Bureau was very satisfied with the 2000 A.C.E. design, which this design roughly duplicates. The Census Bureau has no specific variance requirements for the 2010 CCM estimates, since production of adjusted counts is not anticipated.

The Census Bureau did examine some alternative specifications for the design of the CCM PES, using simulation studies of the quality of the resulting net coverage error estimates and assessment of components of census coverage error, especially estimation of the number of omissions and erroneous enumerations at the national level and for 64 poststrata (see Fenstermaker, 2005). The designs were (1) the design described above, with allocations proportional to total state population, but with a minimum of 60 block clusters per state, and with Hawaii allotted 150 block clusters; (2) similar to (1) except Hawaii is allocated only 60 block clusters; (3) a design in which allocations are made to the four census regions to minimize the variance of estimates of erroneous enumerations, but within regions, allocations are made proportional to state size; and (4) a design in which half of the sample is allocated proportional to the number of housing units within update/leave areas, and half of the sample is allocated proportional to each state's number of housing units.

Through use of simulations, for each design and PES sample, national estimates were computed of the rate of erroneous enumerations (and the rate of erroneous enumerations with mail returns, with nonresponse follow-up, and with CFU), the nonmatch rate, the omission rate, and the net error rate. Finally, national estimates of the population were computed, along with their standard errors. The same analysis was done at the poststrata level. One hundred replications were used for the simulation study. The results supported retention of the design that closely approximated the 2000 A.C.E. design, described above.

The panel has not yet come to a consensus on whether to recommend modifications to this design for the CCM PES in 2010. There is some concern that the allocation of a minimum of 60 block clusters to each state is too linked with the need to provide adjusted counts for states and not as targeted toward measurement of the rates of the four types of census component coverage errors. If it is the case that the households that are more problematic to count can be linked to relatively focused geographic regions, it would be interesting to evaluate a design that oversampled those areas to see the impact on the reliability of measurement of census component error rates. This is similar to design alternative (3) above, but what we are suggesting is more targeted than that.

Furthermore, we also think that the Census Bureau needs to give more consideration to its within-state allocations of block groups. For example, the possibility of oversampling block groups in predominantly minority areas with, say, large percentages of renters is an alternative that deserves further consideration. The panel is also not clear why the Census Bureau is not making greater use of their planning database, which provides an indication of the difficulty of enumerating block groups.

The objective of settling on a sample design for the CCM in 2010 is a difficult task. There are two general objectives of the coverage measurement program for 2010. First, there is the primary objective put forward by the Census Bureau, which is the measurement of census component coverage errors at some unspecified level of geographic and demographic aggregation. Second, there remains the need to measure net coverage error at the level of the poststrata used in 2000 in order to facilitate comparison with the 2000 census. To address the first goal, one would like to target problematic domains. However, one has to guard against unanticipated problems that might appear in previously easy-to-count areas. To do that and to provide estimates of net coverage error across the United States, a less targeted design is needed. These various demands individually argue for very different designs, and mutually accommodating them, to the extent possible, is challenging. The panel anticipates providing much more direction on this question in its final report.

Research on the Use of Administrative Records in Support of Coverage Improvement and Coverage Measurement in 2010

The Census Bureau's research program has explored decennial uses of administrative records, that is, data collected as a by-product of administering a governmental program, since the 1980s. Possible uses include (1) a purely administrative records census; (2) improving census nonresponse follow-up either by using enumerator follow-up only when administrative records do not contain the required information or, alternatively, using administrative records to complete information for households that do not respond after several attempts by field enumerators; (3) improving the Master Address File using addresses in administrative records; (4) assisting in coverage measurement, for example, through use of triple-systems estimation (a generalization of

DSE in which the third system is a merged list of individuals from administrative records); and (5) assisting in coverage improvement, for example, by identifying census blocks for which the census count is likely to be of poor quality. We emphasize that the use of administrative records could be the most promising idea for assisting in the measurement of omissions of hard-to-enumerate groups.

However, until recently, partly due to the limited quality of the available administrative records, including the currency of the information, especially for addresses, the computational burden, and concerns about public perceptions, neither these nor other applications of administrative records have been implemented during a decennial census. (An approach to the problem of currency of address can be found in Stuart and Zaslavsky, 2002). As a result, until 2000, there was no comprehensive census test of the use of administrative records for any purpose, although there were earlier assessments of the coverage of merged administrative lists.[9]

Now, however, several of these concerns have been ameliorated. The quality and availability of national administrative records are improving, computing power has increased dramatically, and as a result the very active research group on administrative records at the Census Bureau has achieved some impressive results. The primary program and database, referred to as E-StARS, now has an extract of a validated, merged, unduplicated residential address list with 150 million entries, 80 percent of which are geocoded to census blocks, and another extract of a validated, merged, unduplicated list of residents with demographic characteristics. These lists are approaching the completeness of coverage that might be achieved by a decennial census. Seven national files are merged to create E-StARS, with the Social Security Number Transaction File providing demographic data.

The panel strongly supports this research program, and we think that there is a real possibility that administrative records could and should be used in the 2010 census, either for coverage improvement, for nonresponse follow-up, or for coverage measurement. Potentially feasible uses in the 2010 census include

- **To improve or evaluate the quality of either the Master Address File or the address list of the postenumeration blocks**. The quality of the Master Address File is a key to a successful mailout of the census questionnaires and nonresponse follow-up, and the quality of the independent list that is created in the PES blocks is a key to a successful coverage measurement program. E-StARS provides a list of addresses that could be used in at least two ways. First, the total number of E-StARS addresses for small areas could be checked against the corresponding Master Address File totals or PES totals to identify areas with large discrepancies that could be relisted. Second, more directly, address lists could be matched to identify specific addresses that are missed in either the Master Address File or the

[9]The Census Bureau operates under the constraint that information obtained from administrative records under confidentiality restrictions cannot be sent out to the field to assist enumerators, which prohibits the use of some applications of administrative records.

PES address listing, with discrepancies followed up in the field for resolution. Note that while administrative records could be used to improve the address list for either the census or the PES, to maintain independence they should not be used for both.

- **To assist in late-stage nonresponse follow-up**. The Census Bureau makes several attempts to collect information from mail nonrespondents to the census form. When these attempts fail to collect information, attempts are made to locate a proxy respondent and, when that fails, hot-deck imputation is used to fill in whatever information is needed, including the residence's vacancy status and the household's number of residents. If the quality of E-StARS information is found to be at least as good as that from hot-deck imputation or even proxy interviews, it might be effective to attempt to match nonrespondents to E-StARS before either pursuing a proxy interview or using hot-deck imputation. Especially with a short-form-only census, E-StARS might be sufficiently complete and accurate for this purpose. (It may ultimately be discovered, possibly during an experiment in 2010, that fewer attempts at collecting nonresponse data are needed by making use of E-StARS information after, for example, only one or two attempts at nonresponse follow-up, thereby shortening and reducing the costs of nonresponse follow-up.)
- **For item imputation**. The Census Bureau often uses item imputation to fill in modest amounts of item nonresponse. Item nonresponse could affect the ability to match a P-sample individual to the E-sample, and missing demographic and other information may result in an individual being placed in the wrong poststratum. Item imputation based on information from E-StARS may be preferable to hot-deck imputation. The use of E-StARS to provide item imputation is currently being tested as part of the 2006 census test.
- **To improve targeting of the coverage improvement follow-up interviews**. The coverage improvement interview in 2010, as currently planned, will follow up households with any of the following six conditions: (1) uncertain vacancy status, (2) characteristics for additional people in large households, (3) resolution of count discrepancies, (4) duplicate resolution, (5) persons who may have been enumerated at other residences other than the one in question, and (6) nonresidents who sometimes stayed at the housing unit in question. The workload for this operation might well exceed the Census Bureau's capacity to carry out the necessary fieldwork, given limited time and resources. Administrative records could possibly be used to help identify situations in which field resolution is not needed, for example, by indicating which of a set of duplicates is at the proper residence. (Uses of E-StARS like this are being attempted in the 2006 census test.)
- **To help determine the status of a nonmatch prior to follow-up of nonmatches in the PES**. It is very possible that nonmatches of the P-sample to the census may be resolved, for example, by indicating that there was a geocoding error or a misspelled name through the use of administrative records, thereby saving the expense and time of additional CCM fieldwork.
- **To evaluate the census coverage measurement program**. Many of the steps leading to production of dual-systems estimates might be checked using

administrative records. For example, administrative records information might be used to assess the quality of the address list in the P-sample blocks, to assess the quality of the matching operation, or to assess the quality of the small-area estimation of population counts. (However, any operation that makes use of administrative records cannot also use the same administrative records for purposes of evaluation.)

The administrative records group at the Census Bureau has already had a number of successful applications of E-StARS. First, an administrative records census was conducted in five counties during the 2000 census, and its quality was judged to be comparable to that of the census in those counties. Second, E-StARS was used to explain 85 percent of the discrepancies between the Maryland Food Stamp Registry recipients and estimates from the Census Supplementary Survey in 2001 (the pilot American Community Survey).

The panel considers this important and promising research that should play a key role in censuses beginning in the year 2020, given the potential for cost savings and quality improvement. With respect to use in 2010, since the various suggestions depend crucially on the quality of the merged and unduplicated lists of addresses and people in E-StARS, the use of E-StARS for any of the above purposes in 2010 will require further examination of the quality of the lists, as well as evaluation of the specific application in comparison to the current method used in the census. Until there are rigorous operational tests of both feasibility and effectiveness, it would not be reasonable to move toward implementation in 2010. Given where we are in the decade, it is unlikely that more than one of the above six bulleted applications could have sufficient resources devoted to support incorporation in the 2008 dress rehearsal, which is a necessity for implementation in 2010. Therefore there is a need to focus immediately on one very specific proposal.

The panel recommends that one of the above applications be developed sufficiently to support a rigorous test in the 2008 dress rehearsal with the goal of implementation in 2010 should the subsequent evaluation support its use. Furthermore, the Census Bureau should begin now to design rigorous tests of all the above suggestions for the use of administrative records, very possibly during the 2010 census, as a first step toward decennial census application of administrative records in 2020. We think that administrative records have great promise for assisting in understanding census omissions and therefore need to be used either for evaluation of the CCM or as a part of the CCM program.

Recommendation 2: The Census Bureau should choose one or more of the proposed uses of administrative records (e.g., tax record data or state unemployment compensation data) for coverage improvement, nonresponse follow-up, or coverage measurement and comprehensively test those applications during the 2008 census dress rehearsal. If a process using administrative records improves on processes used in 2000, that process should be implemented in the 2010 census.

We add that evaluations of the use of administrative records are often viewed as involving extensive, resource-intensive fieldwork. However, while fieldwork needs to be involved to some extent, much evaluation of administrative records can be accomplished if the Census Bureau structures its various databases collected from test censuses in a way that facilitates matching.

Furthermore, if data from E-StARS are used successfully in 2010, the Census Bureau should consider more ambitious uses of administrative data in the 2020 census. Specifically, the Census Bureau might use administrative data to replace the nonresponse follow-up interview for many housing units, not just late-stage nonresponse. Under this proposal, the Census Bureau would use data from administrative records to determine the occupancy status of some nonresponding housing units and the number and characteristics of its residents. To do so, the Census Bureau would have to develop criteria of adequacy of the information in the administrative records to establish the existence and membership of the household for this purpose. For example, agreement of several records of acceptable currency and quality might be considered sufficient to use the information as a substitute for a census enumeration, which would reduce the burden of field follow-up.

This would represent a substantial change in what constitutes a census enumeration, of at least the same conceptual magnitude as the change from in-person to mail enumerations as the primary census methodology. However, given that the completeness of administrative systems and the capabilities of matching and processing administrative records has been growing, while cooperation with field operations has declined, these contrasting trends make it increasingly likely that administrative records can soon provide enumerations of quality at least as good as field follow-up for some housing units. Furthermore, unlike purely statistical adjustment methods, every census enumeration would correspond to a specific person for whom there is direct evidence of his or her residence and their characteristics. The long-run potential for such broader contributions from administrative records is a reason to give high priority to their application in the 2010 census, in addition to their direct benefits in that census.

Two possible objections might be raised in opposition to this approach. First, this use of administrative records may be ruled to be inconsistent with interpretations of what a census entails in the Constitution. Second, public acknowledgment that this method is being used might have a negative impact on the level of cooperation with census-taking. These two issues would need to be resolved before the Census Bureau could go forward. Also, this is clearly dependent on the success of the more modest efforts suggested for possible use in 2010.

4

ADDITIONAL COVERAGE EVALUATION RESEARCH USEFUL FOR CENSUS ERROR REDUCTION

The Census Bureau's research program, described in Chapter 3, will lead to important improvements in coverage measurement in 2010 for assessment of components of coverage error. This chapter discusses other research activities that are potentially valuable but not part of current plans. Greater use can be made of data from the 2000 Accuracy and Coverage Evaluation (A.C.E.), and data captured in 2010 can be structured to facilitate exploration of the relationship between census component coverage error and specific census component processes. While some of these suggestions might be implemented in time for the 2008 dress rehearsal to provide guidance for the 2010 census, the design of the latter is relatively firm, and therefore most of the benefits would not be realized until the 2020 census, though plans for implementing these ideas would need to made prior to the 2010 census to collect and save the requisite information.

We begin by discussing the existing research literature on personal and household factors and census processes associated with components of coverage error. We argue that a key product of a census coverage measurement (CCM) program with the objective of census improvement is a database that jointly represents census processes; person, household, and area characteristics; and census component coverage error assessments. This database can support analyses of factors associated with census component coverage error, which would advance identification of census processes that can be improved. We then discuss how the Census Bureau can better use the 2000 data both to guide design of this database and to help complete the design of the 2010 coverage measurement program. We conclude with some thoughts about planning for coverage measurement in 2010 and how to report coverage error to users.

THE RESEARCH LITERATURE ON PERSON AND HOUSEHOLD CHARACTERISTICS AND CENSUS PROCESSES ASSOCIATED WITH COVERAGE ERROR

Demographic analysis and dual-systems estimation for the 1980, 1990, and 2000 censuses were not designed to identify characteristics of individuals, households, or areas that were associated with high or low rates of components of census coverage error, or processes responsible for these errors. These methods are limited for that purpose for at least two reasons.

First, demographic analysis and dual-systems estimation measure net coverage error, which obscures many offsetting census omissions and erroneous enumerations. Second, these coverage measurement programs only disaggregate coverage error by a limited set of variables: demographics (age, sex, race, ethnicity), some modest geographic detail (census region), and other variables that measure urban/rural, mail return rate (high/low), and owner/nonowner status. This is true for demographic analysis since it is limited to the information in the record systems utilized. The level of detail in dual-systems estimation has been limited by the restricted number of poststrata used and therefore to variables included in the poststratification. While many of these variables are associated with reasons for census coverage errors, relatively modest differences in net undercoverage rates between many poststrata in 1990 and in 2000 suggest that many of these associations are themselves modest. Furthermore, none of these factors has been chosen on the basis of potential links to potentially deficient census component processes.

Since the past two censuses conducted coverage measurement primarily to support adjustment, it is commendable that the Census Bureau has also devoted substantial resources to the study of factors associated with census coverage error. Studies of reasons for census omissions include several participant observation studies, first in the 1970 census (Valentine and Valentine, 1971), then during the 1986 Test of Adjustment Related Operations (e.g., Garcia-Parra, 1987), the 1988 dress rehearsal (Martin, Brownrigg, and Fay, 1990), and the 1990 census (Ellis, 1995). In addition, the Census Bureau supported ethnographic studies during the 2000 census (de la Puenta, 2004), as well as the 1993 Living Situation Survey, which assessed response to a variety of residence and household composition cues (see, e.g., Martin, 1999). These studies identified person- and household-level characteristics associated with the misinterpretation of the census residence rules or with noncooperation with the census, which might be due to mistrust of government or fear of exposure of illegal behavior (e.g., Brownrigg and de la Puenta, 1993; Bates and Gerber, 1998; Martin, 1999).

More quantitative studies include Fein (1990), who used logistic regression to identify factors associated with census undercoverage, and studies (e.g., Dillman, Treat, and Clark, 1994) of effects of mail presentation on census mail response (and hence potential undercoverage). Analyses by Ericksen et al. (1991) suggest that census undercoverage was greater in areas with low mail response rates, high crime rates and rampant drug use, or high rates of irregular housing, for individuals with low levels of English literacy or unfamiliarity with surveys (the poor and the less well educated), in housing units that share a common address or are likely to be omitted from the census Master Address File for other reasons, and households that include distant relatives and nonrelatives. Ericksen et al. (1991) also pointed out that coverage improvement programs, in particular those more distant from Census Day, were associated with a high rate of census coverage error, especially erroneous enumerations.

The less extensive literature addressing reasons for whole-household omissions (e.g., Childers, 1992; Moriarity and Childers, 1993; and in particular Ruhnke, 2003), suggests that there are substantial problems in enumerating households in small, multiunit buildings.

There has also been some research on factors associated with erroneous enumerations and duplications, with a good example being work on factors associated with duplicates in the 2000 census, especially with respect to group quarters (see, e.g., Feldpausch, 2001; Fay, 2004; Mule 2001, 2002).

While the research literature that has been only touched on here is considerable, the reasons for census omission and erroneous enumeration still remain poorly understood, as do the census component processes that would benefit from modification to reduce their frequency of occurrence. For example, a recent National Academies study of census residence rules (National Research Council, 2006) reported that little is known about the extent to which the following types of individuals were missed, duplicated, or erroneously included in the census: people with multiple residences and highly mobile populations (including snowbirds and sunbirds,[1] modern nomads, commuter workers and people in commuter marriages, and migrant farm workers), individuals in complex household structures (including children in joint custody, cohabiting couples, and recent immigrants), linguistically isolated persons, people in long-term-stay hotels and motels, people dislocated by disasters, and people residing in unusual housing stock.

The extent to which this research literature has directly motivated changes in census processes is unclear, but it is probably relatively limited, given the nonspecificity of the information collected. However, as mentioned in Chapter 1, some coverage improvement programs were added in 1980 and 1990 due to findings from demographic analysis and dual-systems estimation on the high differential rate of omission of young adult black men, and a number of the design changes for 2010 were consequences of information collected by A.C.E. in 2000.

INTEGRATING CENSUS PROCESS DATA AND PERSON, HOUSEHOLD, AND AREA CHARACTERISTICS WITH CENSUS COVERAGE ERROR DATA

Each stage of the decennial census consists of a number of alternative component processes. For example, there are a number of different ways in which an address for a housing unit can be added to the Master Address File. Also, various areas of the United States are initially enumerated using mailout-mailback, update list-leave, or list-enumerate (and other less common processes). There are various stages of nonresponse follow-up and coverage follow-up. Alternative ways of being enumerated include the Be Counted program (which allows people to provide census data if they believe they were missed in the census), telephone questionnaire assistance, and processes that help

[1]Snowbirds are people who live in northern areas but winter in southern ones. Sunbirds are people who live in southern areas but summer in northern ones.

households obtain foreign language questionnaires in the mail or receive other forms of language assistance, including actual enumeration. Very different techniques are used to enumerate people living in various types of institutions or other group quarters. This outline is only a hint at the many parts of a census process that is in total enormously complicated. For more details, see National Research Council (2004b:Chapter 4.)

As a result, a given household might take any of a number of paths through this census process "tree" to arrive at either a proper enumeration or a coverage error. The path depends on various characteristics of the household and its occupants, for example, the type and location of the housing unit, how complicated the relationships of the residents are, and their interest in cooperation. Recording the census process path taken and the corresponding person, household, and area characteristics is therefore crucial to understanding what factors may be associated with census coverage error.

This argument points to the need for a database that represents the census processes used to enumerate housing units, characteristics of the persons and housing unit and area, along with the assessment of correctness or type of coverage error represented by these cases. The 2006 census test attempted to collect such data, showing the Census Bureau's interest in determining the value of such a database. If a database can be created that contains this information, *properly linked*, statistical models can be developed that are likely to be very effective in identifying those combinations of characteristics and processes that jointly result in higher rates of census coverage errors. As mentioned in Chapter 3, searching for factors associated with census errors can be regarded as a discriminant analysis problem, since one has a number of individuals whom the census did or did not miss, or did or did not duplicate, or did or did not erroneously include, or did or did not enumerate in the proper geographic area (and there could be several definitions of proper area), along with many potential explanatory factors.

THE POTENTIAL FOR IMPROVED ANALYSES OF 2000 CENSUS DATA

The evaluations following the 2000 census of the various census processes and A.C.E. usually did not combine information on census processes with detailed information on person, household, and area characteristics, beyond the factors used for the A.C.E. poststratification. The master trace sample (see National Research Council, 2004a) was created using 2000 census process data to provide at least a portion of the analysis capability outlined here, but its value was limited since it did not include information from A.C.E. on census coverage errors. (For details, see Hill and Machowski, 2003.) The A.C.E. data were carefully analyzed to evaluate the quality of the various sets of adjusted counts that were considered for important uses between 2001 and 2003. However, those analyses were directed at evaluating the reliability of estimates of *net* error, not at assessing predictors of components of census coverage error.

There are many obstacles to further analysis of the 2000 data. Many census coverage errors in 2000 were not errors under the stricter definitions given in Chapter 2.

Furthermore, the 2000 census data are now six years old and therefore may not be accessible or fully documented.

Nonetheless, analyses of the 2000 census and A.C.E. data with the current objective of census component coverage error measurement in mind might provide insights about potential modifications in census or census coverage measurement processes, and they might identify appropriate topics for further research in the 2010 census. A few possibilities include

- Census omissions identified by A.C.E. could be matched to the merged E-StARS administrative records database to assess characteristics that predict omissions.
- Addresses of whole-household omissions could be matched to the 2000 Master Address File database, which includes a history of additions to and deletions from the Master Address File as it was created and improved. Analyses of the matched database could help determine whether the addresses for these missed housing units were ever on the Master Address File and were dropped for some reason.
- The 2000 A.C.E. data might be helpful in estimating how large the coverage follow-up interview might need to be in 2010.

These and other analyses certainly have problematic aspects, and the findings would not be confirmatory, only suggestive. However, these analyses would not require any fieldwork, and they might provide important information. Also, it is true that given the innovative plans for the 2010 census (described in Chapter 3), some deficiencies discovered in the 2000 census might no longer be relevant for 2010. Nonetheless, in many respects the 2010 design is quite close to that used in 2000, and more comprehensive evaluation of the latter would provide a better basis for understanding and evaluating the outcomes of the 2010 census.

Finally, using the 2000 census data in this way will help to clarify what data need to be saved from the coverage evaluation program, various management information systems, and other data associated with the execution of the 2010 census. It will also help to understand how best to structure a database to support analysis of census component coverage errors looking toward 2020.

LOOKING TOWARD 2010

The approach we propose argues that the Census Bureau, in order to satisfy its own goals for coverage measurement in 2010, needs to retain the necessary data from the 2010 census to support analysis of possible relationships between census component coverage error and census processes. The data that are retained should include information from the CCM program in 2010 on omissions, erroneous enumerations, enumerations in the wrong location, and duplicates, as well as the characteristics of the household and the local area in question. In addition, data from the various census processing files should be included indicating the specific census processes that were used to enumerate (or not enumerate) a given housing unit. A properly structured database would link the information on census processes, people, housing units, and local areas, and the information on census coverage error, to support analysis of the combinations of factors associated with census component coverage error. Included in this database would be information to determine whether a person was a P-sample correct enumeration, omission, erroneous enumeration, duplicate, or an enumeration in the wrong geographic area. Similarly, also included would be information to ascertain whether a person was a census correct enumeration, census omission, erroneous enumeration, duplicate, or an enumeration in the wrong geographic area. Finally, it would be possible to determine, if a person was omitted in either the P-sample or the census, whether the whole household was also omitted.

Data from the various management information systems also could be folded in to represent aspects of the quality of the application of census procedures in a given area. Finally, other contextual information about each housing unit and its residents can also be folded in, possibly from the American Community Survey and E-StARS.

The structure of this database is crucially important, but it lies outside the panel's expertise to provide its specifications. We are sympathetic with the challenge, since it is a complex undertaking to determine what data to include and how to link it to other related data. For example, it is likely to be useful to include the detailed information on the history of the formation of the Master Address File (including all of the various operations that can add or remove addresses from the list), the totality of results from nonresponse follow-up (including how many attempts at enumeration were made and whether the ultimate response was a proxy enumeration), the results from the coverage follow-up interview, the degree of item and unit nonresponse, and the various stages of matching of the postenumeration survey to the census. Representing this complexity will not be a simple matter.

Furthermore, it is unclear how much from the census processes can be saved in real time on a production basis. If constraints dictate the need to save data on a sample basis from various sources, it is unclear how that will reduce the utility of the database for answering various types of questions. We are uncertain as to the feasibility of data capture, and we hope to say more on this topic in our final report. However, the key is to try to anticipate the type of analyses that would be useful to carry out and then determine a database structure and contents that facilitates carrying out those analyses.

Once the database is available, two types of analyses should be carried out. First, many hypotheses generated from reports from the field, from census tests, and other sources can be confirmed using these data. For example, one might suppose that households that have been newly constructed are often missed in the Master Address File, or one might assume that linguistic isolation is a major cause of census undercoverage, or one might suppose that children in joint custody arrangements and people in nursing homes are often missed and often duplicated. These types of questions will be easier to address with a properly structured database. Second, in addition to these confirmatory studies, the Census Bureau should also carry out exploratory studies, examining the data for unanticipated interesting relationships between census coverage error and census processes that might indicate a census process that was not effective for a small area or subpopulation of individuals or housing units.

What is important in this analysis is practical significance. The appropriate metric is how many census coverage errors could potentially be corrected through a modification of the relevant census process, both nationally and for important geographic and demographic domains.

While promoting the benefits of the construction of this analytic database, we are aware that feedback loops linking census component coverage errors to specific components of census processes are always going to be somewhat limited in their ability to pinpoint specific problematic components and to suggest alternatives. For example, knowing that there were many erroneous Be Counted enumerations in big cities is not extremely helpful toward identifying an alternative process that would reduce that error, since these cases tend to be problematic under the best circumstances. In addition, some of the situations discovered may be for such small populations that the census coverage measurement program will not have enough observations to support analysis. We intend to discuss in more detail in our final report how a feedback loop for improving census methodology might operate and what can be done to make it more effective.

ASSESSMENT OF CENSUS QUALITY WITH NEW METRICS

In Chapter 2, we suggest that coverage measurement results should be reported to inform users as to the quality of census counts. The appropriate summarization is not specified, except that the Census Bureau needs to provide assessments of net undercoverage for a variety of geographic and demographic domains. This has been accomplished in the previous two censuses with the release of information on undercoverage for census poststrata. With the new emphasis on four types of census component coverage error—rates of erroneous enumeration, duplication, enumerations in the wrong place (at various geographic resolutions), and omission—an important question is the extent to which users could benefit from having more local knowledge of these four types of errors and, if so, how should this be communicated?

It is unclear whether information on these rates for specific domains would be that useful to users, given their understandable interest in net error. Furthermore, what if the analysis of CCM data demonstrated several (nongeographic) predictor variables that were strongly associated with, say, omissions? Should the knowledge of these predictors be made available to users in some way? Should the communication be in the form of research reports without any sense of the amount of error for a given domain? This is another topic that the panel intends to consider for inclusion in the final report.

We urge the Census Bureau to initiate the development of a database that jointly represents person, household, and housing unit characteristics, census processes, and census component coverage error to facilitate the development of statistical models to help link census errors to census processes in need of improvement.

Recommendation 3: The Census Bureau should collect data in the 2010 census to support development of a database that links person, household, and housing unit characteristics, census processes, and the presence or absence of census component coverage error. This database should also represent coverage errors, including erroneous enumerations, enumerations in the wrong place, duplications, and omissions. The use of this database would better identify the sources of high rates of census component coverage error.

Finally, the panel realizes that the various research and development activities already started by the Census Bureau on contamination, KEs, identification of duplication, CCM forms and sample design, analysis of the 2006 and 2008 test results, etc., are challenging. Furthermore, the panel has made a number of suggestions for further research, especially concerning the development of the logistic regression models, and we have suggested a new framework for analysis that will require additional staff resources. Given the importance of all of this research, which in essence is guiding the development of a feedback loop to improve census-taking over time, the panel thinks that the resources currently devoted to this effort are insufficient.

Therefore, we strongly advise the Census Bureau to provide the coverage measurement group with sufficient resources to carry out its current research program, its planning activities regarding the dress rehearsal and the 2010 census, and the activities listed in this report --including searching for covariates for the logistic regression models on net coverage error, greater targeting of the design of the census coverage measurement survey, further development of the small-area random effects modeling of CCM match rate and census correct enumeration rate, use of administrative records in coverage improvement and coverage measurement, further analysis of A.C.E. data to assist in the design of the census and CCM in 2010, and creation of the database on individual and household characteristics, census component coverage error, and census processes to help diagnose reasons for census coverage component error. Unless properly supported, the panel is concerned that resources will be insufficient to carry out the wide variety of research and planning activities needed in moving toward 2010.

Recommendation 4: Given the number of important research activities currently under way, the needed design of the coverage measurement programs in the dress rehearsal and in the 2010 census, and the additional research suggested by the panel, the Census Bureau should provide the coverage measurement group with sufficient resources to carry out its current research program, its planning activities regarding the dress rehearsal and the 2010 census, and the activities listed in this report.

This report provides an overview of the Census Bureau's coverage plans for the 2010 census, along with some suggestions for additional work. In the panel's final report, we hope to provide more direction on the following issues:

- What data to save in 2010 to support the various coverage measurement models, including the feasibility of saving some data on a 100 percent basis, and the possible need for sampling from the output of some management information systems.
- More work on the framework document looking into assumptions and estimation.
- Random effects modeling for small-area estimation.
- Variance estimation for synthetic estimation and related techniques.
- Treatment of non-data-defined cases in logistic regression.
- Allowable covariates in the logistic regression models (both in terms of balance issues for the E-sample and P-sample, and also due to Alho's concern about using covariates related to census processes).
- Sample design for the CCM survey, at both the state and substate levels.
- The products to use to inform about census component coverage error.
- Use of survey weights in logistic regression models.
- Improvements in demographic analysis in 2010
- How to exploit 2000 data more for 2010 design.
- Very generally, how to best operate a feedback loop for census improvement.
- What issues will come up in identifying duplicates in real time?

Finally, we also hope to consider the question of what a coverage measurement program entirely focused on measuring census component coverage error, including the use of administrative records, might look like.

References

Adams, T., and X. Liu (2001). Evaluation of Lack of Balance and Geographic Errors Affecting Person Estimates. Executive Steering Committee for A.C.E. Policy II, Report 2, Oct.11, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Alho, J.M. (1990). Logistic regression in capture-recapture models. *Biometrics* 46:623-635.

Alho, J.M. (1994). Analysis of sample based capture-recapture experiments. *Journal of Official Statistics* 10(3):245-256.

Alho, J.M., M.H. Mulry, K. Wurdeman, and J. Kim (1993). Estimating heterogeneity in the probabilities of enumeration for dual-systems estimation. *Journal of the American Statistical Association* 88(423):1130-1136.

Alho, J.M., and B.D. Spencer (2005). *Statistical Demography and Forecasting,* New York: Springer.

Auer, P.L. (2004). Study for Cases with Insufficient Information for Matching and Follow-up. Unpublished paper, Dec. 15, U.S. Census Bureau, Washington, DC.

Bates, N.A., and E. Gerber (1998). Temporary mobility and reporting of usual residence. *Survey Methodology* 24(1):89-98.

Bell, W.R. (1993). Using information from demographic analysis in post-enumeration survey estimation. *Journal of the American Statistical Association* 88(423):1106-1118.

Bell, W.R. (2001). Report 10: Estimation of Correlation Bias in 2000 A.C.E. Using Revised Demographic Analysis Results. Executive Steering Committee for A.C.E. Policy II, Oct. 16, Methodology and Standards Directorate, U.S. Census Bureau, Washington, DC.

Bell, W.R. (2005). Contamination Between Census Coverage Follow-Up and Coverage Measurement—Considerations for Implementing Dual-Systems Estimation. Unpublished draft, July 14, U.S. Census Bureau, Washington, DC.

Bench, K. (2002). *A Contamination of Census 2000 Data Collected in Accuracy and Coverage Evaluation Block Clusters.* Census 2000 Evaluation N.1, Aug. 22. Washington, DC: U.S. Census Bureau.

Brownrigg, L.A. and M. de la Puenta (1993). Alternative Enumeration Methods and Results: Resolution and Resolved Population by Site. 1990 Decennial Census Preliminary Research and Evaluation Memorandum No. 219, U.S. Census Bureau, Washington, DC.

Childers, D.R. (1992). The 1990 Housing Unit Coverage Study. In *Proceedings of the Section on Survey Research Methodology.* Washington, DC: American Statistical Association.

Coale, A.J. (1955). The population of the United States in 1950 classified by age, sex and color—A revision of census figures. *Journal of the American Statistical Association* 50(269):16-54.

Coale, A.J., and W. Norfleet Rives, Jr. (1972). A Statistical Reconstruction of the Black Population of the United States, 1880-1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility. Unpublished document prepared under a project agreement with the U.S. Census Bureau, Washington, DC.

Coale, A.J., and M. Zelnik (1963). *New Estimates of Fertility and Population in the United States*. Princeton, NJ: Princeton University Press.

Cohen, M.L. (2000). Coverage evaluation. Pp. 95-101 in M. J. Anderson (Ed.), *Encyclopedia of the U.S. Census*. Washington, DC: CQ Press.

M. de la Puenta (2004). Census 2000 Ethnographic Studies. Census 2000 Topic Report No. 15, Census Testing, Experimentation, and Evaluation Program, March, U.S. Census Bureau, Washington, DC.

Dillman, D.A., J.B. Treat, and J.R. Clark (1994). Influence of 13 design factors on completion rates to decennial census questionnaires. *Annual Research Conference and CASIC Technologies Interchange Proceedings*, March, Rosslyn, VA.

DuMouchel, W.H., and G.J. Duncan (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association* 78(383):535-543.

Ellis, Y. (1995). Examination of census omission and erroneous enumeration based on 1990 ethnographic studies of census coverage. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Ericksen, E., F. Leobardo Estrada, J.W. Tukey, and K.M. Wolter (1991). Report on the 1990 Decennial Census and the Post-Enumeration Survey. Manuscript prepared for the secretary of the U.S. Department of Commerce.

Fay, R.E. (2002). Probabilistic models for detecting census person duplication. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Fay, R.E. (2003). Probabilistic models for detecting census duplication at the person and household levels. In *Proceedings of the Section on Government Statisti*cs. Washington, DC: American Statistical Association.

Fay, R.E. (2004). An analysis of person duplication in census 2000. *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Fein, D.J. (1990). Racial and ethnic differences in U.S. census omission rates. *Demography* 27(2):285-302.

Feldpausch, R. (2001). Census Person Duplication and the Corresponding A.C.E. Enumeration Status. Executive Steering Committee for A.C.E. Policy II (Report 6), Oct. 31, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Fenstermaker, D.A. (2002). A.C.E. Revision II Results: Further Study of Person Duplication. A.C.E. Revision II Memorandum Series #PP-51. (Prepared by Thomas Mule.) Dec. 31, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Fenstermaker, D.A. (2005). 2010 Census Coverage Measurement: Provisional Sample Allocation. 2010 Census Coverage Measurement Memorandum Series #2010-C-01. (Signed by Peter P. Davis.) Oct. 26, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Fowlkes, E.B. (1987). Some diagnostics for binary logistic regression via smoothing. *Biometrika* 74(3):503-515.

Garcia-Parra, C. (1987). Causes for Undercoverage in East Los Angeles from an Anthropological Perspective. Report prepared under contract, April, U.S. Census Bureau, Washington, DC.

Griffin, R.A. (2000). Post-stratification classification affected by household coverage. In *Proceedings of the Survey Research Methods Section.* Washington, DC: American Statistical Association.

Griffin, R.A. (2005a). Initial Results of Preliminary Net Error Empirical Research Using Logistic Regression. 2010 Census Coverage Measurement Memorandum Series # 2010-E03. (Prepared by Thomas Mule and Douglas Olson.) April 18, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Griffin, R.A. (2005b). Preliminary Research on Separate Estimation of the Population in Enumerated and Non-Enumerated Housing Units. DSSD 2010 Census Coverage Measurement Memorandum Series # 2010-E06. (Prepared by Thomas Mule.) Nov. 29, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Habermann, S.J., W. Jiang, and B.D. Spencer (1998). Activity 7: Develop Methodology for Evaluating Model-Based Estimates of the Population Size for States. Paper prepared for the U.S. Census Bureau, Washington, DC.

Haines, D. (2002). A.C.E. Revision II Adjustment for Correlation Bias. DSSD A.C.E. Revision II Memorandum Series PP-53. (Prepared by Roger Shores.) Dec. 31, U.S. Census Bureau, Washington, DC.

Hawley, W. (2004). MAF/TIGER Accuracy Improvement Project—An Update. Presentation to WaTrans Partner’s Meeting, March 2, Washington, DC

Hill, J.M., and J.D. Machowski (2003). Census 2000 Evaluation B.6., Master Trace Sample. Sept. 29, Planning, Research, and Evaluation Division, U.S. Census Bureau, Washington, DC.

Huggins, R.M. (1989). On the statistical analysis of capture-recapture experiments. *Biometrika* 76:133-140.

Kostanich, D., and D. Whitford (2005). Discussion of the Proposal to Reduce Respondent Burden for Households in Both Census Coverage Followup and Census Coverage Measurement. 2010 Census Coverage Measurement Memorandum Series # A-07. Nov. 29, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Little, R. (2003). To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. University of Michigan Department of Biostatistics Working Paper Series, Working Paper #4.

Malec, D., and J. Maples (2005). An evaluation of synthetic, small-area census coverage error using a random effects model. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Marks, E.S., W. Seltzer, and K.J. Krotki (1974). *Population Growth Estimation*. New York: The Population Council.

Martin, E.A. (1999). Who knows who lives here? Within-household disagreements as a source of survey coverage error. *Public Opinion Quarterly* 63(2):220-237.

Martin, E.A., L.A. Brownrigg, and R.E. Fay (1990). Results of the 1988 ethnographic studies of census coverage and plans for 1990. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Moriarity, C., and D. Childers (1993). Analysis of census omissions: Preliminary results. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Mule, T. (2001). ESCAP II Person Duplication in Census 2000. Executive Steering Committee for A.C.E. Policy II, Report 20, Oct. 11, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Mule, T. (2003). *A.C.E. Revision II Results: Change in Estimated Net Undercount.* Decennial Statistical Studies Division, A.C.E. Revision II Memorandum Series PP-58. Washington, DC: U.S. Census Bureau.

Mulry, M.H. (2002). A.C.E. Revision II Missing Data Evaluation Final Report (through David L. Hubble, prepared by Anne T. Kearney). DSSD A.C.E. Revision II Memorandum Series #PP-48. Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Mulry, M.H., E. Schindler, T. Mule, N. Nguyen, and B.D. Spencer (2005). Investigation of extreme estimates of census coverage error for small areas. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

National Research Council. (1985). *The Bicentennial Census: New Directions for Methodology in 1990.* Panel on Decennial Census Methodology. C.F. Citro and M. L. Cohen, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2004a). *Reengineering the 2010 Census: Risks and Challenges*. Panel on Research on Future Census Methods. D.L. Cork, M.L. Cohen, and B.F. King, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2004b). *The 2000 Census: Counting Under Adversity.* Panel to Review the 2000 Census. C.F. Citro, D.L. Cork, and J.L. Norwood, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2006). *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census*. Panel on Residence Rules in the Decennial Census. D.L. Cork and P.R. Voss, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Price, D.O. (1947). A check on underenumeration in the 1940 census. *American Sociological Review* 12(1):44-49.

Resnick, D., S. Love, C. Taeuber, and J.M.Staveley (2004). Analysis of ACS food stamp program participation underestimate. In *Proceedings of the American Statistical Association*, *Government Statistics Section*. Washington, DC: American Statistical Association.

Robinson, J.G. (2001). Demographic Analysis Results. Executive Steering Committee for A.C.E. Policy II, Report No. 1, Oct.13, U.S. Census Bureau, Washington, DC.

Robinson, J.G., B. Ahmed, P. Das Gupta, and K.A. Woodrow (1993). Estimation of population coverage in the 1990 United States census based on demographic analysis. *Journal of the American Statistical Association*, 88(423):1061-1071.

Ruhnke, M.C. (2003). *An Assessment of Addresses on the Master Address File "Missing" in the Census or Geocoded to the Wrong Collection Block*. Census 2000 Evaluation F.15. Washington, DC: U.S. Census Bureau.

Schindler, E. (2006). 2010 Census coverage measurement: The Hunt for the magic variables. In *Proceedings of the Section on Survey Research Methods.* Washington, DC: American Statistical Association.

Sekar, C.C., and W.E. Deming (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* 44:101-115.

Siegel, J.S., and Zelnik, M. (1966). An evaluation of coverage in the 1960 census of population by techniques of demographic analysis and by composite methods. Pp. 71-85 in *Proceedings of the Social Statistics Section of the American Statistical Association*. Washington, DC: American Statistical Association.

Shoemaker, H.H. (2005). Results of a Feasibility Study to Match Census Enumerations Coded in A.C.E. as Insufficient for Matching and Follow-up. 2010 Census Coverage Measurement Memorandum Series #2010-B01, April 18, Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC.

Spencer, B.D. (2005). Statistical Analysis Research on Small Area Estimation of Census Coverage. Prepared by Abt Associates Inc. and Spencer Statistics, Inc. for the Bureau of the Census, Activity 1—Deliverable 6, March 22.

Stuart, E.A., and A.M. Zaslavsky (2002). Using administrative records to predict census day residency. Pp. 335-349 in *Case Studies in Bayesian Statistics,* Vol. VI, edited by C. Gatsonis, R.E. Kass, A Carriquiry, A. Gelman, D. Higdon, D.K Pauler, and I. Verdinelli. New York: Springer.

Trussell, J. (1981). Should state and local area census counts be adjusted? *Population Index* 47(1):4-12.

U.S. Census Bureau (2003). Technical Assessment of A.C.E. Revision II. Washington, DC.

U.S. Census Bureau. (2004). The 2006 Census Coverage Measurement Test. Unpublished draft, April 7.

U.S. Census Bureau (2005), Framework for Coverage Error Components. Unpublished draft, May 26.

Valentine, C.B., and B.L. Valentine (1971). Missing Men: A Comparative Methodological Study of Underenumeration and Related Problems. Report prepared under Joint Statistical Agreement with the Brookdale Hospital Center, Washington, DC, and the U.S. Census Bureau.

Biographical Sketches of Panel Members and Staff

Robert M. Bell (*Chair*) is a member of the Statistics Research Department at AT&T Labs-Research in Florham Park, New Jersey. He is a member of the National Research Council's (NRC) Committee on National Statistics and previously served on multiple NRC panels advising the Census Bureau, including the Panel to Review the 2000 Census, and on the American Statistical Association's Census Advisory Committee. He is currently a member of the board of trustees of the National Institute of Statistical Sciences and chair of the American Statistical Association's Committee on Fellows. Earlier, he worked at the RAND Corporation, where he directed a number of studies on social policy issues. His research interests include analysis of data from complex samples, record linkage methods, and machine learning methods. He is a fellow of the American Statistical Association. He has a B.S. in mathematics from Harvey Mudd College, an M.S. in statistics from the University of Chicago, and a Ph.D. in statistics from Stanford University.

Barbara A. Bailar is recently retired from the position of senior vice president for survey research at the National Opinion Research Center (NORC). She now serves as a consultant on survey methodology. Prior to joining NORC in 1995, Dr. Bailar was the executive director of the American Statistical Association in Alexandria, Virginia. However, most of her career was spent at the U.S. Bureau of the Census in Washington, D.C., which she left in 1988, where she was the associate director for Statistical Standards and Methodology. She has published numerous articles in such journals as the *Journal of the American Statistical Association, Demography*, and *Survey Research Methods*. She is a past president of the American Statistical Association and the International Association of Survey Statisticians, as well as a past vice president of the International Statistical Association. She is an elected fellow of the American Statistical Association and the American Association for the Advancement of Science. She received a Ph.D. in statistics from American University in Washington, D.C.

Lawrence D. Brown is the Miers Busch professor in the Department of Statistics at the Wharton School at the University of Pennsylvania. He is a member of the National Academy of Sciences. He has served on the NRC's Committee on National Statistics and its Panel to Review the 2000 Census, as well as the Commission on Physical Sciences, Mathematics, and Applications and its Board on Mathematical Sciences. He chaired the NRC Committee on the Measurement of Research and Development by the National Science Foundation. He is currently on the U.S. National Committee on Mathematics. He is a fellow of the American Statistical Association and recipient of its Wilks Memorial Award and is a fellow and past president of the Institute of Mathematical Statistics. He has a B.S. from the California Institute of Technology and a Ph.D. from Cornell University.

Michael L. Cohen (*Study Director*) is a senior program officer for the Committee on National Statistics, currently serving as study director for the Panel on the Functionality and Usability of Data from the American Community Survey. Formerly, he was a mathematical statistician at the Energy Information Administration, an assistant professor in the School of Public Affairs at the University of Maryland, and a visiting lecturer in statistics at Princeton University. His general area of research is the use of statistics in public policy, with particular interest in census undercount, model validation, and robust estimation. He is a fellow of the American Statistical Association. He has a B.S. degree in mathematics from the University of Michigan and M.S. and Ph.D. degrees in statistics from Stanford University.

Roderick Little is the Richard Remington collegiate professor of biostatistics at the University of Michigan, where he also holds appointments in the Department of Statistics and the Institute for Social Research. Previously he was associate professor and professor in the Department of Biomathematics in the School of Medicine at the University of California at Los Angeles, an American Statistical Association/Census/National Science Foundation research fellow at the Census Bureau, and a scientific associate at the World Fertility Survey. A national associate of the National Academy of Sciences, he has served on a variety of NRC panels, including the Committee on Research on Child Abuse and Neglect, the Committee on Evaluation of National and State Assessments of Educational Progress, the Committee on National Statistics, the Committee on Research on Child Abuse and Neglect, and the Panel to Review the 2000 Census. He has served as coordinating and applications editor of the *Journal of the American Statistical Association*, is a fellow of the American Statistical Association, and received its Wilks Memorial Award. His research interests include statistical methods for missing data and survey research methodology. He has a B.A. in mathematics from Cambridge University and M.Sc. and Ph.D. degrees in statistics from the Imperial College of Science and Technology, London University.

Xiao-Li Meng is chair of the Department of Statistics at Harvard University. Previously, he was assistant and associate and professor in the same department. He has also served as professor of statistics at the University of Chicago and faculty research associate at the National Opinion Research Center. His expertise focuses on Bayesian methods and methods for missing data. He has served as editor of *Bayesian Analysis* and as an associate editor of *Biometrika*, the *Journal of the American Statistical Association*, *Statistica Sinica*, and the *Annals of Statistics*. He received the 2001 Committee of Presidents of Statistical Associations Award for outstanding statistician under the age of 40. He is a fellow of the Institute of Mathematical Statistics and of the American Statistical Association. He has a B.S. in mathematics from Fudan University, Shanghai, and M.A. (1987) and Ph.D. (1990) degrees in statistics from Harvard University.

Jeffrey Passel is a senior research associate at the Pew Hispanic Center. His expertise focuses on immigration to the United States and the demography of racial and ethnic groups. Formerly he served as principal research associate at the Urban Institute's Labor, Human Services and Population Center. Prior to that, he was a demographer at the U.S. Census Bureau, serving last as assistant division chief for population estimates. He was a member of the NRC's Panel on Estimates of Poverty for Small Geographic Areas. He is a fellow of the American Association for the Advancement of Science and of the American Statistical Association. He received the Demographic Diamond as one of the five demographers/social scientists selected by *American Demographics* as the most representative of influential work in the last 25 years, and he received the Bronze Medal from the U.S. Department of Commerce. Passel has authored numerous studies on immigrant populations in America, focusing on such topics as undocumented immigration, the economic and fiscal impact of the foreign born, and the impact of welfare reform on immigrant populations. He has a B.S. in mathematics from the Massachusetts Institute of Technology (1969), an M.A. in sociology from the University of Texas (1972), and a Ph.D. in social relations from Johns Hopkins University (1976).

Donald Ylvisaker is emeritus professor of statistics at the University of California, Los Angeles, having previously been on the faculties of Columbia University, New York University, and the University of Washington. His primary research interest is in the design of experiments; his applied interests have developed as a consulting statistician, frequently on legal matters. Relevant to census methods, he held a joint statistical agreement with the Census Bureau from 1990 to 1991 in which he reviewed the Census Bureau's 1986 Test of Adjustment Related Operations in Los Angeles. He also served on the Advisory Panel to the Committee on Adjustment of Postcensal Estimates in 1992, and he was an associate editor of a special issue of the *Journal of the American Statistical Association* on census methods. He was a member of the NRC's Panel on Research on Future Census Methods. A fellow of the American Statistical Association and of the Institute of Mathematical Statistics, he has a Ph.D. in statistics from Stanford University.

Alan Zaslavsky is professor in the Department of Health Care Policy at Harvard Medical School. He has written extensively on issues concerning the decennial census, including weighting and administrative records. A fellow of the American Statistical Association, he has served on three NRC Committee on National Statistics panels involving decennial census methodology: the Panel to Evaluate Alternative Census Methodologies, the Panel on Alternative Census Methodologies, and the Panel on Research on Future Census Methods. He also served on the NRC's Panel on Estimates of Poverty for Small Geographic Areas, as well as the Census Advisory Committee on Adjustment of Postcensal Estimates. He has an M.S. in mathematics from Northeastern University and a Ph.D. in applied mathematics from the Massachusetts Institute of Technology.