PREPARED STATEMENT

ROBERT M. GROVES, DIRECTOR
U.S. CENSUS BUREAU

CENSUS: LEARNING LESSONS FROM 2010, PLANNING FOR 2020

BEFORE THE COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS
SUBCOMMITTEE ON FEDERAL FINANCIAL MANAGEMENT, GOVERNMENT INFORMATION, FEDERAL SERVICES, AND INTERNATIONAL SECURITY
U.S. SENATE

6 APRIL 2011

Mr. Chairman, Members of the Subcommittee, thank you for inviting me to discuss the lessons learned from the 2010 Census. U.S. Commerce Secretary Gary Locke and I were committed to ensuring a successful 2010 Census and we are both proud that the 2010 Census was completed on time and under budget. With that backdrop, now is an opportune moment to further examine the results and operations of the 2010 Census, as it can help put the first results of the census in context. This discussion also allows me to report on the foundation for the U.S. Census Bureau's activities for the next several years: focusing on innovating to provide high quality statistical information whenever possible at lower costs.

The Census Bureau has a formal program for decennial census evaluations and assessments, which is called the 2010 Census Program of Evaluations and Experiments. There is also the Census Coverage Measurement program, which is also known as a Post Enumeration Survey. The Post Enumeration Survey provides sample-based estimates of undercounts and overcounts. Over the next two years, these programs will provide extensive analyses for the entire range of census operations, from address canvassing through enumeration to dissemination.

Today, however, I will share some information from early empirical performance indicators and a more qualitative lessons-learned process. They are a large part of the perspective on building a better 2020 Census. So far, I am encouraged by the early indicators.

EARLY INDICATORS OF THE QUALITY OF THE 2010 CENSUS

There are three types of performance indicators used to evaluate censuses around the world:

1. Process indicators of how the operations progressed;
2. Results of alternative ways of measuring the population; and
3. Results of the Post-Enumeration Survey.

While each of these indicators is imperfect, they together provide the full set of information we now have in part to answer the question, "How good are the 2010 Census counts?"

Process Indicators. I would like to review several process indicators: the mail return rate, the proxy rate from non-response follow-up, a few indicators from the data capture operations, and item missing data rates.

The mail return rate in the 2010 Census matched the mail return rate for the short form in the 2000 Census. When one combines the long and short form experience of 2000, the 2010 Census mail return rate was actually better, with a higher percentage of households returning their forms. I think several efforts contributed to this particular success, including the mailing strategy, the partnership program, and advertising. Moreover, the mail return rate provided a great platform for the remaining census enumeration operations.

The second process indicator is the proxy rate, which represents the number of households from which we did not receive direct information, but rather information from a neighbor, building manager, or other knowledgeable person. This occurred after we visited the housing unit repeatedly as many as six times, but were unable to make contact with the resident. The proxy report rate for the 2010 Census non-response follow-up operation was approximately 22 percent, compared to 17 percent for the 2000 Census. This is worse than 2000, but we still have many other quality measurements that will tell us more about the quality of the data.

However, I am pleased to report that there are some positive indicators from our data capture operations. We processed over 3.3 billion information fields. We rely on information technology to read these fields, and only when there is an issue do we send the form to a clerk to examine the field and enter the information. This human check is costly, and an important objective for data capture was to reduce our reliance on clerks, and we were successful. The rate of successful machine reads was slightly higher for this census than it was for the 2000 Census. In 2000, the machines failed to read approximately 0.1 percent of the checkboxes. For the 2010 Census, the machines only failed to read approximately 0.03 percent.

For the write-in fields, the difference between 2000 and 2010 is even larger and reflects the increased sophistication of the optical character reading technology. In 2000, the machines failed to read 16 percent of the write-in answers; we reduced that to 12 percent in 2010. We saved $36 million in staff costs because fewer forms were sent to the clerks.

Overall, the percentage of persons for whom usable response data was obtained from either the mailback or a face-to-face interview was similar in the 2010 Census (99.61 percent) as in 2000 (99.57 percent). At a state level, we reported that more states exceeded their usable information rate obtained in 2000 than those doing worse. At counties and census tracts levels we have similar positive results. It is very desirable to find low variation over states, counties, and census tracts on these indicators. The variation at those levels is lower in 2010 than in 2000.

We have just obtained information about response rates to specific questions on the census and can make some comparisons to the 2000 Census. For the questions on age, sex, race, and

Hispanic origin, we saw a slight decrease in the percent of the population that answered the questions versus in 2000. For example, the item with the largest decline is Hispanic Origin; reported by 94.0 percent of persons in 2000 but by 92.8 percent in 2010. This decline is not desirable; it is a combined effect of greater proxy cases and failure of householders to fill out an item; but the differences are small.

Finally, as the committee already knows through our periodic briefings during data collection, most of our field operations ended early and under budget. Out of an approximately $7 billion budget in FY 2010, we completed all operations with $1.87 billion left over. We were lucky in that the contingency funds we had set aside for hurricanes, floods, epidemics, and other possible disasters were not needed. We think a good part of the savings was because the wonderful staff we were able to hire in these times of high unemployment worked more efficiently than our productivity models suggested. We thank the hundreds of thousands of staff who did a great job for the country.

Alternative Ways of Measuring the Population. Another quality indicator is whether we obtain results similar to the 2010 Census using alternative methods. Demographic Analysis is the classic method of establishing an independent estimate of the nation's population and the Census Bureau has conducted this analysis since the 1970 Census. We delivered the results of the Demographic Analysis on December 6, 2010, two weeks before the census apportionment data were released. We used the data on births and deaths, international migration, and Medicare enrollments to provide national totals that are broken down by age, gender, and Black/non-Black status, and for the first time, the Hispanic population under age 20.

For the 2010 Demographic Analysis, we released five series of estimates. In January 2010, we met with prominent demographers from academia and found a lack of professional consensus on the size of the immigrant population. Therefore, for this reason, we released five alternative estimates based on different assumptions on immigration. Moreover, for the first time the Census Bureau has provided these alternative estimates prior to the census apportionment release in an effort to be completely transparent about their independence from the 2010 Census counts.

We presented a range of estimates from a low of 305 million people to a high of 312 million people. The official 2010 Census count released on December 21, 2010, was 308.7 million, just a few hundred thousand more than the middle Demographic Analysis estimate —in fact, within .09 percent (see Figure 1). In other words, the estimates were virtually the same. As with science in general, when independent methods yield the same result, we are more confident in the results.

Figure 1

We are also comforted by another estimate, and that is the number for the total population from our official Population Estimates program as of April 1, 2010, which is also represented in Figure 1. The population estimates are based on the 2000 Census, as the benchmark, and are updated on an annual basis to reflect births, deaths, and migration. These estimates are, however, distinct from Demographic Analysis because they are benchmarked to the previous census, and are not used to measure the accuracy of the census. They are another way of estimating the nation's population, apart from the current decennial census. And once again, like Demographic Analysis, the 2010 population estimates were quite close, within 0.8 percent, to the 2010 Census.

The population estimates provide additional information about regional and state population totals, which along with information from the American Community Survey, can tell us even more about the changes that occurred in the states over the past decade, especially with regard to migration. In fact, these data may be viewed as corroborating evidence to the regional shifts we observed in the official census.

When we drill down to states and counties with the population estimates, we see smaller variations for 2010 than in 2000. For instance, when we compute a difference between the census count and the population estimates at the state and county levels we see smaller mean differences at those levels and less variation over the geographical units. These are desirable results.

The Post Enumeration Survey. The Post Enumeration Survey is an independent large-scale sample survey of approximately 170,000 housing units. The in-person interviews were conducted in August through October of 2010. We compare the results of this survey to the

results of the census to produce measurements of census overcounts, undercounts, and components of census coverage by certain census operations, such as the non-response follow-up and coverage follow-up operations. We will also produce estimates of coverage for the major race groups and for Hispanic Origin.

The final results from the survey will be available in 2012. However, I do have some preliminary information from its address listing and personal interview operations. First, as part of this survey, we conducted an independent address listing to create the frame, and this listing was later compared to the preliminary 2010 Census address list. When we compared the Post Enumeration Survey address list to the census address list, we had a 96.5 percent match compared to a 91.4 percent in 2000, whether or not they were correctly enumerated. Second, the preliminary calculation of the percentage of correct enumerations for the 2010 Census (the percentage of those who should have been enumerated who actually were enumerated) was 96.3 percent, compared to 89.9 percent for the 2000 Census. The percentage of duplicate enumerations for the 2010 Census is 0.79 percent compared to 1.05 percent in the 2000 Census. When compared to the 2000 Census, these are improved results that point toward better overall coverage in the 2010 Census.

Finally, however, there are also some findings with regard to the match rates for urban and rural areas. Just as with the 2000 Census, there is a higher match rate for urban addresses than for addresses in rural areas. Yet, we have seen a reversal in the rate of correct enumerations among the matched units between these areas in 2010, with a higher rate of correct enumerations for the urban areas than the rural areas. We are only just beginning to delve into the results so it would be premature for me to speculate. But, this observation is an example of something that we will pursue as part of the Post Enumeration Survey evaluation.

In summary, the vast majority of the initial quality indicators are coming in positive for the 2010 Census. These are all preliminary indicators; we will be learning much more during this year and next.

BUILDING THE CENSUS BUREAU OF THE FUTURE

I want the committee to know that although recently many of the resources of the Census Bureau were devoted to conducting a successful 2010 Census, we have also been engaged in a variety of organizational change initiatives. The rationale for these is simple:

1. The difficulties of measuring the busy, diverse, and independent American society and economy are increasing every year (that is, it costs more money to do the same things we have done for years);
2. The demands by American business, state, local, and community leaders for statistics on their populations are continually increasing;
3. New technologies are being invented almost daily that can be used to make it more convenient for the American public to participate in these efforts to inform us about the status of the country;

So, we conclude that our current "business model" of collecting social and economic data face many challenges over the long run. We at the Census Bureau know that we must innovate if we are to remain useful and relevant to the country. Further, we know that this innovation is not likely to be funded by added resources; we must become more efficient.

I want to tell the committee what we have been doing to build a Census Bureau that can supply the country credible and cost-efficient economic and social statistics in the future.

Employee-led Cost Savings. Last year, the Census Bureau established the Improving Operational Efficiency program, where we asked all of our employees for proposals to improve efficiency by saving money or staff time. We found that this process was liberating and empowering for the employees. For the FY2010 program the employees provided more than 650 proposals on a wide range of topics, from human resources management to survey operations to IT functions. My deputy and I reviewed every proposal; and an executive leadership team made final selections. I am pleased to report that many of those proposals were so compelling and nearly costless that we simply said "yes" and empowered employees to proceed.

We also received proposals that were much more comprehensive, needed seed funding, and have the potential to fundamentally shift our business processes. We asked employees to prepare a business case for twenty of these proposals, and we have selected twelve of those to move forward into development, which means that we are committing resources to their development. These proposals range from the development of an integrated logistics support system for the field data collections; to the development of a database with contact information for building managers that will enable us to reduce the number of visits to multi-unit buildings; and to the implementation of an enterprise-wide human capital management system. Each of these projects will pay off its initial investment in cost savings within three years. This will allow us to continue to invest in innovative and efficient business practices that will help us better meet the nation's evolving data needs.

Survey Client Cost-Saving Teams. As this committee knows, the Census Bureau collects economic and social survey data for many other federal agencies; we are an important hub for the federal statistical system. Thus, our costs of data collection affect the costs of other agencies. We have an obligation to those clients to ensure that we are cost-effective. Over the past year, we have developed ideas in collaborative teams with our clients for reducing the cost of surveys while maintaining high quality data. As a result, we are transforming our method of tracking costs during data collection, increasing the transparency of our field activities, re-examining the management infrastructure for our surveys, increasing the active analysis of process data to improve operational efficiency, and improving our cost-modeling expertise. Real savings are possible, but involve changes we must make in concert with our clients.

Building Organization-Wide Solutions; Attacking the Silos of the Organization. Every large organization faces the threat of quasi-independent subunits not working together, but instead jockeying for power and resources. We have recognized this problem at the Census Bureau. There are three areas where we have mounted initiatives to minimize this organizational

weakness: the human resources initiatives, the IT enterprise architecture initiatives, and the structural reorganization of the Census Bureau.

Human Resources Initiatives. We want to make the Census Bureau the best place to work for statisticians, economists, computer scientists, geographers, and other staff. We want a diverse, creative, enthusiastic staff to experience new challenges throughout their career. We have instituted a "corporate hiring" program to assure entry-level statisticians that they can move around the Census Bureau, building their human capital and contributing to diverse work environments, instead of being assigned to one area of the Census Bureau for much of their career. We have instituted a Senior Executive Service mobility program to move innovation invented in one silo to another. In all communications with staff we are emphasizing both the personal and organizational benefit of career moves across the different subunits of the Bureau.

IT Enterprise Architecture. We are instituting key reforms in Information Technology (IT) management. Our efforts reflect a greater emphasis on the Internet and cloud computing to support internal operations and public outreach, as well as efforts to consolidate IT resources and establish enterprise solutions. For example, we are in the process of consolidating two major data centers into one center, a center that will support not only the Census Bureau's activities but also those of other Department of Commerce agencies, such as the International Trade Administration.

In addition, we consolidated 52 different data storage systems into seven systems, resulting in immediate and long term cost savings, as well as reductions in maintenance and contractor support. In FY11, these cost savings are $1.0 million. We have also launched a technology innovation center that will provide a laboratory for us to develop and test new technologies. In fact, we are already testing new technologies that will support telecommuting in a secure and cost-effective manner.

We have also expanded our Internet data collection capabilities. By the end of the year, we hope to provide an Internet response option for 60 surveys, allowing approximately 900,000 respondents the opportunity to provide their responses online.

Finally, we are utilizing a combination of census private cloud and public cloud solutions, allowing us to leverage commercially available software and systems, which, in turn, allows us to avoid the cost and delay of establishing these environments ourselves. One public cloud included the Akamai Content Delivery Network. Utilizing this cloud platform allowed the Census Bureau to stand up this capability in less than a month as opposed to it taking 6 months, improved our security posture and enabled the Census Bureau to save $1.5 million.

Census Bureau Reorganization. I thank the Congress for approving a structural reorganization to facilitate all of the initiatives above. This reorganization has given us the ability to reshape our research directorate to perform mission critical research, construct prototype innovation, and collaborate with program divisions to implement better and cheaper methods. It has created a new directorate for the 2020 Census, which will grow over time to fit the functional needs of the 2020 design. Also, we have constructed a multidisciplinary 2020 Census Steering committee of senior executives, to govern the process of integration of IT, Budget, Field and Research

directorate staff with the 2020 Census planning. Finally, we have created a Risk Management Staff that reports to the Director's office, to oversee ongoing review of key programs, and to spread the culture of risk management throughout the full organization.

PLANNING FOR THE 2020 CENSUS

We want to use a reinvigorated, more integrated, more operationally efficient Census Bureau to build a strong 2020 Census. Some of the guiding principles of the 2020 Census planning effort result from lessons learned in the 2010 Census.

Lesson 1: The multi-decade cost increase of the decennial census must be halted. Hence, the first principle for 2020 is that we are attempting to design a 2020 Census that costs less per housing unit than the 2010 Census, while maintaining the quality of the results. We have looked at the cost trend of the last several decades, and we know that this trend is unsustainable. One of the ways we can do this is to update the Master Address List, used as the basis of the census, continually in small steps rather than mount an expensive exercise at the end of the decade. Another way we can do this is to maintain our collaboration with over 257,000 partnership organizations throughout the decade using modern communication tools instead of mounting an independent partnership operation later in the decade. The most important cost savings, however, must focus on the non-response follow-up operation.

Lesson 2: Traditional non-response follow-up procedures are inefficient and costly. We must make the census more convenient to diverse groups in society. Thus, the second principle is that the 2020 Census will be a multiple-mode census, using mail, telephone, internet, face-to-face, and other modes as they emerge. We have to move beyond the mailback questionnaire and the personal interview. We need to ensure that the response options for the census reflect the communication platforms that people are using. The only way to reduce the size of the expensive field non-response follow-up operation is to increase the convenience of responding to the census.

Lesson 3: Systems development that requires first-use perfection must be abandoned. In the 2010 Census we were developing critical systems weeks before their use; important weaknesses were discovered in the early days of production. The third principle is that we need end-to-end tests of production systems, which use all subsystems in the integrated form needed in the production phase. New system capabilities will be developed in an incremental and modular fashion so that users have a chance to test and evaluate mission critical systems well before they are deployed to production environments.

Lesson 4: Too few of the system and procedure developments of the 2010 Census were designed to have residual benefits to other Census Bureau data collections. All of the valued demographic and economic data we produce use operations similar to those of the decennial census. There were too few plans to utilize the systems built for the decennial to enhance the efficiency of our many other survey operations. As a result, the large investment benefited only the decennial program, not the bulk of the Census Bureau. Thus, the fourth principle is to develop systems within similar survey production environments within the Census Bureau, test and enhance them

repeatedly over the decade, ramp them up for use in the 2020 Census, and then continue to use and enhance them in our ongoing surveys. We plan to use the American Community Survey as the chief test-bed for 2020 Census systems development.

Lesson 5: The short form and replacement questionnaire provided cost-saving benefits to the 2010 Census. With the ongoing American Community Survey, the Census Bureau is providing the needed socio-economic, housing, occupation, and commuting data that are important to local communities. Utilizing the decennial census to mainly focus on the key reapportionment and redistricting purposes was wise. Thus, the fifth principle is to build on the success of the reduced burden of the 2010 short form.

Lesson 6: A small number of large tests create intolerable risks for the Census Bureau. The sixth principle is to mount many, small tests throughout the decade. We are committing to a faster cycling of ideas and testing, relying on a lot of small tests versus a small number of large, expensive tests. For example, although we cannot know the full features of the Internet option for the 2020 Census, we will have repeated tests of Internet census measurement throughout the decade, using platforms that will increasingly resemble those available in 2020.

Lesson 7: In my professional judgment, the voluntary partnerships with over 250,000 local and national organizations, coupled with a paid advertising campaign successfully improved awareness of the coming Census. We have empirical evidence of the increasing awareness of the Census as the partnership and advertising campaign rolled out. Thus, the seventh principle is that we would like to keep these relationships warm, to seek input from these groups to inform the 2020 campaign, and to return to them Census Bureau data useful to their organizations.

Lesson 8: Updating of the master list of addresses using the Postal Service list produced a stronger list in 2010. We want to build upon this success this decade. One critical component is the Geographic Support System Initiative (GSS), an integrated program to improve address coverage and provide continuous spatial feature updates, as well as enhanced quality assessment and measurement. Geospatial information undergirds all of the census and survey programs, making it a fundamentally important contributor to providing high quality data. This interdivisional effort works across silos, drawing on support from fifteen divisions within the Census Bureau, including Acquisitions, American Community Survey, Decennial Management, Geography, and Field, as well as the Chief Information Officer. GSS is also an interagency effort, as we have asked the U.S. Geological Survey, the National Geodetic Survey, and the U.S. Postal Service to participate on the working teams. As a part of these efforts, the Census Bureau hopes to initiate programs to work with partners, both public and private, to receive addresses continuously throughout the decade, and to update our systems so that we can more fully leverage GPS and GIS technologies and resources.

We are incorporating these principles into our planning and management in order to most efficiently focus on increasing the convenience of census participation to the diverse groups within our society. Further, in the Fiscal Year 2012 budget request, the Census Bureau is requesting $67 million to fund the first of a three-year research and testing phase of the 2020 Census. This initiative is designed to identify new decennial census systems and methods which will provide cost savings while maintaining high quality of census data and results.

We know of no single method of collecting census data that is optimal for all residents of the U.S. Some residents have told us they do not want people visiting their home; some residents told us to use information they have already provided in other government forms; some residents want to use the Internet at any time of the day on any device they favor to fit their lifestyle; and some want to speak by telephone to someone guaranteed to speak their language and understand their sub-culture. By making the census more convenient, we hope to reduce the size of very expensive field follow-up activities. This is the most expensive part of the data collection, and by concentrating our efforts there, we want to achieve a quality census at a lower cost per household.

I want to thank the committee's members and staff for meeting with me over the past few months to hear updates about the progress of the Census. The oversight inquiries you have provided over the past months helped to make the 2010 Census better. Thank you.

I would be happy to answer any questions.