United Nations ECE/CES/ GE.41/2012/3

Economic and Social Council

Distr.: General
13 March 2012

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Fourteenth Meeting
Geneva, 24-25 May 2012
Item 2 of the provisional agenda
Full enumeration versus sample surveys

The introduction of the American Community Survey and its benefits to the 2010 Census

Note by the United States Census Bureau

Summary

In the United States, recent decennial censuses have collected basic demographic data for the entire population and detailed data for a sample of the population. The 2010 Decennial Census collected essentially only basic demographic data. Since 2005, along with basic demographics, the American Community Survey (ACS) has continuously collected these detailed social, economic, and housing data). The Census Bureau designed the ACS to meet this critical need for more frequent detailed housing, social, and economic characteristics at small levels of geography. This paper will describe how the introduction of the ACS did, in fact, benefit the conduct of the 2010 Census (by simplifying the census process and by providing timely information to support the enumeration). It will also include some ideas for how the U.S. Census Bureau may expand the benefits of the ACS on the Decennial Census as we look toward the 2020 Census.

I. Introduction

1. This document has been prepared by Frank A. Vitrano and Deborah H. Griffin of the United States Census Bureau.

2. In the United States, recent decennial censuses have collected basic demographic data for the entire population and detailed data for a sample of the population. The 1940

census was the first census to use sampling techniques as a means to collect additional data from a portion of the population (U.S. Census Bureau, 1979). The basic demographic information is needed for apportionment and redistricting. Detailed social, economic, and housing characteristics are the basis for the distribution of funding for many federal programs. The collection of these data is driven by the needs of the federal government but state, local, and tribal governments, the media, academia, and many non-profit and private sector organizations also rely on this rich data source. Over time, the Census Bureau and its stakeholders recognized that obtaining these data just once every ten years was not sufficient. The 2000 Decennial Census was the last census in the United States to include the combination of a full enumeration and a sample survey (the long form) to collect all of this information. The 2010 Decennial Census collected essentially only basic demographic data.1 Since 2005, along with basic demographics, the American Community Survey (ACS) has continuously collected these detailed social, economic, and housing data). The Census Bureau designed the ACS to meet this critical need for more frequent detailed housing, social, and economic characteristics at small levels of geography.

3. The introduction of the ACS provided two direct benefits to the conduct of the 2010 Census – simplifying the census by eliminating the need to collect detailed socioeconomic and housing data as part of the decennial enumeration and providing current data that facilitated aspects of implementing the 2010 Census. This paper will describe how the introduction of the ACS did, in fact, benefit the conduct of the 2010 Census (by simplifying the census process and by providing timely information to support the enumeration). It will also include some ideas for how the U.S. Census Bureau may expand the benefits of the ACS on the Decennial Census as we look toward the 2020 Census.

II. Simplification of the 2010 Census

4. In the early planning stage of the 2010 Census the Census Bureau summarized their strategy in U.S. Census Bureau (2001). Included in that document was the following summary:

The Census Bureau has concluded that in order to be successful, it must increase the currency of detailed housing and population data but must decrease the decennial census operational complexity. Therefore, the overall goal for the 2010 Census is to reduce census operational risks, improve accuracy, provide more relevant data, and contain cost. The goal will be accomplished by moving the two most detailed, and therefore, volatile functions – building the address system and collecting long-form data – from the 2010 decennial census data collection to ongoing, corporate initiatives and by developing and testing the 2010 design early in the decade by taking advantage of the opportunities offered by the MAF/TIGER and ACS.

5. This document identified potential cost savings and simplifications in many areas including printing and postage, mail response, administrative needs of the data collection offices, staff training, telephone assistance, and data capture. The Census Bureau considered some design options such as multilingual forms, handheld data collection instruments, targeted mailings, and internet response as feasible only with the elimination of the long form. While designers considered many of these options,, the Census Bureau did not implement all of these operations in the 2010 Census. However, the removal of the long form led to several important innovations.

[1] The 2010 Census included questions about tenure (own/rent) and occupancy status.

A. Mail Response

6. In Census 2000 mail response to the long form was 53.9 per cent, 12.5 percentage points lower than the 66.4 per cent obtained from addresses receiving a short form (Treat, 2003).[2] Census costs are very dependent on the success of obtaining responses by mail. In the 2010 Census, we estimated that every percentage point increase in mail response would save us $85 million (U.S. Census Bureau, 2010a). The Census Bureau used the concise marketing slogan of "10 Questions, 10 Minutes" to encourage all households to complete and return their census forms by mail. That slogan, and the response achieved by mail in the 2010 Census (63.5 per cent), was possible because there was no long form (LeTourneau, forthcoming). Outreach and promotion stressed how easy it was to respond to the census and did not have to address the complex issues of why the government is asking "all these intrusive questions." The 2010 Census included a targeted second mailing for initial non-responders in certain areas. This option was possible without the complexity of the long form.

B. Language Assistance

7. The creation and distribution of a combination English/Spanish questionnaire was possible due to the small number of questions on the 2010 Census form. The 2010 Census was the first census that mailed out bilingual (English/Spanish) questionnaires. Approximately 12.1 million housing units were in the geographic areas targeted to receive bilingual forms (Rothhaas, 2011). The Census Bureau produced and made available upon request other language materials such as language guides in 60 languages (plus Braille).

C. Questionnaire Printing, Labelling, and Assembly

8. In 2001, the Census Bureau estimated that the reduction in printing costs from eliminating the long form could be about 25 percent. This estimate meant that the total number of questionnaires that required design, testing, and printing would be essentially cut in half. The Census Bureau could dramatically decrease the complexity of the printing and mailing operations. Staff involved in the printing noted that in Census 2000 over 70 percent of the quality assurance resources involved the long form questionnaires. These forecasts regarding reduced costs for printing also extended to the total number of printing contracts. Without the long forms, we were able to reduce the total number of printing contracts. The long form accounted for 16 of the 86 contracts in 2000. Eliminating the 16 contracts for the long form left 70 Census 2000 printing contracts that were bundled into 20 contracts in 2010.[3]

9. The added complexities of labelling and ensuring that the appropriate sample addresses received long form questionnaires was no longer an issue for the 2010 Census. As a result, we streamlined the assembly and labelling operations for the 2010 Census compared to those used for Census 2000.

[2] We based these estimates, and the estimates for the 2010 Census, on the mail response rate at the point in time when the workload was identified for non-response follow-up.

[3] In Census 2000, the Census Bureau procured each questionnaire in a separate printing contract. In the 2010 Census, we bundled similar questionnaires into one contract, eliminating duplication efforts in contract writing, management, and quality control efforts.

D. Data Collection

10. In the early stages of planning for the 2010 Census, the elimination of the long form opened the door for data collection to consider the use of a "hand-held computer." Census managers expected this design option to simplify data collection, reduce data processing, and reduce overall census costs. Despite initial development of this technology, and successful use for address canvassing, the Census Bureau chose not to implement this technology for non-response follow-up in the 2010 Census.

11. Even without the hand-held computers during enumeration, staff responsible for planning and overseeing the Census 2000 and the 2010 Census data collection, specifically the nonresponse follow-up operation, noted the following advantages of not having a long form:

- There was less respondent burden with just short forms in 2010 that shortened interview time.
- It was feasible to gather information from a knowledgeable non-household member to meet the minimum data requirements of the short form questions. It was easier to gain compliance in 2010 with just a short form.
- Not having the long form in 2010 removed issues related to errors of the wrong respondent getting the long form. There was always the potential of an office person or field person accidentally giving the long form to the wrong respondent.

12. The local offices focused training for data collection on core data collection and enumerators learned important data collection strategies in lieu of learning about the full set of long form questions. Staff responsible for the training of non-response follow-up enumerators in 2010 noted that in 2000 we spread the enumerator training, including "live" fieldwork, across three days and increased this training to four days in 2010. Even though there was no long form training required in 2010, there were more mandated topics and time was required to collect fingerprints. If there had been a long form in 2010, the training would have exceeded four days. Given the huge workforce—we hired and trained about 600K people, an additional day of training could have easily cost close to $50 million.

13. Additional aspects of data collection include the preparation of materials and the check-in of completed work. Writing the numerous manuals and training guides was easier in 2010 since they did not contain instructions and procedures for completing the additional questions. There was no need to create long form job aids or questionnaire reference guides in 2010. This saved development time and reduced printing costs. Additional observations include:

- The assembly of materials for enumerators was easier without the long forms.
- The assignment preparation activities for each office were much simpler without having to insert long forms into specific assignments (labelling questionnaires) and provide additional instructions on how to create binders and insert long forms for sampling patterns.
- The review of enumerator-completed questionnaires was much quicker in 2010 without the more complex review required for long forms.

E. Data Capture and Data Processing

14. The long form, despite the small sample of all addresses asked to complete it, drove the data capture and data processing requirements in Census 2000. U.S. Census Bureau (2001) noted that, "Census 2000 was awash in paper, scanning 1.5 billion pieces of paper.

Sixty percent of that paper was the long form." In the 2010 Census, we streamlined the capture requirements. Due to the reduction in paper, we used only three processing sites in 2010, while we required four in 2000[4]. Fewer staff and less equipment were required. The vast majority of the data in 2010 could be captured using optical mark and optical character recognition software. We could dramatically reduce the need for keyers to capture written entries in 2010 with the elimination of the detailed questions from the long form, many of which required keying. Coding operations were also very limited due to the elimination of the long form in 2010.

15. We realized reductions in the complexity of data capture and data processing in both the planning and the implementation stages. Reduced risks allowed the focus to be on the critical processing required to produce key census estimates.

III. ACS Data Used by the 2010 Census

16. While removing the long form from the 2010 Census clearly resulted in simplifications, moving that data collection into an ongoing program provided additional benefits for the 2010 Census in the form of information to assist in planning and implementation. With the introduction of the ACS in 2005, we have essentially been collecting data every day of every year since. We pool completed interviews over the year to produce 1-year estimates for geographic areas with population of at least 65,000. In addition, we combine all completed interviews over three years to produce 3-year estimates for geographic areas with populations of at least 20,000. We designed the ACS to provide data for the smallest geographic areas (those comparable to the Census 2000 long form) by pooling together 5 years of data collection. The Census Bureau released the first 5-year estimates in December 2010, pooling data collected between January 1, 2005 and December 31, 2009. With the data released in December 2010, this information at the smallest geographic levels was not available to assist the 2010 Census. However, the 1-year estimates and the 3-year estimates were available and we used that information in various ways to support the 2010 Census.

A. Language Program

17. ACS information supported various aspects of the Language Program. The ACS includes the following three questions on language spoken at home:

- "Does this person speak a language other than English at home?"
- "What is that language?"
- "How well does this person speak English?"

18. The 2010 Census language program planning and implementation used information obtained from these ACS questions in the following ways.

1. 2010 Bilingual Form

19. The Census Bureau distributed bilingual questionnaires to specific collection blocks in areas where we determined that there might be a need for Spanish language assistance. We used data from the 2005-2007 ACS to identify tracts (small geographic areas with a population size of between 1,200 and 8,000 people) to receive the bilingual questionnaire.

[4] The reduction in processing sites was also a consequence of more efficient scanning and better optical character recognition contextual analysis.

We identified households as needing Spanish language assistance if at least one adult (age 15 or over) in the household reported speaking Spanish and not speaking English "very well". Tracts where at least 20 percent of the occupied housing units needed Spanish language assistance received bilingual forms.

2. 2010 Census Language Selection

20. The five primary non-English languages for the 2010 Census were Spanish, Chinese, Korean, Vietnamese, and Russian. This was determined by defining language need households based on ACS data on Housing Units with no person age 15 or older who speaks English "very well." Spanish, Chinese, Korean, Vietnamese, and Russian had an estimated total of more than 100,000 language need households and the Census Bureau defined these as our primary languages. The same data supported the identification of an additional 45 languages that needed assistance, and we produced Language Assistance Guides in these languages.

3. 2010 Direct Mail Postcard

21. The Census Bureau designed a Direct Mail postcard that contained contact information for our Telephone Questionnaire Assistance Centers and directed callers to operators in English, Spanish, Chinese, Korean, Vietnamese, and Russian. The goal of the Direct Mail postcard was to provide information about the availability of this assistance to potential users of it. The process to determine which households would receive the Direct Mail postcard began with 2006-2008 ACS 3-year estimates. We specifically looked at the response data from the "Language Spoken at Home" question to identify tracts where at least 10 percent of the households indicated speaking Chinese, Korean, Vietnamese, or Russian. Our 12 Regional Offices reviewed the identified tracts and provided additional information based on their local knowledge. We linked the final selected tracts to their corresponding postal codes. We sent all housing units within the identified postal code areas the Direct Mail postcard.

4. Telephone Questionnaire Assistance Planning

22. We designed the Telephone Questionnaire Assistance operation for the 2010 Census to provide support in the five primary non-English languages of Spanish, Chinese, Korean, Vietnamese, and Russian. To appropriately staff the Telephone Questionnaire Assistance Centers with sufficient bilingual staff, managers used ACS data to plan workload estimates in each of the languages.

B. Field Logistics and Planning

23. While the Census Bureau planned most of the 2010 Census operations and procedures nationally, the actual enumeration is a local event. We opened 494 temporary Local Census Offices (LCO) to manage the field operations during the census. One of the responsibilities of the LCO managers was to develop integrated tract action plans that included elements from recruiting, partnership, and operations (enumeration strategies such as paired or blitz enumeration) to assist with planning for enumeration in hard-to enumerate areas. To assist them with developing these plans, we created the "Planning Database." The Planning Database assembles a range of housing, demographic, and socioeconomic variables correlated with mail nonresponse and undercount. The database was used to identify hard-to-enumerate areas and areas with potentially low mail response rates, and for recruitment purposes Identifying hard-to-enumerate areas in advance of the census enables focal points for outreach and special enumeration methods (such as using local partners to

promote the census and performing paired enumeration or employing the use of cultural facilitators). As soon as ACS data became available, we added it into the database.

24. Another advantage of having the ACS program in advance of the Census was that we had a well-trained ACS staff that we could, at certain times, move into positions to help assist with the census. Although, based on its size, the census requires the hiring of many new staff, being able to strategically place experienced staff among the new staff provides some quality benefits as the fieldwork goes into full production.

C. Communications Campaign

25. ACS data played a major role in the 2010 Integrated Communications Program. The Census Bureau implemented a more comprehensive research program that tied traditional marketing research to actual response and response characteristics from Census 2000 and the ACS (U.S. Census Bureau, 2010b.) In preparing for the 2010 Census Communications Campaign, we needed to answer several questions. First, what are the characteristics of the hard-to-count population? Second, what are the obstacles to counting these populations? Third, how do research results translate to the 2010 Census Communications Campaign? We ultimately used a variety of sources to help answer these questions. Our intent was to make sure that the 2010 Census Communications Program was more data driven than what we had done for Census 2000. A key component of answering these questions was the ACS. The Census 2000 campaign relied on surrogate information such as civic participation models. For the 2010 Census, however, having the actual behavioral data from Census 2000 and the more recent ACS provided more direct evidence about who responds and who doesn't respond to a survey and what their characteristics are. This information became very useful as the Communications Staff planned their messages and where to direct those messages.

D. Data Capture and Processing

26. Through the Telephone Questionnaire Assistance Program, respondents could request a questionnaire in one of the five primary non-English languages. Again, we used ACS data to help model how many completed forms in these languages we expected to receive for data capture.

27. We used ACS data to support data processing in two distinct ways. First, when 2010 Census forms were missing a response to the race or Hispanic Origin questions, we sometimes used ACS responses, along with Census 2000 responses, to impute the missing information. That is, if we had a reported response from the same household in a prior enumeration, we chose to use that value rather than using a value from a traditional neighbourhood hot-deck. Second, we were able to use the ACS processing system infrastructure to support the processing of the census data coming from the Island Areas. The Island Areas (American Samoa, Guam, the Commonwealth of the Northern Marianna Islands, and the U.S. Virgin Islands) did not have a short-form census. One hundred percent of the population in these areas received a more traditional long-form. The Census Bureau does not conduct the ACS in these locations. In order to facilitate post-data capture processing of these data that looked more like ACS data than 2010 short form data, the ACS Office offered the use of their system environment which had already been certified to be a secure processing environment. The use of this environment reduced time and effort with setting up a new secure processing environment that would have needed to go through stringent security checks before it would be certified to be secure.

E. Experimentation Program

28. As part of the 2010 Census Experimentation Program, we conducted an alternative questionnaire experiment. Small samples of the population received slightly altered census forms, in order to test some potential content changes in a true census environment. A portion of this experiment tested various ways of asking the race and Hispanic Origin questions, which remain a challenge for some segments of the population. ACS data from 2005-2007 were used to derive stratification and sampling intervals to ensure a sufficient oversampling of specific race and Hispanic Origin groups.

IV. Some early thoughts on using the ACS to assist with the 2020 Census

29. As we begin the process of planning the 2020 Census, we will build on the successful use of the ACS in the 2010 Census. There are several ways that we are considering expanding our use of the ACS to assist with the next census. Some of these ideas were opportunities we considered during the 2010 Census, but ACS data for the smallest areas were not available. Other ideas have originated with our specific early planning for the 2020 Census.

A. Field Logistics and Planning

30. As mentioned earlier, the Planning Database is an important tool for identifying local strategies to ensure a successful enumeration. Unfortunately, we were geographically limited in our use of ACS data, as we had not yet produced 5-year estimates. Now that the ACS is annually producing 5-year estimates down to the tract and block group level, we will have an opportunity to study more current information regarding language support, income, household composition, and variables that may be predictors of hard-to-enumerate populations. This should greatly expand intelligence at local levels on what they will be facing when enumeration begins.

31. During the 2010 Census, the Census Bureau temporarily opened 30,000 Questionnaire Assistance Centers and an additional 10,000 locations where respondents could pick up questionnaires if they did not receive a form. We staffed the Questionnaire Assistance Centers part time, so respondents could walk in and get help with filling out the census form. In general, we attempt to place these sites in locations where we anticipate respondent challenges with participating in the census (e.g., literacy, language challenges, and new immigrants). Outreach to stakeholders and local knowledge were the only available information to assist with placing these sites. To the extent that we use locations such as these in the 2020 Census, rich ACS data will be able to provide us with real data on where we should prioritize our efforts.

B. Using the ACS to Facilitate Testing

32. The ACS has a robust evaluation and experimentation program. With this ongoing program, we are continuously looking for ways to improve the data collection effort and the accuracy of the data we are collecting. The Census Bureau sees the ACS as a potential testing vehicle for census studies. Right now, there is serious collaboration that has been taking place between ACS staff and the Census staff on Internet testing. The two programs are learning from each other and helping each other. But we also want to think about the possibility of developing systems that can be used by the ACS or other demographic

surveys prior to the 2020 Census and how that would reduce risk for the 2020 Census if we are able to use systems that have been well-tested in a production environment.

C. Tailoring Response Modes

33. In order to maximize self-response in the census, we plan to research more customization of response modes by geography and demography. The customization of data collection approaches based on geography has been part of our planning for several decades. Once we introduced sending questionnaires through the United States Postal Service to some households in the country, we found ourselves tailoring enumeration approaches based on how rural or urban an area was, as well as whether the address information we had was sufficient for delivery by the United States Postal Service. The introduction of the bilingual questionnaire in the 2010 Census is another example of tailoring response modes. The introduction of the internet and the ability to provide language support in more languages are two areas of likely expansion in the area of tailoring response modes. ACS data throughout the decade will inform us of any trends with self-response. The ACS plans to introduce internet response in 2013. With the ACS's multi-mode approach to completing the survey, we should have valuable information that could help us predict appropriate approaches for different demographic and geographic groups.

D. Re-using ACS Data

34. In order to save money while maintaining quality, we plan to look at the use of pre-existing data sources such as administrative records or data from the ACS or 2010 Census to support data collection. We can accomplish this by expanding our use of existing data, such as the ACS, in our imputation methods.

E. Developing Organizational Solutions

35. We believe there is an opportunity to save costs and reduce risks if we create corporate solutions that can be tested and used in production prior to the 2020 Census. This would also help us avoid the development of throwaway systems that are used once for the census and then become obsolete. Given the similarities in the ACS and census data collection modes, the ACS is a very good starting point for building data collection and data processing systems in a production setting prior to their need during the Census. Too often, we have built systems that we cannot sufficiently test in advance of the census, increasing the risk of failure or, at least the risk of serious problems.

V. Conclusions

36. The ACS is releasing timely information on an annual basis – providing the data user community with much more information that they ever had when the source was the decennial census.

37. The elimination of the long form sample from the 2010 Census simplified census planning and implementation. High mail response and final census response rates are a very likely consequence of the minimum burden placed on households in the 2010 Census. Cost-savings were also possible without the need to collect and process these additional data.

38. ACS data were useful in planning components of the 2010 Census, particularly with the language program and with areas of the program that require different approaches at the local level.

39. With the ongoing ACS, we believe there are additional opportunities to simplify and improve the efficiency of the 2020 Census while maintaining our high level of quality.

VI. References

LeTourneau, E. (forthcoming). Mail Response/Return Rates Assessment. DSSD Census 2010 Memorandum Series #O-C-13.

Rothhaas, C. et al (2011). 2010 Census: Bilingual Questionnaire Assessment Report. 2010 Census Program for Evaluations and Experiments.

Treat, J. (2003). Response Rates and Behavior Analysis. Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No. 11, TR-11. U.S. Census Bureau.

U.S. Census Bureau (1979). Twenty Censuses – Population and Housing Questions 1790 – 1980. U.S. Government Printing Office, Washington, DC.

U.S. Census Bureau. (2001). Potential Life Cycle Savings for the 2010 Census.