

Continuous Measurement - Harry A. Scarr to Association of Public Data Users (APDU)

U.S. Census Bureau
Washington, DC 20233

The speech was presented to the Association of Public Data Users (APDU), October 16, 1994.

Since the idea of continuous measurement began to get serious consideration as a real possibility for providing an alternative source of information for the so-called "long form" items in decennial censuses, there has ensued an intense and fruitful polylog about its cost, effects, nature, desirability, and appropriateness as a major defining element of the statistical community of the United States Federal government. It has also provided an opportunity for the personal dimension of comfort with change to manifest itself in various fora where the notion is discussed. Perhaps that element needs to be addressed right up front: when things change, they are different. And, in the case of continuous measurement, it is no longer a case of Plus ça change, plus c'est la même chose. We are talking about reengineering, to use current jargon, the way statistics are collected and processed in this country.

For years the Census Bureau has been told to improve the timeliness and quality of its data. The Congress has authorized -- but never funded -- a mid decade census as an expression of its frustration with a lack of timely demographic information between decennial censuses. Leslie Kish offered a rolling census to provide, throughout the decade, contemporaneous data. The late Roger Herriot carried Kish's notions several steps further with a variety of designs for collecting information throughout the decade. And, at the moment, Charles Alexander is working with Larry McGinn and the Census Bureau Continuous Measurement Staff to design a specific program in connection with the 2000 census that would provide those data. Even the Congress, in the person of Representative Tom Sawyer, has urged the statistical system to begin considering tradeoffs to assure more timely information.

Congressman Sawyer: "There is a growing perception that the census has become unwieldy and that the level of content collected on the forms may be diminishing the accuracy of the population counts."

". . . the relationship between content and coverage isn't the only reason to consider the redistribution of the data burden throughout the decade. All of us need data that is more accurate. In a period of rapid change, if the numbers aren't timely, they are not accurate."

Background

The constitution of the United States says, in Article 1, section 2:

Representatives and direct Taxes shall be apportioned among the several states which may be included within this union, according to their respective Numbers The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such manner as they shall by Law direct.

Beginning with this instruction to count people and put them where they belong geographically, the census has grown over its two century plus history to include a variety of other data about those people -- including information about how well they are housed.

In spite of the fact that most people consider the decennial census a convenient way to collect all these data, there are limitations to the quality and usefulness of collecting these other data using this kind of a once every decade structure.

To reiterate, under the current census model, data provided by the "long form" quickly become outdated for the purposes for which they are intended. Further, even though sophisticated means of extrapolation are used to update the data there remain severe accuracy problems for information on most cities and counties, not to mention smaller geographic areas like city blocks or school districts. To try to increase this accuracy and make these data more timely is the purpose of "spreading out" these data over the decade. Continuous measurement corrects these problems of "revisions" and "lack of fit" between the time of the problem and the time of data collection.

What is Continuous Measurement

Continuous measurement is a method for obtaining decennial census "long form" data through a series of large household surveys conducted every month. The general features of continuous measurement include collecting only basic data in the census itself every ten years while shifting detailed data needs that the "long form" has traditionally included from the census on to these large, monthly, household surveys.

Continuous measurement consists of specific components:

- conducting a large monthly household survey;
- using this survey operation to help update the Bureau's address list throughout the decade and linking it to our geographic database;
- using information from the continuous measurement survey to increase the effectiveness of other household surveys such as the Current Population and National Health Interview Surveys; and
- combining data from the continuous measurement survey, other household surveys, the previous short form, and demographic estimates derived from administrative sources, through a program of estimation, to create small-area estimates for the data.

Continuous Measurement for the Year 2000

If we were to use continuous measurement as a way of collecting "long form" data for the 2000 census, this is how it would work: A monthly household survey would begin in January of 1999 and be conducted through December of 2001. A sample of 400,000 different addresses each month would be selected from the most current Master Address File and mailed a questionnaire. From 1999 through 2001, the monthly sample would be 400,000 per month. This size sample would mean that continuous measurement "long form" data would be available as early or earlier than "long form" data collected traditionally -- that is, collected during the census year of 2000 as of one date only. The resulting small area estimates will be based on data collected over a three year period of time. After 2001 the monthly sample size would drop from the 400,000 level collected monthly from 1999 through 2001 to 250,000.

Housing units will be spread across the country, with each month's sample being a separate set of housing units. Each month's sample will produce data that will represent the nation and areas with large populations. Responses will be accumulated over months and years to produce data for the small geographic areas for which data are produced in the decennial census.

1. Questionnaires would be mailed to 400,000 different addresses each month, followed up by reminder cards and replacement questionnaires to those people who do not return their questionnaires in a timely manner. We would plan for a return rate of 60% (240,000 complete mail returns). This return rate is based on our experience with test censuses as well as the other demographic mail surveys which the Bureau conducts.
2. Sample housing units that do not respond by mail, would be interviewed by telephone if possible. We are assuming that we could obtain phone numbers for about 50% of the addresses, and would then use our three Computer Assisted Telephone Interviewing centers to complete 70,000 of the 160,000 mail nonrespondents.
3. A sample of the housing units that could not be interviewed by mail or telephone would be visited for a personal interview. This sample would include 20,000 of the nonrespondent addresses as well as all known vacant housing units.

This describes our idea of a "starter kit" for spreading information collection over the decade.

Beyond 2001

After 2001, the sample size of the large household survey will drop from 400,000 to 250,000 addresses per month. Over five years the cumulative mailout sample size would be 15,000,000 housing units -- or about the same as the sample that would have been given the "long form" in a traditional census. The accumulated total sample up to each of the early years of 2002 through 2005 will be higher than 15,000,000 because the 400,000 per month sample in 1999 through 2001 is higher than the monthly 250,000 monthly sample from then on.

The accumulation of data across these monthly samples would provide timely data annually for states and areas of 250,000 population. Annual rolling accumulated estimates for small geographic areas would be based on accumulations of five years of data. That is, for each year new data would be added and the first year data dropped to calculate the estimate.

For example during 2004 we would produce estimates based on accumulating data from 1999 though 2003. In 2005 we would produce estimates based on accumulating data from 2000 through 2004.

Other Household Surveys

After the 2000 census, samples for the Census Bureau's current household surveys -- the Current Population Survey, the Survey of Income and Program Participation, the National Health Interview Survey, the National Crime Victimization Survey, the American Housing Survey, and the Consumer Expenditure Surveys -- would use the master Address File as their sampling frame. These surveys can use information from continuous measurement to target their interviews to the kinds of households that are of greatest importance to each survey. For example, the Health Interview Survey could oversample households, say, containing elderly individuals.

Eventually, we would hope to be able to link these and other surveys with the continuous measurement survey into a coherent measurement system to meet the Nation's data needs.

Benefits of a Continuous Measurement System

Continuous measurement would offer benefits to the decennial census, to the collection of detailed data, and to other current survey operations.

- The Continuous Measurement Survey will shift detailed data needs from the decennial census to continuous surveys during the decade. This allows the decennial census to do a better job of enumerating the population, by removing the distraction and burden of being the principal source of small-area characteristics. As Gene Erickson noted recently in PAA Affairs, "It seems clear to me that use of continuous sampling will provide clearly better data than a long census form." He bases his conclusion on his detailed analysis of the "long form" data quality in the 1990 census, particularly as it adversely affected "variables like the poverty rate for Blacks and Hispanics."
- The Master Address File is maintained and continually improved through sampling and address correction feedback, which means the address file will be more accurate, improving the basic count for the decennial census and the sample frame for the demographic surveys.
- With a continuous measurement system the currency and relevancy of data improves. The continuous measurement system will produce survey data comparable to the present census "long form" sample data, but the estimates will be available annually for all states and Congressional districts, and cities and counties of 250,000 population, and annual rolling accumulated estimates for smaller areas. Thus data will be more frequent than is currently available from the census.
- There is a "better fit" between the time of incidence and time of data collection.
- The Continuous Measurement Survey allows questionnaire changes or sample customizing to address current needs. For example, we can

- provide the effects on per capita income for areas of large natural disasters like the mid-west flood or Hurricane Andrew.
- The timeliness of the data is improved. The Census Bureau expects to produce data within 6 months of the end of data collection.
- The quality of data improves due to: well trained permanent interviewer staff; computerized interviewing; smaller permanent coding staff with continual training; and monitoring and controlling the completed sample for small areas to ensure sample size.
- Integrating the sampling for current surveys into the Master Area Address File and using continuous measurement, would allow surveys to oversample for rare characteristics such as high income, allow the sample frame to be updated during the decade, produce better quality subnational estimates, and improve survey coverage.
- There would be cost savings from improved samples for current surveys by an enhanced ability to select more efficient and better designed samples.
- Continuous measurement will be useful to our economic statistics as well by providing the characteristics of the individual workers matched to the establishment data for the businesses where they work. This will allow us to study the educational levels or other characteristics of workers at a particular establishment over time and to analyze the personal characteristics using detailed data from the establishment surveys.

Current Activities

The Census Bureau has made significant progress in designing a continuous measurement prototype. Current plans for testing the prototype include a 1995 telephone Questionnaire test to learn about measurement problems created by asking questions throughout the year, and to produce a demonstration data delivery system; a 1996 test using a monthly sample size of 25,000 households of the full system in several test sites; and a pilot national development survey in 1997 and 1998.

The continuous measurement staff is currently doing research to:

- determine the accuracy and completeness of the Master Address File;
- determine the success of obtaining telephone numbers for sample housing units by using data files that allow us to look up a telephone number based on an address;
- plan and develop a rapid data dissemination system;
- collect income data throughout the year using moving reference periods; and
- develop and explain methods to calculate the necessary 3 and 5 year averages; and solve various technical problems such as whether the five-year averages would be population-weighted or are dollar amounts to be adjusted for inflation.

The Census Bureau is also working those who use our decennial census data, and especially with the Congress, to ensure that everyone understands continuous measurement, continues to have input into the decisions about its implementation, and has his or her data needs met.

Conclusion

The continuous measurement paradigm may create as significant a discontinuity in the nature of the statistical system as that which introduced sampling as a procedure in the late 30's and early 40's. Let me remind all of you that real change is seldom incremental. Rather, it results from a shift in paradigm -- a shift in the way we think about things. And it is usually unpredictable -- it sneaks up on us, takes us by surprise, and begins to have a life and direction of its own. I think that is what the continuous measurement notion is doing right now. Those who have adopted the paradigm are articulating a vision of what the statistical collection activity of the Federal Government could become, and what benefits it could offer.

The forces that have converged to make this happen are many and varied: Congressional (more timely information; a less burdensome decennial census); organizational (a less hodge-podgy collection of discrete, unconnected surveys with some information movement between them); technical (examination of the quality of the products from the new system, and the quantification of the tradeoffs); serendipitous (the process of rethinking how the decennial census should be done); and economic (more bang for the buck).

I submit that, indeed, the idea has already assumed a life of its own. I further submit that the genie can no longer be quietly put back into the bottle. I believe that, should this administration decide that continuous measurement should go forward in the context of the 2000 census, it can and will be done successfully.