



UNC
COLLEGE OF
ARTS & SCIENCES

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

DEPARTMENT OF PUBLIC POLICY

CAMPUS BOX 3435
209 ABERNETHY HALL
CHAPEL HILL, NC 27599-3435

T 919.962.0674
F 919.962.5824
maryann.feldman@unc.edu

MARYANN P. FELDMAN

S.K. Heninger Distinguished Professor

TESTIMONY

Before the
Commission on Evidence-Based Policy Making—Public Hearing
San Francisco, CA
February 9, 2017

Hello, my name is Maryann Feldman. I am a professor in the department of Public Policy at the University of North Carolina, Chapel Hill.

I am speaking as a member of the community of scholars that seeks to advance the scientific basis of science and innovation policy. My testimony outlines some of the needs of that community related to 1) open science, (2) ready access to administrative data, (3) rationalization of data access and use policies, (4) the important for researchers to understand context or for them to collaborate with practitioners, which also enhances trust, and finally (5) offering suggestions for a governance structure for data access and use while maintaining privacy.

My own research uses a variety of administrative data from public, private and non-profit organizations at the local, state and national level. I am a recipient of funding under the National Science Foundation (NSF) Science of Science and Innovation Policy (SciSIP) program, which is part of the Division of Social, Behavioral and Economic Sciences. The program funds research to develop models, analytical tools, data and metrics that can be applied in the

science policy decision making process and affect the use and allocation of scarce scientific resources.

I am currently on detail to the NSF. The views expressed here do not necessarily represent the views of NSF or the U.S. Government. In my comments, I represent my own views as a researcher and member of the Science of Science Policy (SoSP) community of interdisciplinary researchers and policymakers, who seek to provide a scientifically rigorous basis for informing science policy.

The fundamental question that motivates the Science of Science and Innovation Policy community is, how can the United States use scarce scientific resources more efficiently and effectively to promote scientific and technological progress, economic development and economic growth. This research is empirical and relies on access to data.

It is only by studying the scientific research enterprise, the range of actors engaged in innovation and specific public programs and incentives that scholars can provide evidence to policy makers.

The United States conducts \$456 billion of research and development (R&D) activity annually (NSF 2015). The Federal government, the second largest R&D funder behind business, invests \$122 billion each year through the work of over 15 federal mission agencies (e.g., Department of Defense and Department of Health and Human Services) and 70 sub-agencies (e.g., the Defense Advanced Research Projects Agency and National Institutes of Health), as well as 15 independent agencies such as the National Aeronautics and Space Administration (NASA) and the National Science Foundation (NSF).

Understanding this diverse and dynamic system is a daunting task. Yet our ability to study and understand this field is greater now than it has ever been because of our ability to work with new digital data sources, and benefit from advances in research design and empirical methods. Progress for my community of scholars is dependent on access to administrative data. I applaud the mission of the Commission and want to be sure that the federal science agencies are included in your efforts to make administrative data available.¹

I would like to offer three examples of projects that used administrative data to provide policy-relevant analysis. Federal agencies routinely publish information on funded projects yet access to all applications is rare. Access to administrative records enables analysis of differences between funded and non-funded projects and allows the construction of quasi-experimental research designs to better discern causal effects.

Each of these three projects I mention was conducted through special negotiated agreements that involved contracts or exceptions to gain access to proposal submissions. All have been published in top journals.

First, Ginther et al (2011) conducted the first systematic investigation of racial and ethnic differences in National Institutes of Health (NIH) funding, using 88,000 Research Project Grants (R01), which is the most common mechanism NIH uses to fund science. The NIH IMPAC II grant database provides applicant's self-identified race or ethnicity. Ginther et al (2011)

¹ There is utility to differentiating the difference between administrative data and scientific data. Administrative data capture agency processes, policies, and outcomes in pursuit of their mandated mission. In contrast, scientific data are the product of the agency's efforts.

found, after controlling for the applicant's educational background, country of origin, training, previous research awards, publication record, and employer characteristics, that black applicants were 10 percentage points less likely than whites to be awarded NIH research funding. Published in *Science*, this work was co-authored with Raynard Kington, former Deputy Director of the National Institutes of Health, and Walter Schaffer in the Office of the Director at NIH.

With this evidence, NIH Director Francis Collins, Ph.D., M.D., established the Working Group on Diversity in the Biomedical Research Workforce to develop effective strategies to recruit and promote the professional growth of underrepresented groups in biomedical research, from graduate study to granting of tenure. Based on the findings of the Working Group, NIH made a \$500 million, ten-year commitment to improve diversity through

- Increasing the number of underrepresented minorities who study biomedicine (Building Infrastructure Leading to Diversity);
- Creating a national mentoring network (National Research Mentoring Network); and,
- Hiring a Chief Officer for Scientific Workforce Diversity.

Second, Lui and Agha (2015) examined the success of peer-review panels in predicting the future quality of proposed research using administrative data to track publication, citation, and patenting outcomes associated with more than 130,000 (R01) grants funded by the NIH from 1980 to 2008. Their analysis found that better peer-review scores are consistently associated with better research outcomes and that this relationship was robust even after accounting for an investigator's publication history, grant history, institutional affiliations, career stage, and degree types. A one-standard deviation worse peer-review score among awarded grants is associated with 15% fewer

citations, 7% fewer publications, 19% fewer high-impact publications, and 14% fewer follow-on patents. The results suggest that peer-review is doing a good job of picking projects. Further research can help program managers further refine the process.

Third, Sabrina Howell, in her dissertation and a forthcoming article in the *American Economic Review*, conducted the first large-sample, quasi-experimental evaluation of R&D subsidies using administrative data on ranked applicants to the U.S. Department of Energy's Small Business Innovation Research (SBIR) grant program. The results found that an early stage award approximately doubles the probability that a firm receives subsequent venture capital and has large, positive impacts on patenting and revenue, with stronger effects for more financially constrained firms. The results suggest that the SBIR program is accomplishing its objective: the results are not due to certification, where the award contains information about firm quality, but instead the SBIR program funds are used for technology prototyping, helping commercialize new ideas.

These readily accessible academic studies are in contrast to the often-used contract mechanism, which provides access to administrative data while limiting publication rights. Projects with limited publication rights are not useful for academic careers, thereby deterring academics from investing in such research. As a result, much of the analysis of these data is done by consultants, of varying quality and motives, conducting rather routine analyses, usually not employing the most advanced methods nor publishing the results in the peer-reviewed scientific journals.

These few examples demonstrate the need for ready access to administrative records to inform science policy and address the practices of federal

agencies. In these few studies researchers have negotiated access on an individual basis, overcoming each agency's own unique procedures for accessing non-funded proposals, reviewer and panel rankings, and administrative records, which include additional principal investigator characteristics, project monitoring data and outcomes information.

Imagine what might be possible if procedures for access to administrative data were streamlined making their use relatively painless. This would require amending the *Federal Information Security Management Act of 2002*, and subsequent updates, which provide the framework of information security standards for the federal government's information systems. One straightforward work around is to ask program applicants to provide informed consent to use their data for research purposes.

There are currently no standard procedures for researchers to gain access to federal science agency administrative data. The situation can be even more onerous for obtaining data from state and other publicly funded programs. Existing procedures for gaining access, where they exist, are time-consuming to navigate, difficult to understand and present barriers to the use of administrative data. Certainly, removing legal barriers that limit access for the research community is a needed first step.

In my work with federal agencies, one of the things that surprised me was the great degree of heterogeneity. Each agency has its own policies and procedures for gaining access to administrative data. For example, NIH established internal policies to determine when data from their records management systems can be disclosed for research purposes.² NIH provides different levels of access depending on a researcher's role. Federal

² <https://oma.od.nih.gov/forms/Privacy%20Documents/PAfiles/0036.htm>

researchers can access data if they take yearly training in information security and privacy. NIH hosts a secure, encrypted environment that can be remotely accessed through a virtual private network. NIH staff who wish to access research grants data that are considered confidential and sensitive, such as unfunded grant proposals and reviewer rankings, require additional security background screenings. Personally identifiable information is highly restricted, often limited to a single person record at a time. If the need for expanded access to confidential and sensitive information arises, NIH staff can obtain a special data access agreement, which is limited to one-year duration. NIH staff can also sponsor a 'guest researcher' to obtain a data access agreement to access disaggregated awards data.³ Sponsorship is at the discretion of NIH Institutes, Centers, and Offices. In most cases, these researchers have NIH grant support. External researchers are required to undergo the same clearances required for NIH staff. Researchers must request NIH clearance for any manuscript or presentations showcasing their work using NIH administrative data.

In contrast, many other federal agencies do not have specific policies or procedures in place to grant research access to administrative data. EPA and USDA participate in Census' FSRDC program, providing pollution abatement and food security population survey data, respectively.⁴

The Commission could strengthen efforts by developing overarching directives that coordinate and encourage data sharing with researchers across Federal agencies. One avenue for coordination and harmonization of agency policies and procedures is the Interagency Council for Statistical Policy. In addition, in October 2016, OMB released a notice of solicitation for

³ NIH policy for the guest researcher appointment is available in NIH manual chapter 2300-308-1 at <https://oma1.od.nih.gov/manualchapters/person/2300-308-1>

⁴ http://www.census.gov/about/adrm/fsrdc/federal_partners.html

comments on a new policy directive “Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies.”⁵ Interagency dialogue and assessments could identify how statistical agency policy development could be adapted and adopted for Federal research award and other administrative data.

Another avenue for coordination and information sharing has been the Interagency Working Group on Science Policy, coordinated by the White House Office of Science and Technology Policy (OSTP) and NSF’s SciSIP program. The primary function of this Group has been to identify opportunities to develop tools, theories, and methodologies that will advance their common interest, recommend joint research, data, and evaluation projects and enable Federal agencies to collaborate, coordinate, and leverage resources and efforts.

One question that the commission has asked about is alternative ways to share administrative data, providing access to researchers while guaranteeing confidentiality, privacy and data security. For the science agencies I would like to advocate for a decentralized system of agency specific data enclaves. There are three reasons.

First, there are many nuances to agency operations: the way programs are operated, data are collected, and even how variables are defined. Pece (2016) provides an overview of the challenges of obtaining reliable R&D data from federal agencies. Standardization is required in order to be able to make interagency comparisons. The National Center for Science and Engineering Statistics (NCSES) and the Office of Management and Budget

5

https://www.whitehouse.gov/sites/default/files/omb/inforeg/directive4/frn_comment_stat_policy_dir_4_addendum.pdf

(OMB) are working to establish a Federal R&D Community-of-Practice (R&D COP) to address agency questions on the interpretation of how to apply OECD Frascati manual. Adoption of standards would certainly reduce barriers to research using administrative data.

Second, my experience is that to best inform policy requires an understanding of context and the specific operation of the agency. The best efforts involve collaboration between academics and practitioners. Rather than simply taking data and conducting analysis there are many nuances and definitions to understand and appreciate and to make informed interpretations of the results.

Third, agencies have specific questions to address their mandate and to help them achieve their objectives. Policy relevant research projects are best-formulated and designed through interaction and collaboration. These interactions will benefit the scholarly community by providing more nuanced research questions and a deeper understanding the mechanisms that govern science funding and agency operations.

Some Science policy questions are universal across agencies but there are many hypothesis-driven studies that can be conducted using single agency data, perhaps linked to publications and patents. I think that there is plenty that we might say based on administrative records at the single science agencies. These would be questions about the effectiveness of different program interventions, selection processes and outcomes. Studies conducted at one agency can be replicated at other agencies.

Another opportunity is an expanded use of fellowships, and other temporary hiring authorities and mechanisms to bring in academic and other researchers

into agencies to access administrative data and conduct evaluations. Mechanisms to do this include the Schedule A excepted service authority and Intergovernmental Personnel Act.⁶ The contract terms outline the access to program administrative data and analyses that the researcher would conduct. There seem to be various mechanisms in place across Federal agencies. Sharing these experiences across the Federal Government could provide a set of lessons learned to improve adoption and use.

I believe that a decentralized system allows for experimentation. Over time better practices will emerge and there will be more confidence in the relationships. This model would also create a dialogue between researchers and the policy community that is likely to increase trust on the part of the agencies and to provide researchers with more interesting and informed questions.

One question that I ask myself is if these administrative data need to be linked to other sources and what would those other sources be. Ultimately there is a hope that individual researchers can be tracked over their careers, however building such longitudinal data takes time. U-Metrics at the University of Michigan is one effort to accomplish this task. With greater familiarity with agency data, new questions emerge. A viable second phase could link agency administrative data with other records, such as those at the Census Federal Statistical Research Data Center (FSRDC). Indeed, Sabrina

⁶ More on Schedule A available at <https://www.opm.gov/policy-data-oversight/disability-employment/hiring/#url=Schedule-A-Hiring-Authority>; the Intergovernmental Personnel Act Mobility Program provides for the temporary assignment of personnel between the Federal Government and state and local governments, colleges and universities, Indian tribal governments, federally funded research and development centers, and other eligible organizations, more information available at <https://www.opm.gov/policy-data-oversight/hiring-information/intergovernment-personnel-act/>

Howell has such a project underway to extend her EPA work by matching to applicant and other firms over time and working with the FSRDC.

Great work has been done at the FSRDC but if administrative data will be housed there then the Commission should address some of the current barriers and limitations. The application and vetting process is very lengthy and time consuming – unsuited to academic clocks which are short and tick loudly. Currently, proposals must demonstrate value to the Census Bureau under Title 13 of the US Code. That requirement might be too stringent considering there are many worthy projects that may not have direct value to the Census or another agency but might have great social value.

Currently, the federal agencies are working together to provide the data on funded projects to researchers and the public in a standard format for funded projects with the Federal RePORTER platform.⁷ I invite you to take a look at this website and its capabilities to query on topics across the participating agencies. There are negotiations to add other agencies. Indeed these efforts would be aided if the Commission could encourage or otherwise incentivize all the federal science agencies to provide their data.

Finally I would like to address something that I have noted that really concerns me and has potential to derail this undertaking. There is fear at many agencies that allowing researchers to work with their administrative records will reveal negative findings about the agencies. At a time when the federal science agencies feel under attack this is certainly understandable. But only by looking at data we can identify problems, bottlenecks and improve processes and efficiency. If we accept the premise that government programs are necessary to address market failures, facilitate network externalities,

⁷ <https://federalreporter.nih.gov/>

reduce barriers to collaboration and take on fundamental projects that are beyond the scope of most American businesses, then innovation and American competitiveness can be improved. It is only by studying the scientific research enterprise, the range of actors engaged in innovation and specific public programs and incentives that can scholars provide evidence to policy makers.

Administrative data related to science and innovation policy cover a large number of administrative units beyond the federal agencies. At least one entity operates in each of the 50 states, with a mission of promoting innovation, entrepreneurship or science-based economic development. On top of this, consider all the activity in firms, research universities, institutes and labs and other non-profit and quasi-government organizations that have some stake in the outcomes related to science, scientific progress and commercial outcomes. Each has administrative data that could be studied to improve our understanding of the research and innovation enterprise. The federal effort in which the Commission is engaged can lead the way and set a model that states and other entities can follow.

References:

- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333, 1015–1019.
- Howell, Sabrina T. (forthcoming). Financing Innovation: Evidence from R&D Grants. *America Economic Review*.
- Li, Danielle, and Leila Agha (2015). "Big Names or Big Ideas: Do Peer-Review Panels Select the Best Science Proposals?" *Science* 348: 434–438.

- National Science Foundation, National Center for Science and Engineering Statistics (2015). *U.S. R&D Increased in 2013, Well Ahead of the Pace of Gross Domestic Product*. Info Brief NSF 15-330.
- Pece, Christopher V. (2016). Putting the Cart Before a Lame Horse: A Case Study for Future Initiatives to Automate the Use of Administrative Records for Reporting Government R&D. OECD Bly Sky Meetings.