UNITED NATIONS **E**

Economic and Social Council

Distr.
GENERAL

CES/2001/21/Add.6
15 May 2001

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Forty-ninth plenary session
(Geneva, 11-13 June 2001)

SMALL-AREA DATA FROM THE AMERICAN COMMUNITY SURVEY

Supporting paper submitted by the United States[1]

I. BACKGROUND

1. The United States conducts a decennial census of population and housing in years ending in zero. Most housing units receive a "short form" questionnaire, collecting very little information other than the name, sex, age, and race of the residents. A random sample of about one-sixth of the housing units receive instead a "long form" questionnaire, which also collects a variety of demographic and economic characteristics, such as income and poverty, educational attainment, disability status, industry and occupation, and housing characteristics. This long form sample is the main source of detailed socioeconomic information below the national level, for states, counties, municipalities, neighborhoods, and American Indian Reservations.

2. The U.S. Census Bureau has plans to replace the long form in the 2010 census, by a large ongoing intercensal survey, called the American Community Survey (ACS). This is part of a package of planned improvements in the census process, many of which will be facilitated by having only a short form in the

[1] Prepared by Charles H. Alexander, U.S. Census Bureau. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

census.

3. The need for more frequent census-type information for Acommunities,@ both geographic areas and population subgroups, has been discussed at least as far back as a 1941 proposal for an Aannual sample census.@ Interest in updating the census description profiles used in allocating Federal Government funds has been renewed more recently (Melnick, 1990). There has been increased interest in measuring change at the state level to evaluate programs such as welfare reform, and in measuring changes at smaller geographic levels for planning by local and tribal governments.

II. THE PLANNED AMERICAN COMMUNITY SURVEY DESIGN

4. The proposed ACS uses the Arolling sample@ design proposed by Kish (1990). The design uses a Arolling sample@ of approximately 250,000 addresses, spread across the entire Master Address File each month. A different sample of addresses is included in each month's mail-out, so that the sample cumulates to about a 2.5 percent sample over the course of a year, and a 12.5 percent sample over five years. The basic survey estimates are annual averages of the number of people or households with specific characteristics. People are included at their current residence at the time their data are collected. As with recent census long form samples, the ACS will have a higher sampling rate in small government units, and a lower rate in large census tracts. No address will be in sample more than once in a five-year period.

5. The Census Bureau is proposing that the main replacement for the census long form data will be the series of 5-year moving averages from the ACS for communities of all sizes. For larger communities, or for any analyses where sample size can be sacrificed for greater timeliness, annual average estimates or 3-year averages can be used.

6. The ACS is planned to start nationwide in 2003. In 1999 the program began in 36 Acomparison counties@, using the ACS design with a five percent annual sample, so that results can be compared to Census 2000. Some of these counties were included in previous years of the ACS A1996-1998 Demonstration Period,@ with three counties having been in sample since 1996.

7. The ACS is a mail survey with follow-up of nonrespondents by telephone and, for a subsample, in person. The data collection for a given monthly sample of addresses takes place over a three-month period. In the first month, there are repeated mailings. In the second month, there is an attempt at a telephone interview for addresses where no mail form has been returned, and where a telephone number can be obtained. In the third month, a random sample of one-third of the remaining nonrespondents are contacted in person.

III. THE ACS SAMPLING FRAME -- A MASTER ADDRESS FILE

8. A continuously maintained address list for the entire nation is a prerequisite for an unclustered sample survey like the ACS. Traditionally, the United States has not had a national list of residential addresses, except at the time of the census. During the 1990's, the Census Bureau began maintaining a Master Address File (MAF) by updating the 1990 census list using new mailing addresses, from a file of mailing addresses maintained by the U.S. Postal Service. The MAF is linked to the Bureau=s geographic database known as the Topologically Integrated Geographic Encoding and Referencing (TIGER). In the two years before Census 2000, the coverage of the MAF was completed by additional block canvassing and address listing operations. Local and tribal governments were invited to compare the MAF to their own address lists to help further

improve coverage.

9. The Census Bureau is planning further enhancements to the coverage of the MAF. These will include detection of changes using satellite imagery, increased use of sources of addresses other than the U.S. Postal Service, and intercensal opportunities for local and tribal governments to provide updates to the MAF. Also, ACS field representatives will be able to update the MAF and TIGER when they find changes in the field, using a computerized system that lets them Amap spot@ new addresses into the TIGER database.

IV. OTHER SOURCES OF SUB-NATIONAL INTERCENSAL DATA

Post-censal Demographic Estimates of Population and Housing

10. The Census Bureau has a long-standing program to develop post-censal estimates of the population of the United States, each state, county, and functioning governmental units. The national, state, and county population estimates are developed using a method called the Acohort component method.@ A major assumption underlying this approach is that the components that constitute population change -- births, deaths, international migration, and domestic migration -- can be represented by administrative data series in a statistical model. In order to build the model, Census Bureau demographers begin with most recent decennial population base and estimate each component of population change separately. The vital statistics data on birth and death, coupled with estimates of international and domestic migration developed using various forms of administrative records, provides the basis for the estimates of the components of population change.

11. The Census Bureau calculates subcounty estimates using a Ahousing unit method@ in which the change in the number of housing units at the subcounty level is used to distribute the county population to subcounty areas. More recently, the Census Bureau has extended the estimates program to develop estimates of the number of housing units for states and counties. The number of housing units in each area is derived by updating the housing unit count from the most recent decennial census, using information from building permits and other sources.

12. Throughout the population estimates process, the Census Bureau works with members of the Federal State Cooperative Program for Population Estimates (FSCPE). These members, designated by their respective state governors, work in cooperation with the Census Bureau to produce subnational population estimates. FSCPE members supply some of the information used in the population and housing updating process and review and provide comments on the resulting population and housing estimates.

13. The population estimates for counties are used as control totals for post-stratification in the last states of the ACS weighting (Dahl, 1998) to improve the ACS estimates. Although the flow of information between the ACS and the post-censal estimates program is currently one-way (the estimates are used as controls), the Census Bureau is examining ways to use ACS information to improve the demographic estimates. When ACS data become available nationwide, they will be used to measure changes in the vacancy rate and the average number of persons per household to improve the population estimates from the housing unit method. Changes in the number of addresses on the MAF will probably become an increasingly important variable in the models for the housing unit method. Also, the ACS can detect shifts in race or ethnic distributions to improve the demographic models.

Small-Area Models Combining Survey Data and Administrative Records

14. More recently, statistical models, combining data from national household surveys with data from administrative records systems, have been used to make small-area estimates for specific characteristics. The Census Bureau=s Small Area Income and Poverty Estimates program uses a prediction model, where the dependent variables are estimates of income or poverty from annual supplementary questions to the national labor force survey, the Current Population Survey (CPS), and the dependent variables include information derived from income tax records and from measures of food stamp program participation. The modeling is done separately at the state and county level, after which the county estimates are adjusted to sum to the state estimates. An empirical Bayes estimator combines the estimate from the prediction model with the direct estimate from the household survey, for states and for counties that are in sample for the CPS. For counties that are not in sample for the CPS, the model-based estimate is used. Fully Bayesian methods are being tested for state estimates.

15. The U.S. Bureau of Labor Statistics has a Local Area Unemployment Statistics program that similarly uses Current Population Survey estimates of unemployment as the dependent variable, and information from local unemployment insurance records as the preditor variable. These models make strong use of the time series structure of these variables, and make monthly estimates for cities and towns as small as 25,000 population. For both of these programs, American Community Survey data can be useful, as an additional predictor, once the ACS is in place for all counties.

V. WHY THIS DESIGN FOR THE ACS?

The Alternative of Expanding the National Labor Force Survey

16. An alternative to having a separate mail survey, like the ACS, to produce intercensal small-area data, would be to expand the CPS sample and have one large personal-visit survey to produce both annual small-area data and estimates of short-term change in labor force characteristics. A rolling sample design with approximately 2 million addresses per year (without sampling for nonresponse followup) could in theory serve both objectives.

17. The unit cost of this design would be substantially higher than the current CPS. The CPS now uses a cluster sample, conducts most interviews by telephone, using the telephone numbers obtained on the first of eight interviews at each address, and has a shorter interview than a combined ACS/CPS survey would require. It is estimated that a combined ACS/CPS survey using this design would cost several times as much as the cost of the current CPS plus the projected annual cost of the ACS operations.

18. Ultimately, the barrier to combining the CPS and the ACS is the requirement for a measure of month-to-month change in unemployment from the CPS, released early in the following month, combined with the size of the ACS sample that is required to get estimates for all small communities in the large U.S. population. The CPS requirements cannot be met with a mail survey, but the ACS requirements are too expensive to meet other than with a mail survey. If a rolling quarterly measure of unemployment were acceptable, or if the national population were smaller, a combined design would be more attractive.

The Alternative of a Mid-Decade Census

19. With the failure of a mid-decade census to be funded in the United States in the 1985 or 1995, this alternative was not extensively considered when the ACS was designed in the mid-1990's. For the purpose of updating census profiles for allocating funds for small areas, a mid-decade sample census is arguably about as Atimely@ as the five-year averages proposed for the ACS. However, a mid-decade Asnapshot@ is not effective for monitoring year-to-year changes at the state level, or to monitor trends for smaller areas, to help in local planning. These new uses are important reasons for doing the ACS.

Alternatives Relying Mainly on AIndirect@ Estimation

20. Another alternative would have been to rely less on Adirect@ estimation from a large survey, and more on Aindirect@ model-based methods combining information from administrative records and smaller surveys. The Asmaller surveys@ could be the CPS and existing surveys, or they might include a smaller mail survey such as a much reduced ACS, to provide the dependent variables for methods such as those described in Section IV.

21. Research on such methods is part of the long-term vision of the ACS. As these methods develop and become accepted by data users, the need for some of the ACS sample may gradually be replaced by information from statistical models. However, the development of these methods is not far enough along to eliminate the need for large samples to produce estimates of a variety of characteristics for very small areas. A particular limitation of the model-based methods is that so far they can only produce estimates for a few characteristics at a time, and cannot replace the large number of cross-tabulations produced by either the census long form or the ACS.

Why This Data Collection Design?

22. The uniform spread of the sample was needed to provide comparable estimates for all levels of geography each year. An alternative design with different states being interviewed in different years was previously explored (Herriot, Bateman, and McCarthy, 1989), but was rejected because of the difficulty in making comparisons among states, when each year some states had direct estimates and others did not.

23. The choice of a mail survey with followup was based on experience with similar data collection methods in the census. The particular multiple-mailings approach used in the ACS was selected based on research following the 1990 census. The limitations of the questions that can be asked by mail were not a barrier, since the ACS objectives involve the topics covered by the census long form survey, which is also done by mail.

VI. POTENTIAL IMPACT ON THE 2010 U.S. CENSUS

24. The simplification of the 2010 census process by eliminating the long form, combined with the improved accuracy of the MAF and TIGER from the enhancements to those systems, provide new opportunities for conducting the next census.

25. Increased use of the Internet by respondents will be facilitated by having only a short form. Having only a short form will also facilitate automated data capture, and will reduce the potential cost of using computer-assisted interviewing by the large staff of enumerators.

26. The enhancements in the MAF/TIGER system will enable enumerators to make greater use of Global

Positioning System technology to reduce their work of accurately finding nonresponding addresses, and in other ways will make the enumerator=s job easier and improve the accuracy of small-area data from the census. The enhancements will also let us make greater use of Geographic Information Systems maintained by state, local, and tribal governments.

27. The decennial census operations will be able to build on the intercensal ACS operations, as well as the ongoing interaction with local and tribal governments from the MAF/TIGER program. The experiences of ACS field representatives can give advance notice of where special enumeration procedures, such as interviews in languages other than English, are needed. The ACS field staff, who will be locating addresses and conducting interviews in all parts of the country, can serve as a cadre of leaders during the census. The operations for the decennial census, and the Census Bureau=s intercensal programs, can be integrated more closely than ever before.

REFERENCES:

Dahl, S. A. (1998). Weighting the 1996 and 1997 American Community Surveys. *Proceedings of the American Statistical Association Survey Research Methods Section, 172-177.*

Herriot, R.A., Bateman, D.C., and McCarthy, W.F. (1989). The Decade Census Program - A New Approach for Meeting the Nation=s Needs for Sub-national Data. *Proceedings from the American Statistical Association Social Statistics Section, pp.351-355.*

Kish, L. (1990). Rolling Samples and Censuses. *Survey Methodology,* 16, 1, pp. 63-79.

Melnick, D. (1991). The Census of 2000 A.D. and Beyond. *Review Major Alternatives for the Census in the Year 2000.* U.S. Government Printing Office, Washington, D.C. August 1, 1991, pp. 60-74.

- - - - -