

## CHALLENGES TO EVIDENCE-BASED POLICY MAKING IN THE DECENTRALIZED U.S. STATISTICAL SYSTEM

Nancy A. Potok, Amy O'Hara, and Ron Jarmin  
U.S. Census Bureau  
Washington, D.C., USA  
[Nancy.a.potok@census.gov](mailto:Nancy.a.potok@census.gov)

*One goal of the current U.S. presidential administration is to accomplish more evidence-based decision making to drive National policy. The FY 2014 President's Budget contains funding for rigorous evaluations across government as well as evaluation capacity building. The stated purpose is to build knowledge so that spending decisions are based on strong evidence that investments yield the highest social returns. One promising approach to conducting data-driven evaluations cited in the budget documents is using administrative records to conduct low-cost evaluations. The US Census Bureau has extensive experience linking statistical survey data to administrative data to create powerful data sets. However, in a decentralized statistical system, many challenges exist to sharing records and data among federal statistical agencies. This paper discusses both advances and challenges in increasing evidence-based policy making.*

### INTRODUCTION

The U.S. federal government relies on evidence-based policy development, execution and evaluation to maximize policy effectiveness and return to taxpayers. A measurement infrastructure must identify policy gaps and quantify and evaluate program impacts. The U.S. Federal Statistical System is a large and critical part of this infrastructure.

Comprised of 13 principal statistical agencies and over 100 statistical programs within Executive Branch agencies, the Federal Statistical System is decentralized with most activities focused on small specialized domains (see Office of Management and Budget (OMB) (2014) for a full description). The legislative authorities of each entity dictate its statistical domain and scope for data sharing. This has important implications for the utility of federal statistical information for evidence-based policy making which we discuss further below. These implications are particularly important for the U.S. Census Bureau, which is the largest statistical agency in the Federal Statistical System and the leading source of quality data about the United States people and economy.

This paper proceeds as follows. First, we briefly discuss how information at statistical agencies can provide evidence for better management and evaluation of programs. Second, we describe the infrastructure that enables this type of analysis. Next, we outline challenges to fully utilizing the U.S. Federal Statistical System infrastructure for program evaluation. Finally, we discuss the challenges of educating policy makers on the opportunities and implications of using and sharing administrative records and other data in moving to evidence-based decision making.

### USING STATISTICAL INFORMATION FOR EVIDENCED-BASED POLICY MAKING AND EVALUATION

There are three primary ways that information produced by the Federal Statistical System and international statistical offices can aid evidenced-based policy making. First—and perhaps the most important—official statistical data provide reliable social and economic indicators that describe the well-being of society across a broad array of important dimensions (e.g., GDP, poverty, infant mortality, educational attainment, etc.). Second, some statistical activities, such as the U.S. Census Bureau's Survey of Income and Program Participation (SIPP), measure how particular government programs impact outcomes for individuals assisted by those programs. Third, statistical agencies possess rich data assets that can be linked to administrative data from program agencies to measure program effectiveness.

Broad social and economic indicators are useful for identifying the potential need for government programs, but are less useful determining whether such programs are effective. Directly measuring program participation and outcomes in the same survey (as is the case with SIPP) is costly and not practical for most policy interventions. To meet the growing demand for

evidence-based policy analysis, agencies are increasingly looking to link administrative program data with other statistical information to monitor and improve program outcomes. This is evident in several ongoing initiatives at the federal level (see OMB 2013).

These efforts include a recent OMB workshop (“Workshop on Using Individual-Level Administrative Data to Build Evidence about Policies and Programs, while Protecting Privacy,” January 17, 2014) that focused on linking program participation data to data at statistical agencies to build evidence. Another example is the ongoing interagency Business Technical Assistance Evaluation Working Group. These efforts expand upon previous work demonstrating that reliable estimates of program impact can be generated from quasi-experimental analyses linking administrative program data to the rich data infrastructure available at statistical agencies such as the U.S. Census Bureau (e.g., Jarmin, 1999; O’Hara et. al., 2010; Brown & Earle, 2012; Ordowich et. al., 2012; and Rutledge, 2013).

A key advantage of linking data on agents (e.g., individuals, households, business establishments or firms) participating in programs to the micro data maintained by statistical agencies is that analysts can construct control groups. Measuring outcomes consistently across agents receiving program treatments and control groups can yield reliable results through quasi-experimental methods. This evaluation methodology is cost effective relative to other experimental designs or fielding large scale surveys. These studies also yield benefits to statistical agency programs as they produce valuable insights on the quality of statistical data.

## INFRASTRUCTURE

The Federal Statistical System has invested in infrastructure to support research design and evidence building to conduct program evaluation. Statistical agencies identify stakeholders and work closely with program agencies to select methodologies and data sources to meet evaluation needs. Data are available from federal and state programs through legal agreements that dictate security standards and permitted uses. Records are matched using data integration methods across program files, surveys, and censuses. The resulting linked data sets permit analyses including descriptive statistics and modeling to build evidence.

The Census Bureau has invested in computing resources to conduct efficient record linkages, handling billions of records each year. Data are acquired from federal agencies including the Social Security Administration, Internal Revenue Service, United States Postal Service, and Centers for Medicare and Medicaid Services. State program data are acquired through agreements with agencies including State Workforce Agencies and Departments of Human Services to obtain program data from unemployment insurance wage and food stamps data systems. The Census Bureau also contracts with commercial data vendors to obtain information to enhance census and survey operations and develop new data products.

Data from federal, state, and private sources are processed uniformly to append unique person or business identifiers, as well as unique address match keys. Based on project needs, these identifiers allow data to be linked across years or data sources to enhance census and survey products. Such evaluations require that appropriate linkage variables exist in all files. Statistical agencies typically maintain high quality linkage keys, as these are necessary in many statistical programs. Program agencies, however, do not always maintain identifying information for program participants. Data with identifiers may remain at the state or local level where programs are administered. Privacy and confidentiality concerns are often cited for reluctance to retain or share sensitive information. Statistical agencies have a long history of maintaining the confidentiality of the census, survey and administrative data under their control. They can provide program agencies guidance on best practices in privacy and confidentiality and can ensure that any program data introduced into the data infrastructures of statistical agencies will be secure.

## CHALLENGES TO USING DATA INFRASTRUCTURE

There are challenges involved with building and using this data infrastructure. To acquire the data, agreements are needed that articulate legal authority, permitted uses, and data security provisions. Acquisitions are complicated and may take years to negotiate all terms. The Census Bureau is uniquely positioned to acquire data from government and private sources. The Census Act (Title 13, U.S.C.) states that, for the efficient and economical conduct of censuses and surveys,

information from government or private sources should be used to the maximum extent possible instead of conducting direct inquiries of businesses and households. In addition, the Privacy Act provides authorization for other agencies to share data with the Census Bureau. It contains 12 exemptions from the requirement to gain informed consent from an individual prior to “disclosing” his/her information outside the collecting agency. One exemption is for the transfer of records to the Census Bureau for purposes of planning or carrying out a census, survey, or related activity.

Given the language in the Census and Privacy Acts, the Census Bureau has the authority to request data from government and private entities. However, the government and private entities may have conflicting (or unclear) authority or policies that prevent data sharing. For example, the Family Educational Rights and Privacy Act of 1974 (FERPA) prohibits disclosure of personally identifiable information from education records to the Census Bureau for statistical purposes, unless that information has been appropriately designated as directory information (20 U.S.C. § 1232g). In addition, the Department of Education has been restricted in sharing information from the electronic versions of the Free Application for Federal Student Aid with the Census Bureau by the privacy restrictions contained in the Higher Education Act of 1965, as amended (20 U.S.C. § 1090(a)(3(E))). In other cases, the data may be shared but with restrictions on their use that would preclude their effective use for building evidence for policy making.

Negotiating data agreements is challenging with a federal agency; the challenge is amplified when negotiating state-by-state to acquire program data. For example, it took the Census Bureau over a decade to obtain agreements with all fifty states for its Longitudinal Employer Household Dynamics (LEHD) program, which combines Unemployment Insurance earnings and the Quarterly Census of Employment and Wages data provided by states with other Census Bureau data. Acquisitions involve Subject Matter Experts, Lawyers, Privacy Officers, and Information Technology Staff, all of whom are critical for successful negotiations and data delivery. Agreements between statistical agencies and administrative/program agencies must clearly state that data will be used for statistical purposes, not for enforcement or surveillance. This must be clear to avoid perception issues: when analyzing programs or conducting evaluations, only statistical information will be produced that will not affect the eligibility or delivery of benefits for any individual.

There are technical challenges when conducting evaluations. Data from government programs often lack data standards and are shared without metadata. The Census Bureau leads other agencies in the Federal Statistical System in developing methods to document data, develop classification systems, and address missing data when using administrative records data. State agencies may lack information technology staff who can extract data from their system. Many program agencies use software to administer their program and produce compliance reports; they may not be able to manipulate data in the system. State offices may also be overwhelmed by maintaining records for multiple programs. For example, many state Departments of Human Services manage complex and changing food security and health programs. The Census Bureau works with state agencies, enabling secure and legal data sharing while educating states about data protections and risk management.

A final set of challenges involving the use of administrative records for program evaluation involve their reliability and availability. Administrative records data are sensitive to external factors such as changes in law and economic conditions. For example, the Economic Stimulus Rebate administered through the Individual Income Tax system induced millions of households to file tax returns. Similarly, Medicaid data systems in 38 states will likely show the expansion of the program due to the Affordable Care Act. In addition to understanding how program changes affect the data systems, these examples also highlight the need to have timely data delivery and processing to measure and assess programs as they change. Depending on the research design and evidence needed for the evaluation, the variability and timeliness of administrative records may not be an issue.

## CHALLENGES TO EDUCATING POLICY MAKERS

The complex landscape regarding the legal, regulatory, and quality issues surrounding data sharing creates many barriers to building linked data sets to inform policy in a timely way.

First, policy makers are often well informed about areas in which they are subject matter experts, but relatively uninformed about all the statistical data that are available to help evaluate the potential effects of new policies and the performance of existing programs. Current metrics may rely on one or two data sets that provide an incomplete picture of program and policy outcomes.

Next, there may be little connection between policy makers and the people who collect and maintain administrative records and other data within highly structured, bureaucratic organizations. Thus, even if a policy official is interested in learning more about program effectiveness, that person may never talk to the people in the agency who keep the administrative records and may be unaware of the more powerful data that could be gained from data sharing.

Third, agencies are often proprietary about their data as an organizational culture artifact. That is, the agency does not see sufficient value in sharing data with other agencies to overcome the institutional reluctance to share something that is "owned". This tendency can be exacerbated by Privacy Officers who are more focused on protecting data than on creating new linked data sets.

Finally, data sets that originate in 50 different states may be governed by different state laws and are maintained in different software and hardware, making sharing more difficult. Each state will have varying levels of quality and accuracy in its data, as well as a different attitude toward sharing data with the federal government. Often, many people in different parts of state government are involved with different, but related data sets.

Getting to a point where multiple players in multiple locations and agencies take a similar view toward creating linked data sets to enable evidence-based policy making requires a massive educational effort. Privacy Officers often come from the legal or technology professions and have limited knowledge of statistics. Exacerbating the problem is the lack of consistency at the university level on how to educate current students on these linked data sets. While some universities give degrees in program evaluation (often in the public policy schools), there may not be much emphasis on using data that are linked. Often, research is focused on a single data set, e.g. from a particular survey. Researchers may become experts in data from one or two surveys, but do not realize the possibilities for expanding useful information if data sets are linked. In addition, many statistics programs are not developing new methods for measuring quality and reliability on data sets that consist of both statistical data from surveys and administrative records and possibly, commercial sources (third party data).

## CONCLUSION

Executive branch departments and agencies are building capacity and producing evidence to inform spending decisions. The Federal Statistical System, with guidance and leadership from OMB, is addressing the challenges of using administrative data for program evaluation. The Census Bureau is embracing opportunities to use administrative data to conduct evaluations that drive policy development and analyses. The Census Bureau offers statisticians an infrastructure to produce robust social and economic measurements to inform policy.

However, if real advances are to be made in utilizing data to drive public decision making, a broader, more comprehensive educational effort is needed. Statisticians, data stewards, evaluators, researchers, analysts and policy makers need a common framework for learning about the issues and benefits of expanding the use of linked data sets. An educational framework is needed, both at the university level and for current professionals that can create a common understanding across disciplines, agencies and states of how these data are created, protected, and used to drive public investments. In that way, more resources can be devoted to developing new data sets and carrying out the data analyses rather than trying to educate the players in the system anew each time a new need for data arises. This will take a concerted effort on the part of current practitioners to work with colleagues in academia to begin to develop the educational framework and start to develop common knowledge baselines.

## REFERENCES

- Jarmin, R.S. (1999). Evaluating the impact of manufacturing extension on productivity growth. *Journal of Policy Analysis and Management*, 18(1), 99-119.
- O'Hara, B., Tangka, F., & Bauder, M. (2010). New health insurance estimates improve measuring program participation in the National Breast and Cervical Cancer Early Detection Program. *SAHIE Working Paper*. Washington, D. C.: U.S. Census Bureau. Available at: <http://www.census.gov/did/www/sahie/publications/files/New-HI-Est-Improve-Measuring-Program-Participation-NBCCEDP.pdf>
- Brown, J.D., & Earle, J. S. (2012). Do SBA loans create jobs? Estimates from universal panel data and longitudinal matching methods. *Working Papers 12-27*. Washington, D. C.: Center for Economic Studies, U.S. Census Bureau.
- Ordowich, C., Cheney, D., Youtie, J., Fernández-Ribas, A., & Shapira, P. (2012). Evaluating the impact of MEP services on establishment performance: A preliminary empirical investigation. *Working Papers 12-15*. Washington, D. C.: Center for Economic Studies, U.S. Census Bureau.
- Rutledge, M. S. (2013). The impact of unemployment insurance extensions on disability insurance application and allowance rates. *Working Papers 13-10*. Washington, D. C.: Center for Economic Studies, U.S. Census Bureau.