

List of Census Bureau Research Topics

(Adopted FY 2006)

Following over two years of discussions by top leadership (Deputy Director, Associate Directors, Division Chiefs, and Senior Technical Staff) on general research at the Census Bureau, problems grouped under eleven (11) *research topics* were identified during FY 2006 to receive the highest priority for research. (Two of the original 12 research topics were combined, resulting in the current list of 11.)

RESEARCH TOPIC A: DECENNIAL CENSUS COVERAGE

Problem:

The decennial census suffers from errors due to omission of people who should be counted and to erroneous enumerations of people, the latter including enumeration of people who should not be counted at all, enumeration of people in the wrong place, and enumeration of people multiple times (duplication). These problems are substantial. For example, there were an estimated 5.8 million duplicates in Census 2000. Coverage measurement of Census 2000 was afflicted by some of the same problems. The 2000 Accuracy and Coverage Evaluation Survey (A.C.E.) was unable to accurately determine some people's places of residence, resulting in an underestimate of erroneous enumerations, many of which were duplicates. In fact, the failure of the A.C.E. to identify a substantial number of the census duplications as erroneous enumerations was a primary reason the 2000 A.C.E. estimates were viewed as inadequate for coverage adjustment of the Census 2000 results.

Discussion:

The goals of this research are to reduce coverage errors in the census, and also to improve the Census Bureau's understanding of these errors by improving census coverage measurement. Specific topics for research include the following:

- 1) Research to prevent and correct for duplication at all stages of the census and coverage measurement process, from address list development to final coverage estimation.
- 2) Research to improve determination of Census Day residence:
 - a) Development and testing to improve the survey instruments and questions, including alternative presentations and formulations of residence rules.
 - b) More basic research on errors in, e.g., recall and reporting of moves and other problematic residence situations.
 - c) Cognitive and qualitative research and field experiments to evaluate the impact of survey questions on the quality of census coverage measurements.
- 3) Development of coverage measurement methods for group quarters.
- 4) Statistical research on improving coverage estimates, including separate estimation of census omissions and erroneous enumerations.

Topics (1) and (2) aim at improving the data collected in both the census and in census coverage measurement. Topic (3) aims at addressing the Census Bureau's lack of knowledge of the coverage of the group quarters population. Topic (4) seeks to improve coverage measurement, particularly in regard to providing more realistic estimates of omissions and erroneous enumerations than have previously been available. Accomplishing these tasks would further the Census Bureau's understanding of census coverage issues, helping the Census Bureau to improve future censuses.

RESEARCH TOPIC B1: NONRESPONSE - DEMOGRAPHIC SURVEYS

Problem:

Survey nonresponse rates have been increasing, leading to concerns about the accuracy of (demographic) survey estimates. For example, from 1990 to 2004 initial contact nonresponse rates have approximately doubled for selected household surveys, including the Consumer Expenditure Quarterly (from 12% to 23.3%), the Current Population Survey (from 5.7% to 10.1%), and the Survey of Income and Program Participation (from 7.3% to 14.9%). Errors introduced by unit nonresponse may bias survey estimates when nonresponse is high and those who participate in surveys are different from those who do

not. Standard nonresponse adjustment procedures typically assume that nonrespondents are similar to respondents, but the literature does not always support this assumption.

Discussion:

General topics for research include:

- Strive to increase response rates by improving data collection procedures. Research on data collection procedures may develop methods for increasing response rates (or at least preventing further decreases) either generally or for specific surveys. This would be a true improvement, however, only if methods developed to increase response do not simultaneously degrade the quality of the data collected. Efforts to increase response rates also can lead to higher costs of data collection due to additional interviewer training and various incentive programs.
- Strive to better understand the nature of nonresponse and its effects on data quality. Doing some targeted follow-up of nonrespondents may yield information on how nonrespondents differ from respondents, as may obtaining information from other sources (such as administrative records) that contain information also collected in a survey (assuming that the records can be linked). Such information could be used to develop a better scientific understanding of the nature of survey nonresponse. It may also be useful for research on Topics 1 and 3.
- Develop better procedures to adjust for nonresponse. If nonresponse rates cannot be reduced to negligible levels, the Census Bureau will need to evaluate current nonresponse adjustment procedures and do research to develop improved procedures. Possibilities to explore include comparing hot deck and model-based imputation procedures, and exploring models for non-ignorable nonresponse. The Census Bureau should also pursue implementation of methods for survey variance estimation that account for error due to imputation.

Success on Topic 1 would reduce field costs and improve the accuracy of the Census Bureau's estimates. Success on Topic 2 would let the Census Bureau better inform data users about data quality and limitations, and would also facilitate work on Topics 1 and 3. Success on Topic 3 would improve the accuracy of the Census Bureau estimates and/or the relevance of its variance estimates.

RESEARCH TOPIC B2: NONRESPONSE - ECONOMIC SURVEYS

Problem:

Maintaining or improving response rates is an ongoing goal and constant challenge for economic surveys. Response rates for the 2002 Economic Census declined from 1997.

Response rates for current surveys range from approximately 30% to 95%. Response in the 2002 Economic Census was 84%, with a target of 86% for the 2007 Economic Census.

Discussion:

Improving the response rate in the 2007 Economic Census is a strategic goal. Response rates for the Economic Census and selected other surveys are critical PART (Program Assessment and Rating Tool) performance measurements.

The Economic Directorate of the Census Bureau believes that reducing respondent burden is key to improving response. Ongoing research and assessments are underway using such techniques as cognitive testing, respondent debriefings and usability testing for economic survey data collection instruments. Nevertheless, several issues remain to be addressed.

- What are the attributes of respondent burden from the respondent's perspective? Are these measurable? What are the measures? Are the components reducible via survey design? If so, how? To what degree do current activities reduce respondent burden?
- Does the hypothesized relationship between respondent burden and response rates indeed exist? How can this relationship be defined and measured?
- What activities can/should be undertaken to evaluate the effect of burden reduction on response rates? What are their costs and benefits?

Of specific interest is how effective the following strategies are in reducing burden:

- Tailoring questionnaires by company size;
- Aligning data requests with companies' accounting practices and record-keeping systems;
- Developing a company-centric approach for data collection from large companies;
- Developing more effective electronic reporting options that respondents will use.

Response may be positively affected by follow-up strategies. The effectiveness of current follow-up strategies needs to be systematically evaluated. Few directorate resources have been devoted to keeping up with recent discoveries or advances in response motivators. Research into new or alternative strategies and their effectiveness in economic surveys is needed.

Results from both lines of research - burden reduction and follow-up strategies - would benefit the economic programs through improved response rates, which could improve data quality by reducing potential nonresponse bias. Improvements would help the Census Bureau in its attempts to meet the Office of Management and Budget's new response rate standards and its own PART targets. More effective follow-up strategies have the potential to result in significant cost savings in conducting economic surveys. (For example, the 2002 Economic Census follow-up costs exceeded \$1.5 million.)

RESEARCH TOPIC C: REPORTING UNIT RESEARCH - ECONOMIC SURVEYS

Problem:

A problem for economic surveys is the potential mismatch between the Census Bureau's definition of statistical (reporting) units and the structural units of a company. The organization of a company's records may make it more difficult or impossible for the respondent to provide data according to the Census Bureau's desired statistical units. This could lead to poor quality estimates. For example, companies in some services industries cannot report data by geographic area for products or services distributed via a network.

The extent of problems associated with reporting unit definitions is, to some degree, unknown or unclear, as they are not routinely assessed for economic programs. Research is needed to evaluate the effect of mismatches on published statistics. If changes in reporting units are overlooked, the potential consequences are severe. The Census Bureau's economic censuses and surveys collect data on or process administrative data representing 180,000 multi-unit enterprises encompassing 1.7 million establishments, 5.5 million single unit establishments, and 17.6 million non-employers.

Discussion:

Proper definition of statistical units is key to obtaining high quality economic statistics and minimizing respondent burden. This will be facilitated through research to:

- Assess company record-keeping practices;
- Assess the quality and utility of administrative data;
- Assess how administrative data can be more fully used to identify structural changes;
- Improve the timely identification and incorporation of administrative data into Census Bureau programs;
- When administrative and survey data are used to develop statistical estimates, assess how definitions used for administrative data compare with those used for survey data and how these compare with data available in company records;
- Identify the correct respondent(s) for the requested data (noting that multiple sources may be required);
- Assess the effectiveness of the Customer Relationship Manager program; and
- Design efficient processes to aid data-gathering when multiple data providers are needed.

Benefits for the Census Bureau from research would include:

- Timely updates to the Business Register with timely impact on sample frames;
- More efficient/effective communication with respondents, improving the timeliness of data reporting, which could lead to earlier publication and dissemination of data products;
- Improved data quality when reporting among multiple data providers is better facilitated;
- Burden reduction through effective use of administrative data, potentially improving response rates and reducing the costs of nonresponse follow-up.

RESEARCH TOPIC D: EDITING - ECONOMIC SURVEYS

Problem:

The Census Bureau needs objective measures to assess whether its programs are over-editing their data. Over-editing has cost and quality implications. It can lead to resource problems, in terms of programmer development time and analyst review time. Equally important, it can be viewed as a quality problem – are the Census Bureau's editing procedures truly improving the reported data or are they (subtly) biasing the data based on preconceived notions of what the Census Bureau believes to be “correct”? Should the Census Bureau be identifying the failure as a problem to begin with? Finally, many Economic

Directorate programs subject the same data sets to several stages of review, with some of these stages having possibly minimal impact on the final tabulated data. This has both cost and quality implications.

At a minimum, all Economic programs should:

- Calculate edit-failure rates by respondent and by classification variable value;
- Implement and use audit trails; and
- Develop Standard Operating Procedures for analysis of both sets of measures (hopefully using statistical quality control procedures).

This problem affects most programs in the Economic Directorate of the Census Bureau. The key issue is that the directorate has no standard set of objective measures of the efficacy of edits and their associated review processes. In fact, one of the charges to the Business Process Improvement Team's Edit Efficiency sub-team was to outline the type of auditing needed in the future, and to identify five pilot programs to test these audits.

The Census Bureau believes that the overall cost of editing and multi-stage review is high; though there is no specific cost information available. With objective measures in place, the cost and quality issues could be assessed. The Edit Efficiency sub-team identified a potential savings of 10% of the resources allocated to data analysis in each of the six programs given the recommendation to calculate edit-failure rates and in each of the nine programs given the recommendation to measure the effect of edit process on the resultant data.

Discussion:

The Economic Directorate needs a method to analyze the editing process and evaluate the quality of the edits and related review processes so that it can eliminate or modify the existing procedures and conduct research on ways to improve specific edits, such as macro-editing and selective editing.

If research could help solve this problem, the Census Bureau would be changed as follows:

- It would reduce the time spent on the overall editing process while maintaining or improving tabulated data quality.
- It would free the analysts to actually do analysis.
- The Economic Directorate would have a more defensible approach to edit-review processing.
- It would promote the development of repeatable procedures that could be implemented by outside-data users, consistent with the Office of Management and Budget (OMB) Quality Information Guidelines.

RESEARCH TOPIC E: PRE-RELEASE REVIEW OF DEMOGRAPHIC DATA

Problem:

Data review of continuing surveys, such as the Current Population Survey (CPS), take a significant amount of time and resources. With the implementation of the American Community Survey (ACS), this data-review work threatens to overwhelm staff. Currently, ACS data review takes about 6 calendar months, during which time about 25 people work from 25 percent to 75 percent of their time on review of the data and data products. As ACS expands to review of full implementation data, and adds Puerto Rico, group quarters, and 3- and 5-year product lines, it will not be possible to complete the work within the current review time frame, without dramatically diminishing its quality.

The dimensions of the problem are huge.

- 1) The large-scale size and cost of getting review done.
- 2) Completing reviews efficiently and on time.
- 3) Making the mundane/routine parts of review simple and direct to do.
- 4) Keeping analysts focused on what is a highly tedious task.
- 5) Codifying behavior of routine review activities, perhaps using automation and standardized tools.

Discussion:

If the Census Bureau were to solve this problem, there would be benefits. Review would become more efficient. The frequency of errors would be reduced and problem situations would be more easily identified. Staff members would be freed for more useful activities, reducing burnout. If gold standard processes can be identified, they can be applied to other

programs, thus paying for these innovations over time. With more focused and structured review, it is possible that ACS products could be reviewed and cleared in far less time and using fewer human resources.

RESEARCH TOPIC F: SURVEY ESTIMATION

Problem:

Research to improve survey estimation techniques, including evaluation of alternative estimation strategies, is needed to keep up with changing circumstances, such as increased nonresponse and demands for small-domain estimators. The continued development of the American Community Survey (ACS) poses new challenges for estimation because the nature of the estimation problems it faces are different from those of other surveys, even the census long form sample. (ACS estimates are desired for traditionally small domains from a very large national sample available yet without corresponding 100 percent census counts in most years.)

The Census Bureau identified a general problem as well as several specific problems in this area. The general problem is that while the Census Bureau's design-based estimation paradigm was developed and works well for estimation in certain situations with large samples (e.g., for many national level estimates), it does not work so well when pushed beyond this realm (e.g., into small area estimation, dealing with large amounts of missing data, dealing with outliers, etc.). In such settings, other approaches (e.g., model-assisted or model-based estimation) may offer opportunities for improvement. Movement in this direction is hampered, however, by such things as tradition, bureaucratic obstacles, lack of staff knowledge of other approaches, and difficulties with implementing a new approach in the Census Bureau's production environment. Addressing this general problem will require additional staff training (e.g., via available short courses), perhaps complemented by recruiting of new staff with backgrounds in alternative estimation approaches, as well as devoting attention to resolving bureaucratic and other obstacles to implementing new or different estimation methods.

Discussion:

Particular survey estimation problems identified for research include the following:

- 1) Study optimum use of population and housing unit controls for the ACS and other surveys.
- 2) Investigate bias and uncertainty in pop controls with an eye to developing error estimates for pop estimates; also, study the impact of these errors on survey estimates.
- 3) Do research on estimators that incorporate administrative data to improve ACS estimates for very small areas.
- 4) Do research on improving small area estimation for other Census Bureau survey applications (SAIPE, SAHIE, etc.).
- 5) Investigate the feasibility of using model-based or model-assisted estimation techniques in the monthly residential construction program (to use additional information from the large sample of building permits to improve estimation of housing starts, completions, and sales).

The Census Bureau can link the general and specific problems by noting the potential of other estimation approaches to address the specific problems just noted. For example, model-assisted estimation, in the form of generalized regression estimation, potentially can reduce the variance of direct ACS estimates without appreciably adding bias, providing a potentially useful tool to assess the average bias in the controls noted in 1) and 2). It is also relevant to the problem noted in 3).

RESEARCH TOPIC G: MEASUREMENT ERROR RESEARCH AND PREVENTION

Problem:

Recent Census Bureau experience points to measurement errors (that is, errors of observation arising from the interviewer, the respondent, the questionnaire, or the mode of data collection) as major sources of inaccurate and inconsistent data. In Census 2000, the coverage reinterview failed to identify a large fraction of the duplicate enumerations in the census (see Decennial Census Coverage Topic A.2) due to inaccurate measurements of Census Day residency. In 2000, the percentage of Hispanics reporting their race as White was 48% in the census and 63% in C2SS, perhaps due to differences in survey mode and interviewer training. Subtle and untested changes in the Hispanic-origin question in Census 2000 led to loss of detailed origin information for about 12% of Hispanics in the mailout universe.

Discussion:

The errors can be very large, and their origins are not always well understood. Very large and poorly understood discrepancies undermine the credibility of census data more broadly.

Many questionnaire design flaws that give rise to measurement errors can be identified and corrected through pretesting. (For example, the flaws in the Hispanic origin question in Census 2000 would probably have been caught had the final version

been cognitively tested.) Research to evaluate and improve pretesting methods is needed to support the Census Bureau's policy of testing all questionnaire changes. Basic research is needed to better understand sources of measurement errors. For example, why were moves and second residences unreported and/or unreliably reported in coverage reinterviews? It is necessary to understand the reasons for such errors, in order to correct them.

- 1) Carry out research on the sources and magnitude of measurement errors:
 - a) Evaluate data quality, through regular compilation of data quality indicators (e.g., item nonresponse rates), regular or special evaluations (reinterview, record check studies), and studies of respondent difficulties and misinterpretations (e.g., cognitive interviews, respondent debriefing studies).
 - b) Evaluate effects of mode of data collection on quality and comparability of survey data in order to develop guidelines for standardizing survey instruments across modes.
 - c) Conduct research on fundamental sources of survey measurement problems (e.g. recall error), drawing on theory and methods in relevant scientific disciplines, such as psychology and linguistics (see also Topic J).
- 2) Conduct developmental research and methodological studies to support measurement error reductions through improved questionnaire design and procedural improvements:
 - a) Develop and qualitatively test questionnaire revisions designed to solve the problems identified in step (1).
 - b) Conduct field experiments to evaluate solutions.
 - c) Develop translation methods and protocols.
- 3) Evaluate and further refine questionnaire pretesting methods.

Several benefits would follow if the Census Bureau were to solve this problem. The Census Bureau could produce more complete, consistent, and higher quality data, with less time and cost devoted to editing and review (See Topic D, Editing Economic Surveys, and Topic E, Pre-Release Review of Demographic Data). A better understanding of the nature and causes of measurement error would support development of improved statistical estimates and better data collection instruments and techniques. It would also help avoid the major data problems that now cause occasional embarrassment, and (when they still occur) would help the Census Bureau to explain them more credibly.

RESEARCH TOPIC H: RESEARCH ON TIME SERIES METHODS AND SEASONAL ADJUSTMENT

Problem/Discussion:

Currently the Census Bureau's Economic Directorate collects data for more than 1,400 monthly and quarterly economic data series, and uses time series analysis methods for seasonal adjustment of over 1,000 of them. However, time series analysis and modeling techniques could also be employed for other purposes, e.g., to assist in editing, imputation and estimation, and to improve the Census Bureau's published measures of statistical uncertainty. Research in the following areas, while not motivated by what are presently perceived as major problems, nevertheless represent opportunities for improvement that could help the Census Bureau save time and effort and improve the quality of published estimates. For most of these topics, *the change to current Census Bureau procedures - implementation - poses a greater challenge than does doing the research.* Also, for these topics it is generally not possible to estimate the extent of the benefits without further exploratory trials or research.

- 1) Compare time series forecasts with tabulated data for the most recent period to identify values that may require an edit response. This may reduce the tremendous labor cost currently required for editing to maintain economic data quality. This approach has been implemented in the Foreign Trade Division (FTD) and the Service Sector Statistics Division (SSSD), but might also be beneficial in other areas.
- 2) Develop statistical measures of uncertainty (e.g., variances or descriptive statistics for revisions) for seasonally adjusted estimates.
- 3) Develop and apply models for sampling error autocovariances, for possible use in such things as improving survey estimates, doing model-based seasonal adjustment, or developing variances of seasonally adjusted estimates.

Topics 2 and 3 are related. Providing measures of uncertainty in the Census Bureau's estimates is very important, but, in most cases, the Census Bureau currently provides variance estimates only for the Census Bureau's original, unadjusted data. Uncertainty measures specifically for the seasonally adjusted data would be more relevant (though research should also investigate how different these would be from those for the unadjusted data).

- 4) Use time series forecasting procedures to address issues with systemic delayed reporting of data. Some work has been done here and at Statistics Netherlands on this topic, but more research is needed.
- 5) Further investigate modeling of and adjustment for time-varying trading-day effects. Close to half of the Census Bureau 1,000+ seasonally adjusted series have trading-day effects, and for some of these series, especially the Service Sector Statistics Division's series, the Census Bureau expects that the trading-day patterns change over time. Crude methods are currently used to deal with this problem. More refined methods based on time series models have shown promise, but more research is needed, including investigation of how prevalent the problem is.

RESEARCH TOPIC I: DISCLOSURE AVOIDANCE METHODS

Problem/Discussion:

The Census Bureau wants to protect the confidentiality of its respondents' information, as required by Title 13 of the U.S. Code, while releasing as much high quality data to the Census Bureau's users as possible. In order to do so, the Census Bureau needs to develop new and improved disclosure avoidance methods. For demographic data, the goal is to counter the growing potential to identify individuals based on the expanding amount of personally identified information on the Internet. For economic data, the goal is to reduce the number of table cells that are suppressed while maintaining data protection. In addition, the Census Bureau is releasing new types of data products such as maps and indices, requiring new disclosure avoidance methods. The Census Bureau should also document the basis for existing disclosure avoidance methods and evaluate the effects of all of the methods on data quality.

These activities would involve considerable effort because of the vast amount and great variety of data products the Census Bureau releases. This problem is important to the Census Bureau because disclosure protection is required by law, because publicized disclosure could result in decreased response rates, and because disclosure avoidance procedures affect the quality of all of the Census Bureau's data products. Moreover, the Census Bureau promises respondents that the Census Bureau will protect the confidentiality of their information. In short, improved disclosure avoidance methods would benefit the Census Bureau in terms of releasing more high quality data while avoiding disclosure.

RESEARCH TOPIC J: INVESTIGATION OF LINKS OF CENSUS BUREAU PROBLEMS TO BEHAVIORAL SCIENCE RESEARCH

Discussion:

Progress in many of the behavioral sciences has been rapid in recent years. These disciplines include several areas in psychology, cognitive neuroscience, and linguistics. For example, the basic science in memory, language, attitudes, and affect has each progressed far beyond the understanding of these areas by most statisticians and survey methodologists. The value of basic science in memory seems self-evident, but the scope of current results and open questions-familiar to readers of *Science*, *Nature*, and a large number of neuroscience and psychological journals-appear virtually unknown to survey researchers. Although sociolinguists are represented in small numbers in survey research, the Census Bureau lacks a critical number of researchers who can represent the current findings from the study of syntax, linguistic pragmatics, or psycholinguistics. Social and cognitive psychologists are revising the conceptual understanding of attitudes, but the implications of these changes appear unrecognized by survey researchers. Advances in the study of affect now touch many other areas of psychology, including cognitive psychology.

Incorporating new knowledge from these fields could potentially improve the Census Bureau's censuses and surveys. Development of more coherent theory (e.g., of survey response or of recall failure in surveys) may be furthered if survey methodologists are exposed to the scope of relevant developments in the behavioral sciences. "Building a bridge to the behavioral sciences" may help the Census Bureau realize some of the potential benefits to its work from the recent developments in these fields. Several suggestions are offered:

- 1) Develop models more closely aligned with current behavioral science, and test the models using past data or new experiments.
- 2) Current survey research staff (particularly younger staff), especially those in the behavioral sciences or

related disciplines, may be encouraged to keep up in their fields and read in some related disciplines. This could be enhanced by making modest investments in appropriate books and key journals for the library, and by encouraging staff to write papers or present informal talks reviewing relevant developments in the behavioral sciences.

- 3) Fund outside researchers in the behavioral sciences to work on problems relevant to the Census Bureau's applications. This is more likely to be successful if the outside researchers have agency collaborators who already have a working knowledge in the same field.

RESEARCH TOPIC K: RESEARCH TO IMPROVE POPULATION ESTIMATES

Problem:

Discrepancies noted in comparisons of population estimates to Census 2000 results identified the need to improve the annual estimates of international and internal migration. In addition, the use of detailed annual population estimates as controls on estimates from the American Community Survey (ACS) imposes new requirements on the population estimates. This research will examine alternative data sources and improved use of the American Community Survey data to estimate international and internal migration.

Discussion:

Increased use of detailed population estimates as the basis for funding mechanisms and as controls for the American Community Survey has led to increasing scrutiny on the data, methods and results of the detailed population estimates, especially in the most vulnerable areas of international and internal migration. If these identified issues continue to remain unresolved, the reputation of the overall Census Bureau will be affected.

- 1) The proposed research on international migration will focus on measuring the annual inflow of migrants to the United States, using selected data from the annual ACS. Estimating the annual outflow of migrants from the United States requires research into the development and validation of new methods combining statistical modeling with demographic data sources. Research on internal migration will focus on integrating new data sources and statistical modeling to identify and reduce bias inherent in the current methods to measure internal migration.
- 2) Research is needed on measuring population on a current residence basis and on developing models to reconcile the current residence population with the usual residence population.
- 3) Research will examine bringing additional data sources into the subnational population estimates. These sources will include the ACS, but might also include administrative sources on employment and taxes.
- 4) Research is needed on improved methods to achieve integrated and consistent population estimates at different levels of geography. The current approach begins at the county level, with the estimates controlled only at the national level. Control at other levels and a more integrated approach are possible.

Improvements in the annual measurement of international and internal migration will lead to improved population estimates and survey controls, which can improve estimates from numerous demographic surveys, especially the American Community Survey. This would enhance the overall quality of Census Bureau products and the agency's reputation.