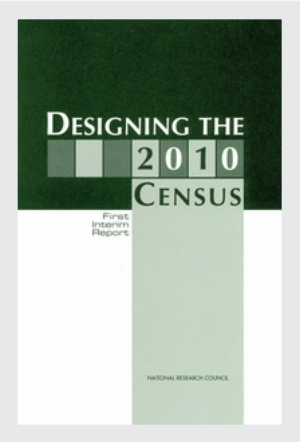


# Designing the 2010 Census: First Interim Report



## DETAILS

52 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-06944-1 | DOI 10.17226/9802

## AUTHORS

Michael L. Cohen, Benjamin F. King, Panel on Research on Future Census Methods, Committee on National Statistics, National Research Council

# DESIGNING THE 2010 CENSUS

## First Interim Report

Panel on Research on Future Census Methods

Michael L. Cohen and Benjamin F. King, *Editors*

Committee on National Statistics

Commission on Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

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 is supported by funds provided by the Bureau of the Census, U. S. Department of Commerce, under contract number 50-YABC-8-66016. 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-9709489). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number 0-309-06944-0

*Additional copies of this report are available from*

National Academy Press
2101 Constitution Avenue, NW
Washington, DC 20418

Call 800-624-6242 or 202-334-3313 (in the Washington Metropolitan Area).
This report is also available on line at http://www.nap.edu

Suggested citation: National Research Council (2000). *Redesigning the 2010 Census: First Interim Report.* Panel on Research on Future Census Methods, Committee on National Statistics. Washingotn, D.C.: National Academy Press.

Printed in the United States of America

# THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

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. Bruce M. Alberts 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. William 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. Kenneth I. Shine 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. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

**PANEL ON RESEARCH ON FUTURE CENSUS METHODS**

BENJAMIN F. KING (*Chair*), Boca Raton, FL
DAVID A. BINDER, Statistics Canada, Ottawa, Ontario
MICK P. COUPER, Survey Research Center, University of Michigan
WILLIAM D. KALSBEEK, Department of Biostatistics, University of North Carolina
SALLIE KELLER-MCNULTY, Los Alamos National Laboratory, Los Alamos, NM
MICHAEL M. MEYER, Amazon.com, Seattle, WA
DARYL PREGIBON, AT&T Laboratories-Research, Florham Park, NJ
KEITH F. RUST, Westat, Inc., Rockville, MD
JOSEPH J. SALVO, New York City Department of City Planning, NY
JOSEPH L. SCHAFER, Department of Statistics, Pennsylvania State University
ALLEN L. SCHIRM, Mathematica Policy Research, Inc., Washington, DC
JOSEPH SEDRANSK, Department of Statistics, Case Western Reserve University, Cleveland
DONALD YLVISAKER, Department of Statistics, University of California, Los Angeles

MICHAEL L. COHEN, *Study Director*
AGNES E. GASKIN, *Senior Project Assistant*

**COMMITTEE ON NATIONAL STATISTICS**
**1999-2000**

JOHN E. ROLPH (*Chair*), Marshall School of Business, University of Southern California
JOSEPH G. ALTONJI, Department of Economics, Northwestern University
LAWRENCE D. BROWN, Department of Statistics,Wharton School, University of Pennsylvania
JULIE DAVANZO, RAND, Santa Monica, CA
WILLIAM F. EDDY, Department of Statistics, Carnegie Mellon University
HERMANN HABERMANN, Statistics Division, United Nations
WILLIAM KALSBEEK, Department of Biostatistics, University of North Carolina
RODERICK J.A. LITTLE, School of Public Health, University of Michigan
THOMAS A. LOUIS, Division of Biostatistics, University of Minnesota School of Public Health
CHARLES F. MANSKI, Department of Economics, Northwestern University
WILLIAM NORDHAUS, Department of Economics, Yale University
EDWARD B. PERRIN, Department of Health Services, School of Public Health and Community Medicine, University of Washington
FRANCISCO J. SAMANIEGO, Division of Statistics, University of California, Davis
RICHARD L. SCHMALENSEE, Sloan School of Management, Massachusetts Institute of Technology
MATTHEW D. SHAPIRO, Department of Economics and Survey Research Center, University of Michigan

ANDREW A. WHITE, *Director*

# Acknowledgments

The Panel on Research on Future Census Methods thanks the many people who contributed to the preparation of this, the panel's first interim report. First, we thank the staff of our sponsor, the Bureau of the Census, who provided detailed and informative presentations on the wide variety of aspects of the census process relevant to the panel's charge. We are especially indebted to Kenneth Prewitt, Cynthia Clark, Ruth Ann Killion, Jay Keller, Juanita Tamayo Lott, and their entire staff, who were always available to the panel members and staff to answer our many questions. We also thank other Census Bureau staff who played a role in the preparation or presentation of material to the panel: Charles Alexander, Theresa Anguiera, Debbie Bolton, Geraldine Burt, David Hubble, Charlene Leggieri, Joe Knott, Randall Neugebauer, Cleo Redline, and Henry Woltman.

The panel also thanks John Czajka for his presentation on potential uses of administrative records in the 2010 census and Alan Zaslavsky for his presentation on sampling, data collection, and estimation in future censuses.

Eugenia Grohman, associate director for reports of the Commission on Behavior and Social Sciences and Education, enormously improved the report through her technical editing.

This report has been reviewed in draft form by individuals chosen for their diverse perspective and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. 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 wish to thank the following individuals for their participation in the review process of this interim report: Margo Anderson, Department of History, University of Wisconsin, Milwaukee; Stephen Fienberg, Department of Statistics, Carnegie Mellon University; Hermann Habermann, Statistics Division, United Nations; Albert Madansky, Graduate School of Business, University of Chicago; D. Bruce Petrie, Canadian Institute for Health Information, Ottawa, Ontario; Samuel H. Preston, School of Arts and Sciences, University of Pennsylvania; James Trussell, Woodrow Wilson School of Public and International Affairs and Office of Population Research, Princeton University; and James Walker, Department of Economics, University of Wisconsin, Madison.

Although the individuals listed above have provided constructive comments and suggestions, it must be emphasized that responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Benjamin F. King, *Chair*
Panel on Research on
Future Census Methods

# Contents

# Executive Summary

The Panel on Research on Future Census Methods was formed to examine alternative designs for the 2010 census and to assist the Census Bureau in planning tests and analyses to help assess and compare the advantages and disadvantages of them. Toward this goal, the panel was asked, in its first interim report, to examine whether the auxiliary information that is planned to be collected (and retained) during the 2000 census could be augmented to help guide the Census Bureau in its assessment of alternative designs for the 2010 census.

## MODIFYING THE MASTER TRACE SAMPLE

The planned master trace sample database will collect information from a variety of auxiliary census data systems to provide comprehensive information for a sample of housing units, showing how each was processed through every step of the 2000 census. We believe that the master trace sample database has the potential to be the single most useful source of information for assessing alternative designs for the 2010 census. This is due to its integration, at the individual household level, of information from all key census processes, so that all their interactions can be examined.

The current (preliminary) plans for the master trace sample are limited in two ways: first, by viewing it only as the result of census operations applied to a sample of census addresses although some operations are carried out prior to the assignment of an address; second by not explicitly mentioning some census processes that would be important to represent.

Relevant processes that may not be currently included in plans for the master trace sample are information on "Be Counted" forms that cannot be geocoded, information on the workings of the primary selection algorithm, and processing information on the Accuracy and Coverage Evaluation (ACE) program. Augmenting current plans to include representation of additional census processes will permit better understanding of the advantages and disadvantages of alternative designs for the 2010 census.

**Recommendation: The current plans for the master trace sample database should be augmented so that data for all key steps in the process—starting with address assignment and ending with a final disposition for each case—are included in the master trace sample data base.**

## SETTING PRIORITIES FOR EVALUATION STUDIES

Many of the evaluation plans described to the panel ended with data collection and provided little detail about how the data would be used. The Census Bureau should develop these plans further to include how the data will be analyzed; how the results obtained will inform decisions about the 2010 census design; and what resources, in terms of data collection costs and staff expertise, are required. Understanding the costs and technical resources required and the potential information value of each proposed evaluation study will help in setting priorities and allocating resources to the studies being proposed. The Census Bureau is planning a large number of evaluation studies, and there may be insufficient staff and expertise to adequately use the data being collected. For setting priorities, the objectives should be both to assess 2000 census operations and to provide information to support 2010 census design decisions. The results of this process will allow the the Census Bureau to make informed decisions about which are the most important and the least resource-intensive evaluation activities following the 2000 census.

**Recommendation: The Census Bureau should develop a detailed plan for each evaluation study on how to analyze the data collected and how to use the results in decision making concerning 2010 census design. The Census Bureau should then use these plans to identify the benefits and resources required for each evaluation study, set priorities among them, and allocate sufficient resources for the careful completion of all or, at least, the highest priority evaluations.**

## MODIFYING THE DESIGN OF AREX 2000

The Administrative Records Census Experiment in 2000 (AREX 2000) will test, in a limited set of sites, the use of administrative records as a supplemental method of data collection. The overall goal of AREX 2000 is to determine whether a merged list of administrative records can achieve sufficient coverage to justify using administrative records in the 2010 census for: (1) whole household nonresponse follow-up, (2) reduction of differential undercoverage, (3) imputation for item nonresponse, (4) address list improvement, and (5) the support of special procedures to target individual groups. However, the current design of AREX 2000 will not sufficiently advance the current state of knowledge as to the feasibility and advantages of the use of a merged list of administrative records for the purposes listed above. The currently planned comparison of aggregate counts could miss information on the kinds of individuals or households that are omitted by either the census or the merged AREX 2000 list.

The current plans should be modified to include a match of census households with the merged list. The field follow-up of unresolved matches could be restricted to a sample of cases to control costs. In ACE block clusters, one could also use matching to determine whether individuals that are enumerated by the ACE survey and missed by the census appear on administrative records lists, which would help to assess whether administrative records could be used to reduce census undercoverage.

**Recommendation: The Census Bureau should modify the plans for AREX 2000 to include a match of the merged administrative records list with the 2000 census households. Costs of this new activity could be controlled through use of field follow-up of unresolved matches on a sample basis.**

## MEASURING THE ACTIVITIES OF LOCAL COMMUNITY GROUPS

For the 2000 census, the Census Bureau has increased greatly the participation by various community organizations, which is expected to increase the mail return rate and overall responsiveness to the 2000 census. The result will be lower costs and, possibly, improved data accuracy by reducing the frequency of proxy response and amount of differential undercoverage.

Community organizations will vary in their degree of effort and the methods used.

Some groups might make greater use of "Be Counted" forms, speeches at public gatherings, or door-to-door interaction, but such variation can complicate evaluation.

Therefore, to the extent possible, the Census Bureau should examine ways of collecting information about the activities of these groups. This information could come from activity diaries after the 2000 census is concluded, local census office monitoring of the activities of local groups, or interviews with the staffs of local groups. Without such information, analysis of the 2000 census will be confounded, possibly to an important degree, by the unknown activities of these groups.

**Recommendation: The Census Bureau should examine various methods for collecting information on the activities of local organizations that are intended to produce a more complete census count.**

In addition to these four recommendations, the panel proposes a number of other changes for the Census Bureau, which we summarize here:

For the master trace sample data base:

- use a two-stage sample design;
- oversample ACE blocks, list/enumerate and update/leave households, and households in areas that are hard to enumerate;
- improve the quality of the information on the number and dates of attempts at enumeration;
- set priorities for the retention of master trace sample input files;
- provide for the accessibility and availability of the database;
- increase the resources for developing the database; and
- collect sufficient information to support a model of total census error.

For AREX 2000:

- examine whether administrative records could be used as the primary stage of nonresponse follow-up for clearly matched households, leaving field follow-up for the remaining households;

- obtain approval for acquisition of food stamp and Medicaid files for the AREX 2000 test area;
- begin legal and legislative processes for greater access to administrative records;
- increase the resources devoted to carrying out this experiment;
- examine how to improve race and ethnicity data available in administrative records; and
- assess the accuracy of the matching and unduplication used in this experiment.

With respect to auxiliary census data systems that provide information on census processes:

- save the census visual images for at least the households in the master trace sample database; and
- reevaluate the decision to use custom software for the various data systems that monitor and control various aspects of census processes.

For the 2000 census:

- expand the Content Reinterview Survey to measure response bias;
- evaluate the quality of imputation and editing;
- collect comprehensive cost data on all components of census processes;
- collect comprehensive information on the accuracy of ACE matching; and
- continue systematic observation.

Finally, while the 2010 census is well in the future, simple steps taken now could have a very big payoff later (as is the case for many of the above proposals). For the 2010 census and the American Community Survey (ACS) we propose:

- use the ACS as a census testing platform;
- conduct a census short form-ACS match study;
- improve the address list in several ways;
- determine the feasibility in 2010 of use of a targeted replacement questionnaire;

- continue work on legislative and legal initiatives concerning use of sampling for nonresponse follow-up, modifications to administrative records systems, and developing greater access to administrative records;
- form an ACS advisory group; and
- refine development of a model linking responses to the questions on race and ethnicity used in the 1990 and 2000 censuses.

We recognize that implementation of the panel's recommendations and proposals must take place in the brief amount of time available before the 2000 census is in high gear. Although we have indicated some priorities, we understand that it will almost certainly be necessary for the Census Bureau to further set priorities among the recommended and proposed changes.

# Introduction

## BACKGROUND

The 2000 census will provide a great deal of auxiliary information (i.e., information in addition to that requested on census forms) on how well various 2000 census operations were carried out. This information is relevant to the task of our committee, namely, assessing how modifications to the design of the 2000 census might work in 2010. This auxiliary information comes from several sources. First, in monitoring the 2000 census operations, the Census Bureau will be collecting information, often on a temporary basis, as part of several data systems that govern various census processes. For example, one data system, the Operational Control System 2000 (OCS 2000), will be monitoring nonresponse follow-up. (Other data systems are described below.)

Second, the Census Bureau is also obtaining useful information for planning the 2010 census from three additional sources:

- evaluation studies—All major systems used in the 2000 census will have associated evaluation studies to determine how successfully they were carried out. These are often based on quality control systems put in place for this purpose. Others have suggested that the use of decennial census administrative records, especially cost data that result from various internal audits that are conducted, may be useful as an evaluation tool. The panel has not explored this matter in any depth to date.

- experiments—For each recent decennial census, the Census Bureau has used the unique environment created by the census to conduct several experiments. These experiments are used to examine the feasibility and value of a potential change in the next census.
- a master trace sample—The Census Bureau is planning on collecting a master trace sample that will provide comprehensive information for a sample of housing units as to how each was processed through each step of the 2000 census.

## THE PANEL'S CHARGE

The Panel on Research on Future Census Methods was convened to assist the Census Bureau in selecting and evaluating alternative designs for the 2010 census, focusing on plans for the census tests and for research and analysis. The panel was convened in 1999 to help the Census Bureau determine whether the auxiliary data planned to be collected in the 2000 census are adequate to support the test plans and research and analysis.

Because many of the plans concerning census data systems and experiments were nearly final by the panel's first meeting in June 1999, the panel focused on minor modifications for the census data systems and the census experiments. Since the plans for the master trace sample were not final, the panel was able to consider a range of possible improvements. Finally, the panel received only limited information on the evaluation studies to be used for assessing the quality of the 2000 census and so could not assess how they might be modified to provide more information for evaluating alternative designs for the 2010 census.

The panel's charge for this first interim report required predicting, to the best of its ability in late 1999, the alternative designs that will be the most promising candidates for the 2010 census. (Given the testing cycle, this means predicting the primary alternatives for the 2010 census as viewed in 2004.) To make this prediction, one must forecast, among other things: the degree of responsiveness of the U.S. residents to mail questionnaires from the government, with or without an advertising campaign; the concerns of the U.S. residents about privacy and confidentiality; the number of persons expected to have affiliations with more than one household; the quality, content, and availability of administrative records; the degree of general access to the Internet; and the effect of other technological innovations. Each of these dynamic factors—and undoubtedly others—that are

extremely difficult to predict will influence the effectiveness of different designs for the census in 2010.

Lott and Keller (1999) have provided a list of important considerations relevant to the ultimate design of the 2010 census. Using this and other information,[1] the panel believes that alternative designs for the 2010 census can be usefully distinguished on four dimensions:

- the degree to which sampling will be used in the 2010 census, including a postenumeration survey, sampling for coverage improvement,[2] and possibly sampling for nonresponse follow-up;
- the degree to which administrative records will be used, for master address list improvement, for help with specific groups, for coverage improvement, for coverage measurement, as sources of census data (including long-form information), and for imputation for item and unit nonresponse;
- the degree to which greater use of technology could benefit either data acquisition or data dissemination; and
- interaction with the American Community Survey (ACS), which includes both improvements to the master address file and eliminating the need for a census long form.

Although these dimensions are useful as a starting point, the panel recognizes that they do not fully describe all the promising alternative designs for the 2010 census.

## THE PANEL'S ACTIVITIES TO DATE

In the time available, the panel has been able to schedule two full panel meetings, for which the Census Bureau provided presentations on future census designs, the American Community Survey, census 2000 address list

[1]Three presentations at the panel's first meeting were especially helpful: the future of sampling in the census, by Alan Zaslavsky; the future use of administrative records in the census, by John Czajka; and anticipated advances in technology relevant to the census, by panel members Daryl Pregibon and Michael Meyer.

[2]Coverage improvement programs are census programs that attempt to address potential omissions in the census through additional efforts to include specific types of housing units or specific types of individuals, such as additional follow-up of housing units initially identified as vacant and the parolee/probationer programs in 1990.

formation, census 2000 experimentation, census 2000 evaluation plans, plans for the master trace sample, and census 2000 auxiliary data systems. In addition, a working group of the panel has met with Census Bureau staff to discuss the initial plans for the master trace sample and statistical aspects of the plans for the experiments to be carried out during the 2000 census. Finally, a working group has participated in two conference calls with Census Bureau staff concerning the auxiliary data systems for the 2000 census.

In light of the relatively recent convening of our panel, we commend the Census Bureau for its conscientious efforts in assisting the panel in its work by providing excellent background material and technical presentations that have brought the panel up to speed. In addition, we commend the Bureau for getting such an early start on planning for the 2010 census, though, as mentioned above, the lack of detail on the specific evaluation studies that was provided limited the panel in its ability to review that important component of the auxiliary data collected in 2000. Realizing that the briefings of our panel drew resources away from other important census activities as April 2000 fast approaches, we are very grateful for this assistance. We look forward to future panel meetings and to the opportunity to carry out our charge as the results of the 2000 census are analyzed and planning for 2010 moves forward.

## STRUCTURE OF THE REPORT

The panel's activities and deliberations have resulted in this report, with analysis and recommendations to the Census Bureau. The next three sections cover the key topics noted above: the master trace sample; the 2000 census research experiments, especially the experiment on administrative records; and the 2000 census data systems. In addition, although the main charge to the panel for this report focused on the plans for auxiliary data collection for the 2000 census, the broader charge is to examine alternative designs for the 2010 census, and consistent with the broader charge, the final two sections cover other 2000 census issues and planning for the 2010 census and the ACS survey.

# The Master Trace Sample

## BASIC STRUCTURE

The master trace sample is a probability sample of addresses that will be selected during the various stages of formation of the final 2000 master address file. The plan is to keep all of the relevant information on census processes for each address and its occupants in the sample by retaining the full history of values for each data field as the census progresses, in addition to other information related to field enumeration. Thus the information will include both the data collected from the respondents and data related to census operations, such as number of follow-up attempts, thereby showing how the various stages of data collection, processing, and treatment work. The resulting database is intended to be used to examine a wide variety of questions concerning census operations, including evaluation of potential alternatives to current census procedures for use in planning for 2010.[3]

The master trace sample will have two main components: a 0.5 percent systematic sample of addresses and all addresses from a sample (of unknown rate) of block clusters in, and not in, the Accuracy and Coverage Evaluation (ACE) program.[4] The sample of block clusters will facilitate

[3]The current plans for the master trace sample are not fully documented; therefore, it is possible that some of the panel's suggestions may already be included in the design.

[4]The ACE survey is the 2000 census version of the postenumeration survey used in the 1990 census.

analysis of effects that occur at a local level, such as those deriving from the local administration of the census (mainly nonresponse follow-up, coverage improvement programs, and other local field work, that occur at a local level). The 0.5 percent systematic sample of 600,000 addresses will contain roughly 500,000 housing units that receive the short-form questionnaire and 100,000 housing units receiving the long-form questionnaire (ignoring "Be Counted" forms, which are all short-form questionnaires). Of the 100,000 long forms, 25,000 will be subsampled for inclusion in the Content Reinterview Survey (a sample of housing units, not in ACE block clusters, originally receiving the long form, that are asked to fill out the long form a second time to measure response variance). Of the systematic sample, 0.3 percent (1,800 addresses) are expected to fall in ACE block clusters (0.3 percent being the overall ACE sampling rate).

In the current plans (see Bureau of the Census, 1999), the master trace sample selection has five separate samples: a sample of addresses in the mailback universe on the Decennial Master Address File (DMAF); a sample of addresses from update/leave areas; a sample of addresses from list/enumerate areas; a sample of addresses discovered during various field operations, especially nonresponse follow-up; and a sample of addresses generated from other sources, such as "Be Counted" forms.

The information collected for these addresses will include input from five sources: files used to track and control census operations; files generated as a result of data capture and processing; files generated as a result of final editing and allocation (filling in of missing responses) of the response data; files from Accuracy and Coverage Evaluation responses; and data from the Content Reinterview Sample. A sample of questionnaires will be double-keyed, with differences reconciled, to measure errors due to the data capture process.

The information collected in the master trace sample will be linked to two additional databases that will provide information on field operations. The master trace sample database will make use of enumerator-level data extracts from the PAMS/ADAMS[5] and OCS 2000 databases; the former manages the compensation of enumerators and other field staff; the latter monitors the completeness of field enumeration at the level of the local

[5]PAMS is the preappointment management system; ADAMS is the automated decennial administrative management system.

census office. The PAMS/ADAMS database is planned to be linked to the master trace sample database through use of the enumerator's social security number. This link will facilitate research on the cost and quality of the nonresponse follow-up operation at various stages (initial, closeout, final attempt) and may show whether there is any relationship between enumerator characteristics (e.g., enumerator productivity or experience, wage rate, local turnover rate) and data accuracy. Another link to an administrative records database will permit comparisons of census data with administrative records data for the master trace sample addresses; in this role, this administrative records database will be considered a component of the master trace sample database.

The Panel on Decennial Census Methodology (National Research Council, 1988) recommended that the Census Bureau collect a master trace sample in conjunction with the 1990 census. The Panel on Alternative Census Methodologies (National Research Council, 1999) repeated this recommendation for application to the 2000 census. We believe that the master trace sample database has the potential to be the single most useful source of information for assessing alternative designs for the 2010 census. This is due to its integration, at the individual household level, of information from all key census processes, so that the interactions of various factors can be examined. Therefore, we strongly endorse the recommendation of the Panel on Alternative Census Methodologies for the 2000 census, and we note that the current design is much broader than described by the previous panels, especially with regard to the plans to incorporate information on enumerators and from administrative records. Since even more limited versions of a master trace sample are difficult to implement, the Census Bureau deserves strong praise for its efforts toward making this version of the master trace sample a reality. In addition, the planned sample size will greatly facilitate detailed analysis of factors underlying census accuracy. Because of its breadth, it would be useful if the master trace sample could be overrepresented in various postcensus studies, as is being done with the Content Reinterview Survey.

However, some relevant processes of the 2000 census may not be covered by the current plans for a master trace sample database, and these processes could be represented through use of relatively modest additions to the current plans.

**Recommendation: The current plans for the master trace sample database should be augmented so that data for all key steps in the pro-**

**cess—starting with address assignment and ending with a final disposition for each case—are included in the master trace sample data base.**

In particular, the panel recommends that the master trace sample be modified to include information on "Be Counted" forms that cannot be geocoded, on the individual components of the primary selection algorithm,[6] on specific aspects of ACE operations and processing (e.g., the effects of imputation routines used), and any other census processes that are not currently represented, possibly through not being linked to an address on the DMAF. Some of these other processes that may not be represented include coverage improvement activities and components of telephone questionnaire assistance.

"Be Counted" forms (census questionnaires that are widely distributed in public places) must have an address that can be located in census geography (i.e., geocoded) or they are rejected and are not included in the census. It would be useful to be able to assess the process of determining which addresses can and cannot be geocoded and the types of households or individuals that are on forms that cannot be geocoded. The primary selection algorithm is used to determine which information relative to a census address is the information to be used for that address when an address has multiple submissions, and it is also used to identify duplicate entries. While the results of the primary selection algorithm are included in the plans for the master trace sample, it is not clear that the operations of the components of this algorithm will be able to be examined so that improvements to the algorithm can be identified. Finally, by ACE operations, we mean the initial ACE interview, the ACE follow-up interview, various stages of computer and clerical matching, and imputations for nonresponse and unresolved matches.

## OTHER MODIFICATIONS

In addition to its major recommendation for augmenting the master trace sample, the panel proposes several other modifications to the plans for the sample.

[6]The results of the primary selection algorithm are included in the master trace sample through the differences between the decennial response file 1 (DRF1), which contains every response record received, and the decennial response file 2 (DRF2), which contain addresses with assigned identification numbers and links all data records into one "form."

### Use of a Two-Stage Sample Design

Methods of statistical inference that make full use of two sampling plans, one a systematic sample at the level of the individual household and one a sample of block clusters, are not straightforward. Analyses that would make full use of both samples would be greatly facilitated through use of a two-stage sample design: first sampling block clusters and then sampling individual housing units from within a block cluster. Also, a change to a two-stage sample design could facilitate the use of hierarchical models that incorporate random effects associated with cluster membership (see, e.g., Goldstein, 1995, Sedransk et al., 1997), which could provide a more thorough understanding of factors that affect the data. While the panel understands that time is short, the Census Bureau should consider implementing this design change.

### Oversampling ACE Blocks and Stratification

The panel has three suggestions concerning the design of the master trace sample. The current plans are to apply proportional sampling to both ACE and non-ACE block clusters. Given the additional information collected in ACE block clusters, there is much to be gained from oversampling ACE block clusters in the master trace sample. Also, some block clusters represent more difficult challenges to census processes. It would be useful to stratify the master trace sample of block clusters on some measures related to enumeration difficulties and then to oversample those strata that represent greater difficulties. This stratification would provide a larger sample size for areas for which they are probably needed. In addition, given the possibly greater nonresponse and undercoverage for list/enumerate and update/leave areas and other difficulties with "Be Counted" forms, the Census Bureau should consider oversampling of addresses for these areas and for people returning "Be Counted" forms. As in the case of the previous proposal, the panel recognizes that the data collection will begin very soon, so that this change may no longer be feasible.

### Improving Information on Enumeration

Important information on the number and timing of enumeration attempts at each address will be collected in the master trace sample. This information could be used to examine important issues concerning

nonresponse follow-up, such as the benefits of sampling for nonresponse follow-up.[7]

This information could also help determine the optimal number of attempts at enumeration before accepting proxy information, and it might suggest strategies for selecting optimal times and days for field enumeration. If there are differences in accuracy associated with the number of attempts (or the mode, whether field, telephone, or Internet), they might have important implications for changes in census processes. If this enumeration information is generally missing or is collected in a haphazard manner, analysis will be substantially limited.

The panel is aware that some field staff resist this request for enumeration information as an additional burden on the already difficult job of nonresponse follow-up. If this attitude is widespread, the accuracy of the data will suffer. This attitude may be most prevalent in those areas in which a full understanding of census processes is most crucial, the areas in which the mailout-mailback return rate is lowest. While the job of nonresponse follow-up enumeration is difficult, with limited opportunities for training and supervision, the panel's proposal is for a very small amount of information for each enumerator. Given the high value of this information, the panel strongly urges that the Census Bureau examine methods for encouraging field offices to give this a high priority. A memorandum to each local census office explaining why this is being added to the list of enumerator responsibilities might increase the response rate and overall accuracy for the data.[8]

Unfortunately, this and related proposals may not be able to be fully considered at this stage of planning for the 2000 census. If so, every effort should be made to give the collection of this information by enumerators a much higher priority for the 2010 census, making its collection a routine part of each enumerator's duties.

---

[7]Although this approach is currently not permitted by Title XIII, it might be possible if the law is amended before the 2010 census.

[8]However, unusual attention to these data might result in enumerators providing falsified responses consistent with a "perfect" performance. To avoid this effect, the reason for collecting this information should be made clear.

### Retaining the Master Trace Sample Input Files

The retention of input files for use in the master trace sample database needs to be given a high priority. The immediate objective must be to collect and preserve all of the information that will be put into this database. The panel believes that the Census Bureau agrees with this objective, so we offer this proposal mainly to provide emphasis and reduce the chances of an omission. It would be natural during the intense activity that is typical of decennial censuses to overwrite or erase a data file due to the need to collect more contemporaneous information for monitoring some census process. To avoid this possibility, the panel suggests more emphasis and planning be given to the need to capture perishable files as soon as their immediate utility ends.

### Setting Priorities for Structure, Access, and Research

The Census Bureau has postponed final consideration of the database layout for the master trace sample database in favor of final planning for the data collection. The panel thinks that this is sensible. A database structure that can accommodate all of the planned links will be a challenge, though we are confident that a format will be identified that will be satisfactory.

Toward that end, the structure that is ultimately selected should facilitate access, that is, it should be designed to facilitate the most likely analyses. Efforts should also be made to keep the structure as uncomplicated as possible so that the database is available and accessible by Bureau staff for analysis as soon as possible to help guide 2010 planning. To help meet this goal, the Census Bureau needs to be specific about what needs in order of priority will be met by the master trace sample and then operationalize those needs through the database design. An assessment of priorities is a preliminary step for the master trace sample database that will help in many areas, such as in deciding which additional inputs should be included. The panel hopes to be able to consider the issue of needs in its future work.

Individual data files feeding into the master trace sample database should be made immediately available inside the Census Bureau for "univariate" analysis for help in evaluating the 2000 census. Finally, creation of a public-use master trace sample data file should be explored, possibly addressing confidentiality concerns through a variety of masking techniques. One advantage of doing this is that it would effectively extend the

Census Bureau's analytic capabilities. If the confidentiality concerns are nonstandard, the creation of a public-use file should be given low priority.

### Increasing Resources

The panel stresses that the master trace sample database may be the single most useful source of information for assessing alternative designs for the 2010 census. The panel believes that insufficient resources have been allocated to this important and difficult undertaking. The group within the Census Bureau that is currently responsible for collecting the data files to support the master trace sample database and designing the database is relatively small, and most if not all of them have other duties. To the extent possible, additional people should be allocated to this effort, or, barring that, more of the people currently charged with this work should have it as their sole responsibility.

### Collecting Information for a Model of Total Census Error

The panel believes that the Census Bureau's current plans are to make use of a model of total census error to assist with census evaluations and decisions, including the decision on whether and how to make use of the 2010 analog to ACE to support adjustment of census counts. Without a priori support for or use of any particular version of this model, evidence is needed to show that the plans for the master trace sample, augmented by the planned evaluation studies for the 2000 census, contain information that will support reliable estimation of all components of error for the 2000 census.

# Research Experiments

## THE NON-AREX 2000 EXPERIMENTS

The Census Bureau is planning on carrying out five experiments during the 2000 census. The Administrative Records Experiment (AREX 2000) is separately described and discussed below. The four other experiments, for which we have no suggested changes, are briefly described here.

• *Incentives for Alternative Response Modes* For a sample of households, the Census Bureau will examine whether offering free time on a telephone calling card increases the frequency of respondents' use of the telephone or the Internet to provide their census responses. There will be a control group and two treatment groups, both of which are offered the calling card. Each group is made up of three panels, one for each response mode: (1) reverse computer- assisted telephone interviewing, (2) automated spoken questionnaire, and (3) the Internet. While households may respond with the paper questionnaire, the calling card is activated after one of the above alternative response modes (selected for each panel) is used.

• *Use of Noncognitive Tests of Enumerators* For a sample of enumerators, the Census Bureau will examine whether the use of a test of noncognitive characteristics can help to predict enumerator effectiveness. A sample of local census offices will administer a noncognitive test to a sample of their new hires. (This test will not be used in the selection

process.) The test to be used is the Wonderlic Employee Inventory, which measures long-term job commitment, trustworthiness, courteous job performance, conscientiousness, emotional maturity, safe job performance, and freedom from disruptive substance abuse. The goal of the experiment is to see if there is an association between the results from this test and employee turnover and work performance. The test would have to be judged to be job-relevant, have no between-group differences, and be criterion-related to be used in the selection process.

• *Alternative Questionnaire Study* For a sample of households, the Census Bureau will examine whether alternative questionnaires with changes to one of the following—the overall appearance of the census questionnaire, the order and wording of questions on ethnicity and race, and the wording of the question concerning the definition of household residence—result in improved levels or accuracy of response.

• *Study of Attitudes on Confidentiality and the Use of Social Security Numbers* For a sample of households, the Census Bureau will examine attitudes toward requests for social security numbers on census forms and notification of their use in linking to administrative records to support statistical analyses. There are two components. The SSN-notification component examines the effects of a request for social security numbers or a public notification on mail or item response rates in the decennial census. There will be different panels, depending on whether the household receives the long form or the short form, whether there is a request for social security numbers, whether the request is for the entire household or only for the person completing the form, and whether there is a notification that mentions the Census Bureau's possible use of statistical data from other federal agencies. The SPA 2000 component is a household survey that assesses the public's attitudes toward the Census Bureau's possible use of administrative records. The objective is to measure the public's opinion (of the federal government and the Census Bureau) of expanding uses of administrative records and of the Census Bureau's possible interest in collecting social security numbers in the future. SPA 2000 will collect data twice, once before advertising and once immediately after April 1, 2000.

## THE AREX 2000 STUDY

### Description

The Administrative Records Experiment in the 2000 Census (AREX 2000) is an experiment that will test, in a limited set of sites (Baltimore City, Baltimore County, and three counties in Colorado), the use of administrative records as a data collection method. The goal of this test is to better understand the potential for merged lists of administrative records to deliver content and geographic detail, especially sufficient coverage, that meet reapportionment and redistricting requirements. Two methods, which differ by whether individuals are linked to households, will be used in the experiment to generate block-level population counts. The "top-down" approach merges a number of administrative records files, unduplicates them, and then assigns individuals block-level codes using the MAF/TIGER (Master Address File/Topologically Integrated Geographic Encoding and Referencing) system. The "bottom-up" approach matches residences found on the merged administrative records list with those on the census master address file, resolving any inconsistencies. This approach therefore forms households. The resulting population counts, disaggregated by race, Hispanic origin, and gender, will be compared with census counts.

Four evaluations are planned: (1) comparison of population counts (disaggregated demographically) to census counts; (2) analysis of the components of the implementation of the experiment, including effects on coverage and costs; (3) comparison of household level information for the "bottom-up" method; and (4) assessment of the feasibility of the use of administrative records for the purposes of delivering geographically disaggregated count and content information that meets reapportionment and redistricting requirements. Several eventual uses of merged lists of administrative records are under consideration: nonresponse follow-up, reducing differential undercoverage, imputation for item nonresponse, address list improvement, and special procedures to target individual groups. For nonresponse follow-up, the question to be answered is whether there are advantages to the use of (pre-follow-up) census data and AREX in comparison with the current use of (pre-follow-up) census data combined with nonresponse follow-up data.

### Analysis and Recommendation

The panel questions whether the goals for AREX 2000 are consistent with the current plans. Specifically, how will this experiment inform the Census Bureau sufficiently to support more refined testing of administrative records (with respect to any of the above purposes, e.g., for nonresponse follow-up or for reducing differential undercoverage) for 2010? Clearly, assessing coverage of the merged list, possibly in conjunction with the census, is the first crucial step toward any of the possible uses of administrative records. A two-way match of the list of census enumerations and the merged administrative records list is a very important addition to this experiment.

**Recommendation: The Census Bureau should modify the plans for AREX 2000 to include a match of the merged administrative records list with the 2000 census households. Costs of this new activity could be controlled through the use of field follow-up of unresolved matches on a sample basis.**

Without a match, one could imagine a situation in which administrative records were considered deficient due to a lack of coverage, but matching (with additional later work) could have shown that missing people were generally included on an administrative records list that had been omitted in the test. Inclusion of that administrative records list would have eliminated the problem. More generally, a match might indicate why the merged list was inferior. Conversely, one could imagine a situation in which administrative records had excellent coverage in comparison with the census, but the process used to merge the lists was deficient, or the lists were of poor quality, so that a positive assessment of coverage was erroneous and the merged list was not nearly as useful as it seemed.

Furthermore, without a match, how will AREX 2000 adequately demonstrate the value of a merged list of administrative records for nonresponse follow-up? Distributions of information on characteristics will not indicate whether the answers are of high enough accuracy to be used for this purpose. Matching the census and AREX 2000 files will provide important additional information on the ways in which administrative records data are inferior or superior to census data, which could be easily missed by comparison of aggregates. If costs are a primary concern, sampling could be used to limit the costs of field reconciliation of differences. In ACE

block clusters, one could also use matching to determine whether individuals that are enumerated by the ACE survey and missed by the census appear on administrative records lists, which would help to assess whether administrative records could be used to reduce census undercoverage. This matching is a first step toward measurement of correlation bias and is also a preliminary step toward implementation of triple system estimation.

In addition to its recommendation for matching, the panel has six other proposals for AREX 2000. First, the Census Bureau should investigate whether AREX 2000 could be used to assess the possibility of using administrative records, in nonresponse follow-up, to supply census data only for the individuals who have a good match, i.e., those cases with a high likelihood of matching as derived from the Fellegi-Sunter algorithm, reserving field enumeration for the remaining, more difficult cases. The panel has not examined the overall feasibility of this approach; if it can be done with only a modest increase in costs, the Census Bureau should consider it.

Second, given the possibly limited utility of the current combination of administrative records used in AREX 2000 to cover the non-elderly poor, the Census Bureau should explore the process of obtaining approval from each state (and some counties) to acquire their food stamp records, as well as the process of obtaining Medicaid files. While obtaining 50 separate approvals from states for food stamp records may be time consuming, the great majority of them might be straightforward. Moreover, the immediate concern is to acquire files for the test areas, which lie in only two states. It would be extremely useful if the Census Bureau could acquire food stamp and Medicaid files for the test areas to assess the value of these lists for these purposes. The Census Bureau should give high priority to getting approval for the acquisition of these files for the test areas (which has been generally feasible in the past). (Besides providing information on the non-elderly poor, both food stamp and Medicaid files are the two obvious candidate record systems for augmenting Internal Revenue Service [IRS] data files for improving coverage.)

Third, given that facilitating acquisition of administrative records through legislative or legal change can sometimes take a considerable amount of time, the Census Bureau should assign a relatively high priority to beginning these activities as soon as possible. This is especially important if food stamp records are to be used.

Fourth, the second evaluation that is planned—analyzing components of the implementation of the experiment, including effect on coverage and

costs—does not appear to be fully thought out. It has not yet been fully described to the panel, and questions remain as to exactly how the data files are going to be merged, the effects resulting from the current choice of data files, how costs will be measured, etc. There is a concern that AREX 2000 is too specific a test on which to base what might be a stop-go decision with respect to the use of administrative records in the 2010 census. Some of the specificity of this test may stem from what appear to be limited resources allocated to it. The panel is concerned that AREX 2000 might not have adequate staff support. Given the importance and complexity of this experiment and given the additional suggestions made here, the panel believes that additional resources are needed. Therefore, the panel suggests that the Census Bureau consider allocating additional resources to this experiment and that it plan additional tests of administrative records for these purposes to learn as much as possible about the choices that are related to better or worse outcomes.

Fifth, the panel is concerned that the above uses of census data often require information on race and ethnicity. The panel strongly recommends that the Census Bureau, very likely in conjunction with other federal agencies and in conjunction with academia and research organizations through directed research, examine how to improve the race/ethnicity data available from administrative records and what the social and political implications would be as a result of the addition of this information to administrative records.

Sixth, these applications obviously require the matching and unduplication of records both from within a single administrative records system and across systems. Therefore, the Census Bureau should include in AREX 2000 an assessment of the accuracy of the matching and unduplicating operations used and an assessment of which data from duplicate records are the most accurate.

# Auxiliary Data Systems

The Census Bureau has created several data systems for the 2000 census to monitor and control various aspects of address list formation, mailout-mailback, nonresponse follow-up, data capture, and the compensation of enumerators and other field employees. While the primary goals of these systems are monitoring and control, they also can be used to examine census processes, and so they are useful for identifying areas in which alternative methods might have advantages for the 2010 census. The panel examined these systems to determine whether data that were being collected, often on a temporary basis, could be retained for later analysis.

The panel's proposals for retention of auxiliary census processing data focus on additions to the master trace sample. The richness of that planned database makes it a preferred environment for analysis. Therefore, rather than suggest retention of, say, information on the primary selection algorithm, the panel has suggested that information from relevant census processes be folded into the master trace sample database. The panel has no other suggestions concerning retention of census data; however, the panel does have suggestions on how the data systems themselves might be modified with respect to improvements in satisfying their primary purpose.

The panel was briefed on nine main data systems:

(1) the decennial master address file (DMAF), which includes every master address file residential address;

(2) the decennial response file 1 (DRF1);

(3) the decennial response file 2 (DRF2);

(4) the census unedited file (CUF), which is essentially a composite of information from the DMAF and the DRF2 files;

(5) the census edited file (CEF), which is also essentially a composite of information from the DMAF and the DRF2 files;

(6) the operational control system 2000 (OCS 2000), which monitors field data collection activities, including nonresponse follow-up, in the regional census centers and the local census offices;

(7) the management information system (MIS 2000), which tracks on a daily basis the status, cost, and timing of each census operation in terms of the anticipated completion and budget in the master activities schedule;

(8) the preappointment management system/automated decennial administrative management system (PAMS/ADAMS), which covers enumerator hiring and compensation; and

(9) the data capture system (DCS 2000), which monitors the receipt of census forms and monitors which questionnaires have been successfully processed in one of the data capture centers.

In addition, summaries from the management information system will feed into a data warehouse that Census Bureau staff will use to adjust resource allocation and make other day-to-day decisions.

These systems represent an impressive collection of software specifically developed for running the 2000 census. Many of them are updates from earlier versions used for the 1990 census, and most of these systems were successfully tested in the dress rehearsal.

The panel has two proposals for future censuses. First, the Census Bureau has decided to discard the visual images of the census questionnaires.[9] Even for the entire census, but certainly for the master trace sample,

[9]One of the first steps in data capture for the 2000 census will be to "photograph" each census form, i.e., obtain a visual image. (The physical form will be destroyed early in the process.) Optical mark and optical character recognition operates from this visual image of the census form. Once the data are captured from these visual images, the visual images will also be discarded. There are groups that are concerned over the plan not to save these electronic visual image files. They argue that not saving them would be a major change in policy from the past (see *Federal Register*, 1999). The panel did not discuss this aspect of the issue for the current report.

the visual images represent a relatively modest amount of data that could provide important information in planning the 2010 census. They could provide information about the technology needed for data capture, errors that are made in data capture, and issues concerning ease of response, such as problems with the use of foreign language questionnaires. Therefore, the Census Bureau should reconsider its decision, particularly for the master trace sample cases.

Second, the software systems were developed on site by Census Bureau staff. This approach was certainly the correct decision in the 1980s given the unusually large size of census data sets relative to the typical commercial applications at that time. In addition, there are obvious benefits to using custom software that is specifically targeted to the various census applications. Also, custom software provides the Census Bureau with greater understanding of the workings of the software, facilitating modifications and maintenance. However, continuing advances in technology have made large data sets more common with the associated development of commercial products, and there are considerable benefits to making use of more standardized commercial software to the extent possible. The greatest disadvantage of custom software is that it makes the Census Bureau too dependent on a small number of employees who fully understand it. Maintaining custom software also requires that the Census Bureau retain expertise in both using and modifying the software. It is difficult to hire people for either task because of the lack of outside expertise (and because the skills then acquired are not easily transferred to other jobs). Therefore, the panel proposes that the Census Bureau reevaluate its decision to use custom software for the 2010 census.

# Other 2000 Census Data Issues

In addition to issues directly related to the panel's charge for this first interim report, the panel's work raised some related issues concerning the 2000 census. This section presents the panel's findings and analysis on those issues.

## SETTING PRIORITIES FOR EVALUATION STUDIES

Many of the group evaluation plans[10] described to the panel ended with data collection, providing little detail on how the data would be used.

**Recommendation: The Census Bureau should develop a detailed plan for each evaluation study on how to analyze the data collected and how to use the results in decision making. The Census Bureau should then use the plan to identify the benefits and resources required for each evaluation study, set priorities among them, and allocate sufficient resources for the careful completion of all or, at least, the highest priority evaluations.**

These plans should include detailed information on how the data will be analyzed, how the results obtained will inform decisions about the 2010

[10]The panel was not provided with a list of the individual planned census 2000 evaluation studies.

census design, and what resources, in terms of data collection costs and staff expertise, are required.

Understanding the costs and technical resources required and the potential information value of each proposed evaluation study will help in setting priorities and allocating resources to the studies. Setting priorities is critical because of the large number of evaluation studies being planned by the Census Bureau; there may be insufficient staff and expertise to adequately use the data being collected. In setting priorities, the objectives should be both to assess 2000 census operations and to provide information to support 2010 census design decisions. The results of this process will allow the Census Bureau to make informed decisions about which are the most important and the least resource-intensive evaluation activities to carry out following the 2000 census.

## MEASURING THE ACTIVITIES OF LOCAL COMMUNITY GROUPS

For the 2000 census, the Census Bureau has greatly promoted participation by community organizations, with the goals of increasing the mail return rate and overall responsiveness to the 2000 census. If the goals are met, the result will be lower costs and, possibly, improved data accuracy, by reducing the frequency of proxy response and differential undercoverage.

**Recommendation: The Census Bureau should examine various methods for collecting information on the activities of local organizations that are intended to produce a more complete census count.**

Community organizations will vary in their effort and methods. Some groups might make greater use of "Be Counted" forms, speeches at public gatherings, or door-to-door canvassing. Such variation may complicate evaluation of the census because evaluation will depend on knowing what activities took place in each area. Therefore, to the extent possible, the Census Bureau should examine ways of collecting information about the activities of these groups. For example, they could be asked to supply a diary of their activities after the 2000 census is concluded, or the workers at each local census office could be asked to monitor the activities of local groups, or the staffs of local groups could be interviewed.

The Census Bureau is also planning on providing mailout-mailback areas with a day-to-day measurement of their return rate of the mailed

questionnaires. It would be useful to examine the effectiveness of this system for encouraging local activities to increase mail response. Therefore, this information on daily response rates should be retained, along with information about the areas, including the final census and adjusted counts. Finally, any adjustment of the census will suffer from not being able to incorporate information about local efforts.[11]

[11]Without these data, some areas with strong local efforts, and thereby possibly reduced undercoverage, will likely have estimated undercounts that are biased upward. Even if this information were collected, its incorporation in an adjustment model would be difficult to carry out.

# Other 2000 Census Proposals

## EXPANDING THE CONTENT REINTERVIEW SURVEY TO MEASURE RESPONSE BIAS

The Content Reinterview Survey involves a reinterview of a sample of addresses that received a census long form. In previous censuses, the Content Reinterview Survey included a panel for which the reinterview was more intensive, so that the study could be used to measure not only response variance, but also response bias. The survey proposed for 2000 involves only a reinterview that mimics the questions used in the 2000 census, so it can only measure response variance. Measurement of response bias is arguably more important than response variance, as it will provide important information for revising questionnaire wording in future censuses.

Response bias is likely to vary by mode of response, which is important given the large number of response modes, including "Be Counted" forms, the Internet, and, possibly, greater use of the telephone. Bias, unlike variance, does not generally diminish with aggregation. Therefore, the panel proposes that the Census Bureau consider expanding or refocusing the Content Reinterview Survey to measure elements of response bias, particularly for census short-form items. To examine this, "Be Counted" forms, telephone responses, and Internet responses—which are all short form responses—should be sampled, and at a greater than proportional rate, in the Content Reinterview Survey.

### Evaluating the Quality of Imputation and Editing

The imputation and editing routines used in previous censuses have received relatively little evaluation in preparing for the 2000 census. This is surprising, given their impact on census data quality. This issue is especially important given the recent advances in techniques for imputing missing values. Data should be collected during the 2000 census to support evaluation of the current editing and imputation procedures and, more importantly, to support research on new editing and imputation methods for the 2010 census. These data could come from administrative records or from expanding the Content Reinterview Survey to include a bias evaluation component, as described above.

### Collecting Comprehensive Cost Data

Costs of the decennial census have grown in real terms for several censuses (National Research Council, 1995), and two major questions are how much the United States should spend on its decennial censuses and what additional advantages are gained through the use of more or less expensive designs. The only way that such questions can be addressed is to understand in detail the sources of the current costs. Therefore, the Census Bureau needs to collect data on census costs in as much detail as possible, to permit their comprehensive modeling for alternative design options for 2010. In addition to the use of the PAMS/ADAMS system as an input to the master trace sample, the detail should also include the costs of forming the address list, advertising, printing forms, contractors, etc.

### Collecting Comprehensive Information on the Accuracy of ACE Matching

The most error prone component of the ACE, which underlies the plans for adjustment of census counts, is probably the matching operation. For the 1990 census, two studies were used to measure the extent of matching error: the comparison of two independent matching groups, and the results of an evaluation follow-up study. As discussed by the National Research Council (1999:Ch. 4), comparison of the matching groups was not a direct measurement of final matching errors, and the evaluation follow-up study was completed so long after the census that responses were obtained for only about 40 percent of the sample. Given the importance of

measuring matching error, the Census Bureau should plan to collect sufficient information so that research on adjustment for the 2010 census can be fully informed on this issue.

### Continuing Ethnographic or Systematic Observation

In the 1990 census, ethnographers (also referred to as systematic observers) collected information for a heterogeneous collection of 29 small areas on the amount and causes of census undercoverage (Brownrigg and de la Puente, 1993). These case studies were useful in providing hypotheses as to possible ways in which the census might be modified to reduce undercoverage. The panel encourages the Census Bureau to develop similar studies for use in the 2000 census, and to make the plans for these studies available for review to various census advisory groups.

### Preparing for Large Numbers of Internet Responses

Finally, the panel is concerned that the number of households using the option of the Internet for responding to the census short-form questionnaire will exceed the capacity of the census webpage servers. The Census Bureau should be very optimistic in its assessment of the number of people likely to use this option on the days immediately after the mailing of the census questionnaires and plan accordingly.

# Planning for the 2010 Census and the ACS Survey

Looking beyond the 2000 census, this section covers planning for the 2010 census and the proposed American Community Survey.

## ACS AS A CENSUS TESTING PLATFORM

The American Community Survey is a proposed national, continuous, mailout-mailback survey of 250,000 households per month, with field follow-up that makes use of techniques closely related to those used in the census. Therefore, rather than rely exclusively on the two or three large-scale census tests, which are always at least slightly limited in their generalizability by the specific locations selected, the Census Bureau could use the ACS as a platform for testing possible changes in the census. This work could serve as preliminary testing to the larger mid-decade tests for the census design.

## A MATCH STUDY OF THE CENSUS SHORT FORM AND THE ACS

The decennial census makes use of one residence rule definition, the ACS uses a second, and a third approach is being tested in the alternative questionnaire study.[12] As the Census Bureau is well aware (based on the

[12]The difference between the census and the ACS residence rules stems from the fact that the ACS defines residence at a specific time while the decennial census defines residence as the predominant residence over a year.

allocation of an experiment to this issue), confusion over residence rules is a source of possibly substantial error in the census. The panel suggests that the Census Bureau evaluate the methods used in an early 1990s study, the Living Situation Survey (Martin, 1999), in which it was shown that a large number of individuals, especially in Hispanic communities, were missed through use of a question similar to the current census question on a household's current residents. The Census Bureau needs to find the residence rule (within the set of rules satisfying legal and other restrictions) that results in the most accurate estimates. To learn more about this issue, the panel proposes an ACS-short-form match study in 2000 to examine this and other short-form measurement error issues.[13] Another possible goal is to make use of some of the ideas in the Living Situation Survey in the 2010 version of the ACE survey.

## ADDRESS LIST IMPROVEMENT

An accurate list of housing units will be needed for both the ACS and the 2010 census, making it a high-priority item. The accuracy of the address list may be the most important factor in determining the overall accuracy of a decennial census. The panel suggests some modifications to various aspects of both the MAF/TIGER software and the process for updating the master address file.

First, MAF/TIGER needs to be modified so that it can accommodate inputs from geographic positioning system (GPS) technology to permit better identification of the location of rural addresses and boundaries in MAF/TIGER. Once this is accomplished, data collection augmented by GPS inputs should begin as soon as possible.

Second, in the field check of addresses for MAF, every third address is directly visited, with indirect observation of the housing units on each side. It would be useful if the Census Bureau could make use of some form of evaluation to determine whether the information collected for housing units directly visited was better than that collected for neighboring housing units.

Third, there needs to be a more specific plan on how to interact with localities to make the Community Address Updating System (CAUS) successful. For the 2000 census, the Census Bureau used information from

[13]The ACS and the census long-form samples will not overlap, so a direct match study of ACS and census long forms is not possible.

the U.S. Postal Service as a source of address updates: the Address List Improvement Act of 1994, permitted for the first time, the transfer of data from the Postal Service to the Census Bureau for decennial census purposes. The Address List Improvement Act also established a program, called LUCA (Local Update of Census Addresses), which permits representatives of local governments (who have signed a confidentiality agreement) to review and provide feedback on the address list for their local areas. This program is based on the theory that local jurisdictions have unique access to administrative data and local knowledge. Ultimately, LUCA is supposed to be an integral part of CAUS, which is a larger program of continuous address list updating. Such updating is required by the ACS, which uses the MAF as a sampling frame, and will also be required for the 2010 decennial census. In order for CAUS to work, input from localities must be of high enough quality to be useful, and participation needs to be substantial and relatively uniform across the nation.

There are several hurdles that the Census Bureau must overcome before any system of continuous address list updates can be put into place, including: establishing the strengths and weaknesses of the Postal Service Delivery Sequence File (DSF) updates; determining how other sources of information, including commercially available lists and input from local users, can be used to overcome these deficits; determining how to facilitate effective data transfer between Census Bureau staff in Washington, staff in the regional census offices, and representatives of localities; and gaining an in-depth understanding of what actually happened in the 2000 LUCA program (and its implications for the entire address list review process). Specifically, more needs to be known about the effectiveness of the DSF update in areas with city-style addresses as a tool for updating the MAF. Evaluations are needed to better determine MAF coverage (e.g., MAF's coverage of units in multiunit structures); determine the presence of newly constructed units in existing and new structures, including data on the lag between completion and reporting; and determine the accuracy of DSF updates for different types of areas.

For the LUCA process in areas with city-style addresses, information is needed on what happened in 2000 by addressing two questions: How many areas participated in LUCA, with participation defined as the actual submission of address information? How were these participants distributed by census region, state, size of place, and other key demographic characteristics, such as population growth and level of undercount or overcount?

Information is also needed about the effectiveness of the LUCA effort in areas without city-style addresses.

Finally, after the above work has been completed, the Census Bureau needs a comprehensive plan to engage localities in CAUS, including: the process that will be used to link Census Bureau headquarters, the regional census offices, and local area representatives; methods for updating the address list; and interface with ACS staff for update and feedback purposes.

## TARGETED REPLACEMENT QUESTIONNAIRES

It was clear from the dress rehearsal that a targeted replacement questionnaire—a second questionnaire that is mailed to housing units that fail to return the initial mailed census questionnaire—would be likely to raise the mail response rate by around 7 percent. Such a rise would significantly reduce census costs and greatly aid subsequent census operations, probably improving the ultimate accuracy of the data collected. Research in the early 1990s (Dillman et al., 1994) demonstrated the value of a targeted replacement questionnaire, and it was therefore included in decennial census plans through 1996. However, it was then discovered that commercial firms were unable to address the number of envelopes for housing units that would be expected not to return a questionnaire in the amount of time needed to keep on schedule. Given this problem and the number of duplicates that would probably have resulted from the use of a blanket replacement questionnaire (mailing of a replacement questionnaire to all housing units, including initial respondents), as demonstrated in the 1998 census dress rehearsal, the Census Bureau decided not to use a blanket replacement questionnaire in the 2000 census. Thus, the 2000 census will include neither blanket nor targeted replacement questionnaires. However, technology is quickly changing, and the feasibility of such an important improvement should be examined as soon as possible. The Census Bureau needs to find out how to make use of a targeted replacement questionnaire in 2010 (see National Research Council, 1999).

## LEGISLATIVE AND LEGAL INITIATIVES

Several legislative and legal changes would be needed to implement some proposed features for the 2010 census. These changes include use of sampling for nonresponse follow-up; modifications to administrative records systems, such as inclusion of the address of residence on tax forms;

and, as discussed above, greater access to administrative records (e.g., food stamp records). These initiatives should receive attention as soon as possible. Also, the Census Bureau should inform policy makers that decisions on ACS have extremely important implications for the 2010 census and, therefore, that decisions about ACS should be made in light of that relationship. ACS has been correctly viewed as a substitute for the census long form, but there are many other relationships that should be considered, especially the continuous update of the MAF through ACS processes.

## AN ACS ADVISORY GROUP

The development of the ACS raises a number of issues related to the quality of and planning for the 2010 census. There are also many other important technical issues raised by the introduction of the ACS into the federal statistical system. Formation of a technical working group could help to address many of these issues.

## LINKING RESPONSES FROM 1990 AND 2000 CENSUS QUESTIONS ON RACE AND ETHNICITY

A sample of the Current Population Survey in the summer of 2000 will be used to develop a mapping between responses based on the question format and sequence relative to the race and ethnicity questions on the 1990 and the 2000 censuses. The panel is concerned about how the model that accomplishes this will be developed and evaluated. The statistical relationship between responses to the 1990- and 2000-style questions may exhibit strong local variation, which the sample may not be large enough to identify or which the model may not accommodate. Plans for model development and evaluation are crucial and may have important implications for the sample design. Moreover, the panel is concerned about how the results from the effort will be reported and disseminated. The Census Bureau could produce a simulated double-coded sample, as was done when the industry and occupation codes were changed from 1970 to 1980, but at present there appears to be no plan to do so.[14]

[14]A simulated double-coded sample is a method in which two different coding systems are applied to a subset of the sample and the resulting joint distribution is used for (multiple) imputation of the missing codes for the remaining cases in the sample (for details, see Clogg et al., 1991).

As a final comment, we note that the Census Bureau has contracted to obtain advice on how to take advantage of new technologies. While this is a start on the need to incorporate the latest technology in census planning and operations, it would be helpful if familiarity with technology were integrated more fully into the staff. Also, it would be extremely useful if some sort of framework were established that made the Census Bureau aware, early on, of opportunities created by advances in technology, though the panel has no specific suggestions. Even more difficult, there is a need for the Census Bureau to be able to predict future technologies that would be useful in alternative census designs. To facilitate the use of the latest technologies in the planning for the 2010 census, testing and decisions of technology-intensive systems should be delayed as long as possible in the cycle for the 2010 census.

# References

Brownrigg, L.A., and M. de la Puente

1993 Alternative Enumeration Methods and Results: Resolution and Resolved Populations by Site. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Bureau of the Census

1999 Master Trace Sample: Research Objectives. Unpublished paper, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Clogg, Clifford C., Donald B. Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman

1991 Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* 86(413):68-78.

Dillman, Don A., Jon R. Clark, and James B. Treat

1994 Influence of 13 Design Factors on Completion Rates to Decennial Census Questionnaires. Paper presented at the Annual Research Conference of the U.S. Bureau of the Census. Arlington, Va. (March).

*Federal Register*

1999 Notice of availability of proposed records schedules; request for comments. *Federal Register* 64(238: December 13):69571-69572.

Goldstein, Harvey

1995 *Multilevel Statistical Models*. London: Edward Arnold and New York, Halstead Press.

Lott, Juanita Tamayo, and Jay Keller

1999 Continuity and Change: The Decennial Census in the 21st Century. Paper prepared for the Joint Statistical Meetings, August 7-12, Baltimore, Md. (September 10, final revision).

Martin, Elizabeth
1999 Who knows who lives here? Within-household disagreements as a source of survey coverage error. *Public Opinion Quarterly* 63:220-236.

National Research Council
1988 *Priorities for the 1990 Census: Research, Evaluation, and Experimentation (REX) Program.* Panel on Decennial Census Methodology, Constance F. Citro and Michael L. Cohen, eds., Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.
1995 *Modernizing the U.S. Census.* Panel on Census Requirements in the Year 2000 and Beyond, Barry Edmonston and Charles Schultze, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.
1999 *Measuring a Changing Nation: Modern Methods for the 2000 Census.* Panel on Alternative Census Methodologies, Michael L. Cohen, Andrew A. White, and Keith F. Rust, eds., Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

Sedransk, J., D. Malec, C. Moriarity, and F. Le Clere
1997 Small area inference for binary variables in the National Health Interview Survey. *Journal of the American Statistical Association* 92:815-826.