

Testimony to the Commission on Evidence-Based Policymaking
“Policy-Driven Demand for Government Evaluation: Data and Capacity”

Submitted by: Katherine O’Regan, Assistant Secretary for Policy Development & Research, Department of Housing and Urban Development
November 4, 2016

A. Importance of the Commission’s work

Good afternoon and thank you Commissioners for the invitation to join you today, to talk about data and capacity issues in government evaluation. Collecting, analyzing and using data and research to inform HUD’s programs and policies is at the core function of my office, the Office of Policy Development and Research. Like many other Federal agencies, HUD has been focused on learning more from the data we collect, and the research we conduct.

We believe that to make the most of government data and fully leverage how these data can inform policy, we need a leap forward:

- by making our administrative data much more widely available while protecting privacy;
- by matching them to other data – across government agencies, across levels of government, and non-governmental data, which is particularly key to expanding cross-domain knowledge; and
- By ensuring the capacity of governmental agencies to leverage linked government data to conduct in-house research.

We have access to our own administrative data but have both internal and external gaps.

- Internal needs include the staff time and capacity to do the research internally - with better infrastructure to support them, to ensure data privacy protections, and to more easily permit linking to data external to the agency.
- External needs include the ability to get outside researchers’ access to our data and, similarly, the infrastructure to support them, ensure data security, and enable linking.

The Evidence Commission’s charge, and recommendations, could create that leap forward, greatly improving the use of Federal data in building evidence for policy. I will start by providing some background on an area we have focused considerable effort.

B. Cross-Agency Data Linking

HUD has initiated several cross-agency data linking efforts to better understand non-housing domains of HUD assisted households, and to support evaluation work. I will focus on two efforts in particular, and some lessons learned.

1. Department of Education, Free Application for Federal Student Aid (FAFSA).

This year, HUD and the Department of Education (ED) signed an MOU that allows HUD to provide ED with data on assisted tenants for matching to Federal Student Aid data.

Benefits

We are using the linked data to learn about application for and receipt of Federal Student Aid among tenants, **and** to conduct some rapid-cycle experiments, aimed at increasing FAFSA take up and college enrollment.

- For the first time, HUD has national data on one measure of college attendance for those we serve.
- We are able to conduct rapid-cycle, in-house experiments.

- We are also able to provide aggregated data to Public Housing Agencies (PHAs) on the FAFSA take up for their assisted households, which some will use to evaluate their programmatic efforts – for the first time.

Lessons

- Those linked-data, however, do not sit in a secured research center with access for HUD or ED staff.
- *Lack of infrastructure inhibits easy in-house work, or for getting more out of the matches.*

2. HUD-National Center for Health Statistics (NCHS) linked data.

The second example is work with the National Center for Health Statistics, linking HUD longitudinal administrative data on assisted households to 14 years of cross-sectional health survey data from two national surveys (*National Health Interview Survey (NHIS)*; *National Health and Nutrition Examination Survey (NHANES)*).

Benefits

The linked data files enable researchers to examine the relationship between assisted housing and factors that influence health status, chronic disease, health care utilization, morbidity, and mortality. Benchmarking data and research have already been produced.

Lessons

- The existence of the CDC research center, with *appropriate infrastructure for linking and for handling privacy* was fundamental to the work.
- The MOU process allowed both agencies to conduct the linkage cost-free, and the interagency MOU underwrote free RDC access for HUD researchers, a significant cost barrier reduction.
 - *Sustainable financial models for data sharing need to be built in to such work.*
- Legal barriers due to data privacy concerns were very challenging.
 - Establishing the MOU between agencies took many years to negotiate.
 - *This project moved forward due solely to the prolonged commitment of a handful of staff.*
 - *As we continue down this path, a more centralized approach is much more efficient.*

Across both experiences, these one-off, time consuming interagency matches are not the most efficient way to exploit data and are difficult to sustain. This has led to our most recent effort.

C. **HUD MOUs with the Census Bureau**

HUD has entered into two agreements with Census to greatly broaden access to HUD data and matching with non-HUD data for approved internal and external researchers, within Census’s Controlled Access Infrastructure.¹

1. The first is an Inter-Agency Agreement (IAA) with the Center for Administrative Records Research and Applications (CARRA) to link data from HUD’s tenant databases and select randomized control trials with the Census’ survey data collection and other administrative data.
 - HUD’s tenant databases are already maintained within CARRA, which permit outside researchers to gain access via a secured Federal Statistical Research Data Center.

¹ Census has the infrastructure to support IT security requirements for research with PII and other highly sensitive data. As part of the infrastructure support, Census provides disclosure reviews and research assistance to authorized researchers with approved research projects.

- Via this IAA, HUD has sent the Moving to Opportunity (MTO) data to CARRA (the HUD voucher demonstration that Raj Chetty and co-authors linked to IRS data);
 - In FY2017, we plan to send the Family Options Study (FOS) data to CARRA.
- To jumpstart research projects, HUD will issue a solicitation for research using these data linked with other administrative data – all accessed through the Federal Statistical Research Centers that are the portal to CARRA.

Benefits

As you all know, and will be learning more about at upcoming convenings, OMB directed the Census Bureau to use appropriated funds for these purposes, and to administer the Evidence-Based Policymaking Commission.²

- Our agreement with Census is the first Federal effort (a ‘pilot’) in a larger expansion of Census’s data linkage infrastructure to support evidence-building and program evaluation.
- Our IAA and data sharing with Census will greatly increase external researchers access to HUD administrative and research data, within a secured infrastructure.

Lessons

While this pilot has just started, and we believe it holds the promise for how best to proceed with cross agency data matching, and making the most of archived research studies, we also believe:

- A sustainable cost structure will need to be developed for a broad group of Federal agencies to use this model.
 - For context, the current pilot costs \$50,000 per evaluation data set, *per year*, with potentially additional costs per researcher.
 - The longer-term costs need a model that does not consume agencies’ current research budget.
2. HUD and Census have also signed a Joint Statistical Project Agreement (JSPA) that will commit HUD and Census to partner on linking housing data to non-housing data sources already acquired by Census, to be available for internal researchers at each agency.

Benefits

- HUD staff working on these projects are able to access the data linked via CARRA remotely from a secured computing environment within PD&R/HUD, for much faster policy/evaluation work.

Lessons

- *Providing Federal staff with access to linked administrative data, those with program knowledge and the ability to move evidence much more quickly into policy is a critical piece of driving evidence-based policy.*

An additional lesson on the benefits of linked data in CARRA has already occurred on the interplay between data availability and what we study. PD&R is conducting our research-agenda setting process, one that engages with stakeholders internally at HUD, and much more broadly. As we winnow down the more than 500 submissions, some ‘big questions’ that might not have been asked - or small ones that might not have been prioritized – will be included in our Strategic Plan because, through data matching with the Census, we considered the research feasible. *The existence of a way to evaluate/study something does affect whether we prioritize it in our research agenda.*

² Evidence-Based Policymaking Commission Act of 2016

As we move towards establishing a clearinghouse for Federal data, such as CARRA, there are ways to make the most of those data.

D. Making the most of data within CARRA

1. Institutional Review Board guidance to permit indefinite data matching at CARRA

The recent work from Raj Chetty has brought attention to the benefits of data matching, and specifically to HUD's Moving to Work or MTO demonstration of the 1990s. In that study, HUD received consent from 4,610 families to conduct data matching against administrative records.

- The consent was not time restricted and it was general about the types of administrative data for the matching.
- This has allowed HUD to track the impact on the heads of households and their children for nearly two decades. It permitted the long term impact work of Raj Chetty. And it is allowing the MTO data to be made available at CARRA to measure impacts over even more years.

HUD's Family Options study, a randomized study to measure the benefits of receiving long-term versus short-term housing assistance for families with children who are homeless, also has research questions that extend out 20 years or longer. However, the IRB for this study limited the consent to 5 years.

- Although HUD will be providing these data to CARRA, the limited consent period applies.

Lesson

- It is reasonable to limit the time frame for consent for generic administrative matching.
- But if research data are eventually protected by the Census Bureau Title XIII, as part of the Federal clearinghouse, we are proposing that IRBs be given model language to allow indefinite consent under those conditions.
- When impacts take 20 or 30 years to materialize, we need to be sure that we will be able to measure them.

2. State data

Demonstrations and evaluations take place in a limited number of locations, and non-Federal administrative data is frequency crucial to many studies. There are many obstacles to accessing –and linking- those data too.

- In the long term evaluation of the Moving to Opportunity demonstration, researchers sought administrative data from multiple states on wages from unemployment insurance programs and on various benefits (TANF and SNAP).
- Data access varied considerably from state to state, with some states providing only aggregate data which limited HUD's ability to look at program impacts for specific subgroups. For one state, we are still working on securing needed data.
 - Note, some of the need to link state data is itself occurring after failure to get agreement on linking across Federal agencies for (in this case, Medicaid data).

Lessons

Consider the role of Non-Federal administrative data in your review

- Census has initiated a pilot with Chapin Hall, bringing in a limited set of state and local data into CARRA, as a proof of concept. This could be expanded.
- Any data inventory the Commission creates should include non-Federal data.

E. Additional Issues for the Commission's Consideration

1. Gathering data and the Paperwork Reduction Act

The Paperwork Reduction Act (PRA) requires that for any data collection involving 9 or more individuals or organizations, the data collection needs to go through a formal process of review that involves both a 60-day public notice in the Federal Register seeking comment to the agency and an additional 30-day notice for the public to provide comment to OMB. HUD's experience is that clearances take between 120 to 180 days, so for most questions, **it is 4 to 8 months before data collection can begin.**

PRA is an effective tool for reducing burden by preventing *unnecessary* data collection from the public. But the length of time to complete the process may also be inhibiting useful data collections that could improve policies and programs, and those sub-optimal policies may impose also impose severe burden on the public.

This is particularly relevant for HUD, which does not operate its programs directly. Instead it provides funding to local organizations to implement its programs. These organizations include Public Housing Agencies (3,400), state and local governments (1,200), lenders (several hundred), multifamily-owners of assisted properties (20,000), and non-profit homeless providers (10,000 or so).

Sometimes it is critical to get new information from these organizations to address an un-foreseen issue. PRA prevents collecting this information in a timely manner beyond nine of these organizations. If these programs were operated by Federal government employees directly we would not be prevented from collecting this information. As a result, it is not uncommon that HUD makes policy decisions and program changes with very little information; often reacting to anecdote rather than a more complete picture.

Lessons

- More robust incorporation of data and evidence in Federal policy requires improvements in PRA. There are various ways this could be done.

As one example: allow 30-day agency reviews without OMB engagement on collections below some thresholds, such as 1,000 responses and under 500 burden hours. This would take a 90 to 180-day process for these collections and turn them into 30 day processes.

There are numerous other suggestions that this commission would likely hear across Federal agencies, along a similar vein.

2. Address weakness of Privacy Act Protections

In a recent Freedom of Information Act (FOIA) case³, a Federal district court ordered HUD to release to the requester the individual records of tens of thousands of voucher-assisted tenants, omitting the most obvious identifying variables (name, SSN, address) but including zip code, census tract, age of all household members, race of head, and a wide variety of other data that in HUD's view could be used to identify households receiving HUD assistance.

³ "The Inclusive Communities Project, Inc. v. The United States Department of Housing and Urban Development," Civil Action No. 3:14-CV-3333-B

This was possible because the HUD administrative records, and records of many other agencies, are protected from disclosure only under the Privacy Act.

- Under FOIA, when records have only Privacy Act protection, the onus is on the government to prove the probability of harmful disclosure. In this particular case, the court felt HUD had not met the burden of proof, and presumably other courts might come to the same conclusion.
- The individuals whose lives are captured in the administrative and survey data that HUD uses for evaluation and analysis are therefore vulnerable in a way that they would not be if these data had the statutory protections that Census data, income tax data, or Medicare data have.
- HUD is concerned that the quality of administrative data may decline, if some individuals will eventually be less inclined to give honest answers to the questions we ask, or may decline to cooperate at all, if they may subsequently become subject to public exposure.

Lesson

- The Commission should consider recommending strengthening in statutory protection for administrative and survey data collected for agencies like HUD.

F. Conclusion

We appreciate the chance to provide these comments to the Commission, and the forthcoming conversations on data and capacity issues in support of government evaluation and policy needs.