

The Impact of Statistics on Government Policy: Discussion

Nancy M. Gordon

U.S. Census Bureau, Retired, 5500 Friendship Blvd., Chevy Chase, MD 20815

Abstract

Although four papers were presented and discussed at this 2017 JSM Invited Paper Session, these remarks cover only the two papers that were submitted for publication in the *Proceedings*.

Key Words: Federal statistical system, confidentiality protection, population census, official statistics, statistical quality, uses of federal statistics, Office of Management and Budget

1. The Session As A Whole

To put my remarks in context, my career has been multi-faceted. I began in mathematical economics and statistics, where I learned from Nobel Laureate Kenneth Arrow about public goods, which is what the federal statistical system produces.

I moved on to examining labor market issues using both theory and empirical studies, and then turned to estimating the impacts of policy alternatives on individuals and families using federal data, and last (but not least) to heading the demographic programs at the Census Bureau – where I led the its efforts to start new programs, especially the American Community Survey or ACS. Please note – without many others’ support, including from the private sector, and *especially* without Katherine Wallman’s tireless efforts, the ACS would not exist today.

Compliments are due to Jim Cochran, who organized this session as well as serving as its excellent chair, and also to the Program Committee, who decided that the impact of statistics on government policy warranted being an Invited Session.

These papers make clear the great importance of the federal statistical system to all of us as citizens or residents of the country. They demonstrate the many uses of federal statistics and administrative records to inform important decisions, not just by local, state, and federal governments, but also by private sector firms and organizations, and by individuals.

However, I fear that a great many, and possibly most, members of the ASA are unaware of the importance of this system to their profession and their association. Moreover, these members likely have no idea that the ASA was founded *because of concern about the state of federal statistics* (that is, the population census) in 1840, as Lynne Billard noted.

Why is the federal statistical system so important to all of us as statisticians?

First, the federal statistical system sets the “gold standards” that all empirical statisticians strive to meet, regardless of their specific fields like biostatistics or data science. Specifically, the Office of Management and Budget, or OMB, issues these standards and directives based on work by the Office of Statistical and Science Policy. Among them are ones on:

- Collecting and processing data, and designing information products
- Disseminating objective results in a non-political and timely way
- Transparency about the methods that were used and limitations that inform fitness for use and
- The principles that guide federal statisticians, such as professional independence and producing accurate, objective, timely, and accessible statistical information.

Second, staff at statistical agencies develop or enhance methods and tools that are adopted by others. These methods include dramatically improved ways to protect respondents' confidentiality, and other statisticians can apply their underlying theoretical work in related fields. Also, under OMB's guidance, the Federal Committee on Statistical Methodology, or FCSM, coordinates interagency research, publishes resulting papers, and organizes conferences on statistical methodology at which their progress is shared with statisticians inside and outside the government.

As a result, when the federal statistical system is under attack, it is an attack on all of us in the ASA. I'm particularly concerned that some ASA members object to the association taking public positions such as defending the budget requests of statistical agencies, even though both the country and the profession need a comprehensive, up-to-date federal statistics program.

Most statistical agencies have had their budgets cut in real terms for several years, and often in nominal dollars, as the ASA's Steve Pierson has documented in *AmStat News*. Without adequate funding, statistical agencies have no choice but to respond by cutting quality – dropping more key products, taking longer to release remaining ones leading to less timely information, reducing accuracy, not implementing better methodologies already developed, and curtailing research on new collection techniques and changes in content to increase relevancy. Even when funding is restored, as was the case for the Supplement on Contingent Workers, we had gaps in data and losses in efficiency.

2. The Papers

Turning to the papers themselves, a general compliment is in order. Both the current President of the ASA, Barry Nussbaum, and the President-elect, Linda La Vange, have emphasized that high quality work is not enough. We must *clearly* explain what we did and what we found. In Barry's words, "It is not what we said; it's not what they heard; it's what they say they heard."

The papers present complex information in an easily understandable manner. As a result, I expect that what members of this audience say they heard will be an accurate portrayal of the material from this session, which is what effective dissemination is all about.

2.1 Lynne Billard – Statistics and the Census

Lynne is a former President of the ASA and was a member of the Census Bureau's 2000 Decennial Advisory Committee, which may have sparked her interest in this topic.

Summarizing all the changes in the population census since 1790 is a daunting task, and Lynne rose to the challenge. Her presentation could only skim the surface, and I hope that you will read the entire paper for two reasons.

First, it serves as an excellent prologue as the federal statistical system wrestles with how to confront a new reality, including competition from private suppliers of information that is available fast and seems just as good as official statistics, as well as growing distrust of government.

Second, as you start or continue your careers, in many cases aiming for positions of influence, you might gain from applying the underlying behavioral traits of key players from the past that resulted in their exceptional effectiveness. Of special interest is Francis Walker, who was the Census Bureau's Director for the 1870 and 1880 censuses. That was more than a century ago, but in my view, his principles remain every bit as applicable now as then.

Some of the most dramatic changes in census history were:

- Adopting scientific methods for 1880, instituted by Walker and continued under subsequent directors
- Expanding the use of technology, starting with Hollerith machines for 1890 and onward
- Asking the bulk of the questions *only* of a sample of households in 1940 (that is, instituting the long form) to reduce burden without a commensurate loss in information
- Substituting the American Community Survey for the long form in 2010, drastically simplifying the operations of the census.

Because of these changes, the census form now has only a few questions, and the Census Bureau gained considerable experience using technology, though not just for the population census. For example, technology transfer from the Economic Census has sped the development of Internet data collection from individuals and households.

As a result, expanded use of administrative records, end-to-end processing, and Internet collection are possible for the 2020 population census, and could generate significant savings compared with using earlier methods. However, lack of funding to date is limiting the Census Bureau's ability to pretest the new methods, which threatens their successful implementation.

Looking to future decades, I hope that we will see another leap forward that resembles that of the 1880s.

2.2 Katherine K. Wallman – Statistics are Evidence; Evidence is Statistics

Katherine has done a masterful job of illustrating several aspects of the federal statistical system, which she led as Chief Statistician of the United States for many years.

These aspects include:

- The breadth and depth of the statistics the federal system produces
- The principles that underlie their production and release and
- Their uses for decision making, primarily by governments, but also by businesses and throughout society, down to individuals and families.

However, now that the page limit for papers in the *Proceedings* has been eliminated, I hope she will find the time to expand some aspects of hers. One area is her discussion of ways *businesses* use data – from demographic sources like the American Community Survey and from economic sources like the National Economic Accounts and the Economic Census.

A second expansion would be including a few examples from the Congressional Budget Office. Doing so would highlight times when analyses based on federal statistics have had a major impact on Congressional actions.

The most exciting part of her paper for me is the last section that begins, and I quote:

“If, instead, data are thought of as a product *from* and *for* the general public to use to make personal, business, and public policy decisions, what would that mean for the federal statistical system in practice?”

Her points are so important, I want to be sure that we follow the age-old edict to “tell them what you're going to tell them, tell them, and then tell them what you told them,” though in a somewhat different fashion.

The federal statistical system is wrestling with how to respond to the public’s demands for

- Ever more detailed information – about particular types of people and firms, and for small geographic areas
- Information that is current, not years old and
- Information presented so that it can be accessed as easily as an app on an iPhone.

She suggests one direction for statistical agencies, which is to streamline their operations and incorporate more administrative records with their survey data, in order to produce more useful products faster.

Another direction, which is not mutually exclusive from the first, is to implement recent advances in confidentiality protection. The agencies must follow this principle – not only is protecting confidentiality the ethical thing to do, but it is also required by law. And from a pragmatic point of view, if it is not followed, agencies would likely see lower, perhaps much lower, response rates, leading to sharply reduced accuracy. However, until recently, protecting confidentiality has conflicted with releasing data products for small groups or small geographies.

One new approach – adding noise to the original data – has been successfully implemented in several programs. Even though this method inherently reduces the accuracy of each adjusted value, it enables information to be released for strikingly more geographic areas, industries, and demographic groups. Furthermore, when the noise-infused values are aggregated for small groups of businesses or individuals, or for small geographic areas, accuracy increases rapidly. (See papers from other JSM sessions for more information.)

Following these directions will be intellectually challenging, and will require that resources be devoted to investing in the future. Moreover, the quality of some resulting products, especially those relying on administrative records, might be so low that agencies would be reluctant to release them.

But when would quality be “so low”? Although standard measures of various aspects of statistical quality exist, one size does NOT fit all. The minimum acceptable quality depends on the intended use of the data, which determines which aspects of quality are most important. For example, an historian might highly value accuracy but not timeliness, two aspects that frequently must be traded off against each other. A policy maker needs information before a decision is made and accepts lower accuracy to get the preliminary results quickly. Hence, a better question is “what is good enough for a specific use?”

Given that a statistical agency cannot visualize all possible uses, nor *should* it try to answer the question of “good enough” for the uses it does recognize, the user will be the one to decide.

Katherine is hopeful that providing sufficient information about the characteristics of the data (called metadata) will allow users to determine “whether the data are good enough,” that they will frequently consider the metadata, and that they will not use the data unless they are “good enough.”

Although usually an optimist, I fear that most people will use whatever is available, regardless of the quality, and probably not even bother to check on quality.

Nonetheless, I think the statistical agencies need to move down both these paths. Recall an earlier remark – the federal statistical system sets the gold standards that other statisticians *strive* to meet. However, other statisticians can't always meet them, and it is time for the federal statistical system to accept that not all its products can meet them either. It is a matter of moving forward in order to meet the needs of the public, or become irrelevant.

I hate to close on a negative note, but feel obliged to acknowledge the “elephant in the room” – namely, the prospect that funding for the statistical agencies will be cut more than ever imagined. If that happens, agencies would not have the funds for investing in either of the two directions just described. Instead, they might have to “start from scratch,” because expensive existing processes could not be maintained. However, it might be too late for that, because staffs would shrink and the necessary skills might not be available.

Let’s hope that the innovative character of people in the federal statistical system, and the final funding levels, will be enough to enable the agencies to move forward quickly and successfully:

- To streamline their operations
- Incorporate more administrative records with their survey data and
- Implement recent advances in confidentiality protection.