Possible 2020 Census Designs and the Use of Administrative Records: What is the impact on cost and quality?

Frank A. Vitrano, Maryann M. Chapin

U.S. Census Bureau
4600 Silver Hill Rd., Suitland, MD 20233

Frank.A.Vitrano@census.gov
Maryann.M.Chapin@census.gov

Introduction

The 2020 Census will occur in a more demographically and culturally diverse nation, with a population characterized by increasingly informal and complex living arrangements. Changes in technology and its uses are occurring at an ever faster rate. Traditional enumeration methods may no longer be acceptable to the population nor be effective in producing a high-quality census.

The Census Bureau must innovate to improve the census, adapting to these changes. This requires significant planning, research, and testing early in the decade. The Census Bureau has a long history of successful methodological and technical innovations. For example, in 1940, the Census Bureau began collecting census information on a sample basis, which continued until Census 2000, the last census to use the "long form" (since replaced by the American Community Survey – the ACS). In 1960, the Census Bureau initiated mailing out forms prior to physically visiting every housing unit and collecting long form information from only a sample of housing units. In 1970, many households were asked to mail back the forms and, in those areas, household visits were made only to nonresponders. The 1990 Census was the first to use electronic rather than paper maps, and that approach has led to widespread commercial applications, such as Google Maps. In 2000, the Census incorporated paid advertising, which helped curb a declining mail response trend, and initiated the use of intelligent character recognition for reading electronically scanned mail back questionnaires. For the 2010 Census, the Census Bureau, for the first time, used an automated handheld listing device for updating the Nation's addresses and pinpointing the location of housing units using the global positioning system. However, the disturbing trend of full life-cycle census costs quickly growing every decade continued with the execution of the 2010 Census. Figure 1 shows the increasing cost per housing unit from 1970 through 2010 (as currently estimated), and where the costs are headed for 2020 without significant changes.
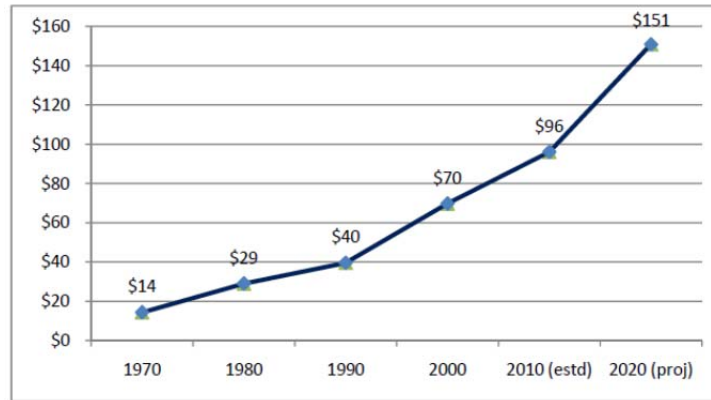
Figure 1: Cost per Housing Unit by Census Year, 1970 - 2020 (2010 dollars)

Source: Bureau of the Census. **Notes**: (1) Chart reflects cost projections based on known FY 2010 savings of $1.87 billion as of October 4, 2010 (some additional adjustments may occur as cost information is finalized/verified); (2) Projected cost per housing unit for 2020 assumes no change in design and real cost growth from 1990-2000 and 2000-2010 averaged (57%); (3) Includes the costs for the MAF-TIGER Enhancement Program and American Community Survey.

2010 Census Cost Drivers

The Census Bureau analyzed costs over the 2010 Census life cycle to determine when the greatest costs occurred. Most of the costs over the 12-year life cycle occurred in FY 2009 and FY 2010. A major cost in 2009 was associated with address canvassing (address list updating and map improvements). A large preponderance of the costs contributing to the peak in FY 2010 were for data collection. This included the costs of follow-up, including costs to hire field office staff and enumerators, and supporting infrastructure. These costs resulted, in part, from labor-intensive field operations for non-response follow-up and paper-based data collection. Also significant was the cost of initial response, including several field operations and an IT infrastructure to support all operations, both supported in large part by major contracts and Headquarters staff. By understanding the primary cost components of the 2010 Census, we can focus our research agenda on the factors that provide the greatest opportunity to achieve our cost reduction goals while still maintaining the necessary quality.

Our analysis concluded that the rising costs of the 2010 Census were largely driven by five factors: (1) the increasing diversity of the population; (2) the demand for the Census Bureau to strive for improving accuracy over previous censuses, rather than examining the benefit-cost tradeoff between quality and costs; (3) the lack of full public participation in the self-response phase of the census, requiring the hiring of a large field staff for non-response follow-up; (4) an acquisition strategy driven by lack of timely requirements, resulting in increased costs, and, (5) substantial investments in major, national updating of the address frame just prior to enumeration (2009). The first two factors, increasing diversity and the demand for absolute accuracy, are beyond the Census Bureau's control. To achieve its cost and quality targets and meet its strategic goals, the Census Bureau must make fundamental changes to the design, implementation, and management of the decennial census. As a result, the research program focuses on the other key cost drivers and what can be done to address them.

2020 Census Research Agenda

The 2020 Census program lifecycle is divided into three planning phases: a Research and Testing Phase, a Development and Testing Phase, and a Readiness Testing, Execution, and Closeout Phase. The objective of the Research and Testing Phase is to develop preliminary designs based on solid evidence and a trade-off analysis aimed at achieving the goal of conducting the 2020 Census at a lower cost than the 2010 Census (per housing unit on an inflation-adjusted basis).

The 2020 Census must strike the appropriate balance between cost and quality. As discussed earlier, field operations (especially for non-response) and the enabling infrastructure dominated decennial census costs.

Information Technology systems, which, in some cases, will no longer be used after the close-out of 2010 Census operations, were also high-cost items. As for quality, the biggest predictors of quality in the census are the address and spatial database and the rates of response by the American public.

Consequently, the 2020 Census research agenda focuses on investigating options for dramatically improving design elements related to these cost and quality drivers. This includes finding the best way to: (1) increase self-response rates, including use of the Internet; (2) reengineer the field operations (especially non-response) to reduce the field labor force; (3) modernize the headquarters IT infrastructure; (4) evaluate the quality of the address database and establish acceptable quality levels during continuous address frame updating; and (5) determine how to best use administrative records to save money and maintain quality.

Ultimately, we have categorized our research agenda into seven areas. They include Improving the Address Frame, Improving Response Strategies, Increasing Efficiency of the IT Infrastructure, Increasing the Efficiency of our Field Infrastructure, Assessing Potential Sources of Administrative Records, Improving Census Coverage Measurement, and Cross-cutting Research. All of these research areas have some aspect of research associated with administrative records, with the exception of Increasing the Efficiency of the IT and Field Infrastructures.

It is important to note that administrative records have been used in decennial censuses for decades. For example, for the 2010 Census, the Coverage Followup Operation used administrative records to identify housing units with potentially undercounted persons or persons not enumerated in a housing unit. Additionally, since 1940, Demographic Analysis has used administrative records (e.g., Vital Statistics on Births and Deaths) to measure census coverage. 2020 is the first time, however, that administrative records are seriously being considered for supplementing respondent data on a national level.

Administrative Records-Related Research Projects

Improving the Address Frame

Last year, the Census Bureau began the Geographic Support System initiative that will enable continuous updating of the Census Bureau's Master Address File and TIGER database. A continuous update of MAF and TIGER will ensure more accurate and timely information for Census Programs such as the ACS and population estimates program. This initiative would also support the goal of requiring a targeted Address Canvassing operation for the 2020 Census. The work of this initiative includes obtaining address and geographic information from many external sources, such as the United States Postal Service, state, local, and tribal governments, and possibly commercial sources. As part of this effort and related administrative records research, the Census Bureau will look to see to what extent do various administrative records sources provide valid living quarters otherwise missed from currently used sources. In addition, administrative records could assist in developing quality measures and determining if a targeted address canvassing, rather than a full nationwide canvass, could be deployed.

Improving Response Strategies

There are several areas of research where we believe that administrative records could improve how we obtain responses from the public during the census. The first has to do with how we contact the public. The 2010 Census Mailout/Mailback Operation included a four-part contact approach: notification, first questionnaire, reminder, and replacement questionnaire. This contact strategy was dependent on having the physical addresses for the housing units. This project expands the idea of respondent contact frames from the physical address to other potential contact methods. This project seeks to obtain contact information such as telephone numbers, cell phone numbers, Post Office Boxes, email addresses, and social network association. The idea is for the projects that focus on response options to test different ways of contacting respondents, such as sending a reminder to fill out the form by email or by phone call. For example, providing in a mailed out letter a link to a site and a control number may not yield as much internet response as providing this information in an electronic format, allowing the respondent to act directly from the contact information (clicking on a link and copying a control number, if necessary). It will also assess whether these alternative frames can be readily linked to a physical address. It is important to note that regardless of how we obtain the data, the census still requires counting people at a physical location for apportionment and redistricting. Furthermore, in this project we will attempt to determine the quality of the relative frames.

The second response strategy improvement research project has to do with how and when we do coding, editing, and imputation. To the extent that we believe we have high quality administrative records data for specific households, can that information be used as cues to improve responses in automated modes such as internet collection or automated enumerator-collected responses? For example, if an administrative records source identifies three persons at a household, but an internet respondent for that household only enters two of the names, perhaps we could ask additional coverage question of that respondent to make sure they included everyone appropriately.

A third response strategy improvement research project has to do with supplementing and supporting the Nonresponse Followup operation with administrative records. This project tests the use of administrative records in the nonresponse followup operation. Nonresponse Followup (NRFU) is the largest and most expensive 2010 Census operation. All households that did not return completed questionnaires and households whose forms were returned as undeliverable by the U.S. Postal Service are followed up on. The NRFU workload as of July 24 was 47.2 million cases. This includes over 600,000 new addresses added during the operation. The national staffing included more than 600,000 field staff. A quality check re-interview was conducted on approximately 5.5 percent of all NRFU addresses.   Not surprising, the size of the NRFU workload is the single largest cost driver in the census. Therefore of great interest is any way to reduce the workload either through increasing self-response or using some other method other than a field followup to obtain respondent data—such as through administrative records. A critical focus of this project is to look at the cost and quality impacts of using administrative records to support the NRFU operation.. One alternative would be to reduce the number of contacts during NRFU and add the use of administrative records. In the 2010 Census, we instructed enumerators to make up to six contact attempts with each household, three in person and three by phone. Instead, we could research reducing the number of contacts in NRFU to reduce the cost of the operation and then research the use administrative records to supplement the operation. Administrative records could also be used to help us more effectively target our field work, in conjunction with other efforts to modernize our field infrastructure. Other research projects, not dependent on the use of administrative records, will consider alternatives to our field infrastructure and to the way Census field work is managed as a means of controlling costs.

Assessing Potential Sources of Administrative Records

This project focuses on assessing the potential sources of administrative records. This includes looking at federal sources such as from the Internal Revenue Service, Medicare Enrollment information, or Selective Service. It also includes researching the potential use of past census data as well as commercial sources of information.   The data would be updated using other files, such as commercial files, the American Community Survey data, and other available records. This project will conduct a coverage study to determine relative coverage and quality among files to find the optimum mix for use in the census

Improving Census Coverage Measurement

There are several projects associated with seeing if administrative records can help us measure the coverage of the population in the census. One project examines whether new data sources and methods could increase the utility of Demographic Analysis by going to lower levels of geography and by improving the quality of race and Hispanic Origin data. Perhaps additional administrative records sources could improve these measures. A second project examines whether our traditional capture-recapture methods for measuring coverage of the population could be made more efficient if administrative records were used. For example, if there were known areas where we had high quality administrative records, perhaps we wouldn't have to independently list and interview them for comparison to the census enumeration. Then, perhaps, we could focus our field work on other sample areas. Or, should we consider the use of multiple system estimates by using administrative records as an independent measure?

Cross-Cutting Research

There are two projects that are cross-cutting in nature but are of critical importance to using administrative records in any widespread way in the enumeration in 2020. The Privacy and Confidentiality Study attempts to get a baseline on public attitudes about using administrative records, understand what the concerns are, and how concerns can be allayed. The Matching Process Improvement project examines the current approaches to matching and

unduplication as well as researching new or emerging methods to ensure that the Census Bureau can accurately conduct entity resolution (i.e., is this Joe Smith the same Joe Smith as on the administrative record?).

2010 Census Match Study

Finally, there is a research project that precedes the 2020 research that is critical to our understanding about the quality of administrative records. The 2010 Census Match Study is part of the 2010 Census Program of Evaluations and Experiments.

The purpose of the study is to assess the accuracy and coverage of administrative record data with respect to the 2010 enumeration of the US population. Several different sources of administrative records data will be researched. Then, administrative records will be used to determine a count of persons and addresses in the United States as of April 1st 2010. The administrative records will then be linked to the 2010 Census files to:
- Compare coverage of addresses in administrative records and 2010 Census files.
- Compare coverage of persons in administrative records and 2010 Census files.
- Compare demographic characteristics of persons in administrative records and 2010 Census files.

This comparison will include both federal records such as Internal Revenue Service information or Medicare Enrollment Information, as well as acquired commercial files. . We attempt to make administrative records look like the decennial census short-form file—addresses conform to those addresses in the Master Address File and persons have basic demographic information contained in the census short form (name, age, sex, race, and ethnicity).

Previous studies using administrative records focused on macro level population counts and demographic characteristics. The question of whether a complete administrative records census is possible – as a replacement for the census – is an interesting one but has too often been the beginning and the end of the discussion. Multiple analyses compare population counts and distributions of age, race, ethnicity, and sex as found in administrative records versus Census 2000, the Census 2000 coverage study--Accuracy and Coverage Evaluation, and Demographic Analysis. However, no study explores the characteristics of individuals absent from administrative records on a micro level that includes state and even smaller regions on a nationwide basis. Past studies also fail to closely examine the characteristics of individuals found in administrative record sources, but not in the Census for smaller geographic areas. Moreover, no previous assessments have analyzed how well the characteristics of people found at various stages within the census process, such as in the Non Response Followup (NRFU) operation compare with the Census Bureau's own administrative record systems. Likewise, these studies left open questions surrounding the accuracy and coverage of the characteristics associated with traditionally undercounted populations such as children and racial minorities. The 2010 Census Match Study aims at bridging these gaps and answering some of the outstanding questions not yet answered by the above research, i.e., what is the coverage for NRFU cases completed after week 3 of the operation or after the 3rd attempt (at the household level.

Conclusions

The U.S. Census Bureau is committed to new and different ideas for planning and conducting the 2020 Census. In order to make informed decisions about what options should comprise the ultimate 2020 Census design, a wide variety of options must be thoroughly researched and tested especially the full implications of using administrative records. Early analysis results indicate that the only known method of substantially reducing the cost of the enumeration is to substantively reduce the NRFU workload by using administrative records. However, there may be a quality tradeoff depending on what is learned about coverage and error structures in administrative records. The Census Bureau is committed to transparency and inclusion throughout the 2020 Census planning process and looks forward to partnering with experts both inside and outside the Census Bureau.