

This month's guest columnist, Ed Spar, is a regular to Amstat News pages as the author of the Council of Professional Associations on Federal Statistics (COPAFS) updates. As COPAFS executive director, Spar is well-versed in the federal statistical system. I'm grateful he accepted my request to write on the health of the federal statistical agencies with his comments on the ongoing and upcoming challenges for these agencies.

~Steve Pierson, ASA Director of Science Policy, pierson@amstat.org

Challenges Facing Federal Statistics in the United States

Edward J. Spar, Executive Director, Council of Professional Associations on Federal Statistics

Are the federal statistical agencies in the United States meeting the needs of the many users? Surveys required for policy purposes in health, education, labor, and other areas are being conducted with well-tested statistical designs that so far have reasonable margins of error. The decennial census, even with an under and overcount, meets the needs of the Constitution and thousands of federal, state, and local data user measures—including labor force data, gross domestic product, the system of national accounts, health, education, and income estimates—are covered by the federal statistical agencies. Estimates of the population are reasonable, even in situations where high immigration and/or internal migration—which have disproportionate influence—take place. The agencies are sensitive to the need to maintain the confidentiality of respondents. Based on the above, it sounds as if the federal statistical system is healthy and on track. But what about the future?

Many new problems are facing the statistical agencies, and it will take an enormous effort to solve them. Indeed, the agencies are fully aware and understand there is a need for innovative thinking. An example of the type of innovation that has already taken place is the U.S. Census Bureau's American Community Survey. This is a replacement for the decennial census long form, and as an ongoing annual survey of about 3 million housing units, it is unique. The ability to have data available every year for national, state, and local geographies is an important step for a dynamic country such as the United States. Another innovative set of data is the U.S. Census Bureau's Longitudinal Employer-Household Dynamics. Using a mathematical model to ensure nondisclosure, data are available at local geographic levels.

Nonresponse

A critical issue being closely monitored is the declining response rates in key federal surveys that measure, for instance, employment, income, consumer expenditures, health, and education. Surveys that were achieving rates in the middle to high 90% range are now attaining response rates well below that. Clearly, the continuing decline in nonresponse will have serious effects on the usefulness of data collected. Either the statistical error will become so high that estimates will be of limited value or, perhaps even worse, the data may lose its value due to biases.

Clearly, the statistical agencies are aware of the problem and much research is being conducted to determine, for example, if address-listing techniques can be of use in conjunction with telephone interviewing. Some work has been accomplished in nonresponse bias, yet much more is required. The issue of conducting telephone surveys, given the elimination of landlines in favor of cell phones, also must be addressed.

Data Retrieval

The data retrieval world has been transformed by the web. The concept of charging for governmental data is no longer realistic, given the data user's assumption that all data online should be free. Also, search engines such as Google have enabled users to retrieve diverse information as an integrated package. However, data integration across federal statistical agencies is limited. For example, there is no way to analyze and reconcile the many measures of



ASA Science Policy Action

ASA signs letter in support of fiscal year 2010 funding for U.S. Census Bureau, health accounts, and National Science Foundation

income between and within agencies. Each agency creates its own web site and its own data dissemination system with little or no regard for the user, who has to go to more than a dozen sites and learn a dozen approaches to get a complete review of the socioeconomic data of the United States. Indeed, if the user wants to integrate the data, it is much easier—but more expensive—to go to a private sector vendor.

At a time when the web is there for the specific purpose of retrieving information easily, freely, and comprehensively, this approach is outdated. The time has come for an integration of data processing and retrieval systems. This should be accomplished, even though the structure of the U.S. federal statistical system is highly decentralized.

The concept of a single system in the United States, and probably most countries, is misleading. In reality, what you have is a confederation of agencies reporting to different jurisdictions that are independent of each other. In the United States, there is limited administrative record data sharing and, with separate Internet sites, little integration of tabulated data sets. Each agency has its own budget, and except for the purchasing of surveys from the U.S. Census Bureau, little in the way of financial interaction. This lack of centralization affects the agencies' influence with Congress and the funding for their programs. (This is not the case during the decennial census cycle, when the apportionment of congressional seats can affect a representative. Other data series, such as employment and inflation, also are closely reviewed.) Would a centralized, single agency help solve this? Put another way, would an agency large enough to be recognized by Congress and the current administration as being critical to the overall health of the nation have a better opportunity to receive the needed resources?

Confidentiality

To perhaps overstate the case, the days of taking censuses and surveys may soon be coming to an end. We may be at the crossroads of relying on administrative records. Using administrative records data brings up confidentiality issues on the part of agencies. Yet, these data may become the basis for measuring health, education, employment, expenditure, transportation, energy use, and many other areas. Using administrative records data will call for public/private sector coordinated analyses and the allocation of talent and research dollars. If the use of administrative data becomes the norm, it is not too *outré* to see a time when no data will be real—model-based estimates will be used to protect identities.

Data at the Local Level

Assuming the statistical accuracy of the American Community Survey is reasonable for small geographical areas' income, education, and employment data, these data will significantly enhance one's ability to measure the effects of policy and social change. Yet, there are many data series in which more local information is needed. For example, users of the National Criminal Victimization Survey conducted by the Bureau of Justice Statistics have overwhelmingly stated the need for state and county information. Similar requests for local education and health data also have been voiced. Most surveys from statistical agencies such as the National Center for Health Statistics and the National Center for Education Statistics produce data at the national level. For national policy needs, this makes sense. To serve the needs of states and local communities, however, national data does little good. The statistical agencies surely understand local needs, and this lack of local information is not based on the agencies' lack of desire to produce it. Rather, we are back to the issue of influence and the need for resources. Where in Congress are the champions for statistics when we need them? I submit that, for the most part, they're gone.

Staffing

Over the next few years, many of the senior staff members of statistical agencies will reach retirement age. At the same time, it is difficult for agencies to hire new personnel and hold on to talented statisticians and economists. The private sector offers both higher salaries and the opportunity to diversify. Indeed, the problem of "stove-piping" within statistical agencies, where talented people are expected to stay in one place for an overly extended period of time, is counter-productive. There is a need to develop a system whereby people can move not only within an agency, but also across agencies. Such a system of diverse training will be required so personnel can develop the skills needed to address the concerns that have been mentioned here.

The challenges I have reviewed are only the beginning. To properly measure the effects of the current, and probably future, economic crises in the United States, timely and relevant data are needed for those who have to make informed decisions affecting all Americans. ■

Make the most of your ASA membership

Visit the ASA Members Only site:

www.amstat.org/membersonly

Partner discounts • Member forums • Enews archive