Commission on Evidence-Based Policymaking
November 4, 2016
Open Meeting

Commissioners Present:
Katharine G. Abraham, Chair
Ron Haskins, Co-Chair
Sherry Glied
Robert M. Groves
Robert Hahn
Hilary Hoynes
Jeffrey Liebman
Bruce D. Meyer
Paul Ohm
Allison B. Orris
Kathleen Rice (via phone)
Robert Shea
Kenneth R. Troske
Kim Wallin

The open meeting was called to order at 8:30 AM. Chair Katharine Abraham and Co-Chair Ron Haskins provided an introduction.

Presentation 1: Evaluation Capacity Considerations from a Federal Evaluation Office: U.S. Department of Labor

Demetra Nightingale, Chief Evaluation Officer, U.S. Department of Labor

- Ms. Nightingale provided an overview of the role of the Office of the Chief Evaluation Officer (CEO) in the U.S. Department of Labor (DOL). The office works with over a dozen operating agencies within the Department on issues related to worker protection and labor standards, employment services, job training, and worker security. Many of the individual agencies in DOL have research, policy analysis, and evaluation offices. The office of the CEO complements and supports these other offices but does not centralize all evaluation functions.
- DOL develops an agency learning agenda that drives their evaluation and analytic activities. The Department's strategic plan, Congressional requirements, and guidance from the Office of Management and Budget (OMB) are the inputs to a learning agenda for each agency. The CEO's office helps agencies to prioritize evaluation and analytic projects.
- DOL performs formal evaluations using experimental and quasi-experimental designs, outcome evaluations through secondary analysis and modeling of data from the Bureau of Labor Statistics and the Census Bureau, program performance analysis, and implementation and management evaluations.
- Ms. Nightingale spoke about the importance of having appropriate outcome variables for evaluation. She said that wages and earnings are important, but other needed outcome measures include compliance rates and workplace injuries and fatalities. It is also important

to have information about labor market conditions and characteristics of individuals, families, and businesses.

- It is particularly important to have microdata, and preferably longitudinal microdata, rather than aggregates to conduct evaluation. DOL needs an agile way to merge data from multiple Federal departments longitudinally so they can use it more readily. Furthermore, it needs a data infrastructure that can be used by Federal departments but also by external researchers.
- Ms. Nightingale's "wish list" to improve data for DOL evaluations includes: 1) direct and less costly access to earnings data; 2) common firm identifiers; 3) streamlined Paperwork Reduction Act (PRA) requirements for evaluation; 4) streamlined Interagency Agreement (IAA) processes for cross-Department data sharing; and 4) clarity and consistency in privacy and security rules among Departments.

Discussion and Questions:

- Co-Chair Haskins asked Ms. Nightingale for the five most important things the Commission could recommend to make it easier for Federal departments to share data. Ms. Nightingale responded that she and her colleagues have a list and will submit it. Co-Chair Haskins further asked how to make it easier for evaluation studies to follow people for longer periods of time to measure longer-term outcomes. Ms. Nightingale replied that long-term follow ups require funding streams and evaluation contracts that span multiple years. She indicated that those kind of flexibilities require statutory or regulatory changes.
- Commissioner Hoynes asked about how the PRA affects DOL's ability to conduct evaluation and about the impediments to more interagency data sharing. Ms. Nightingale said that the PRA adds to the cost of evaluation because DOL has to pay contractors to work on clearance documents for four months to a year. She suggested that OMB could delegate some of its PRA functions to Departments that have the capacity in house. She said that the Commission should look at how data sharing works right now as legal and regulatory restrictions limit the ability to do comprehensive evaluations, the time they have to do them, and researcher access to reanalyze the same data.
- Commissioner Orris spoke about the PRA from her perspective at OMB. She acknowledged that there are delays and indicated that OMB would be open to conversations about improving the process. The statute requires certain processes that balance public transparency with timeliness.
- Commissioner Shea asked about the impact of DOL's evaluation work on policy, whether there was interest from policy makers, and any changes that happened as a result. Ms. Nightingale indicated that the White House and Congress have an increasing interest in evaluation because of their focus on evidence. Budget submission guidance from OMB requires an evidence justification. She said that we need better ways to communicate the results from evaluation to non-researchers. Labor Secretary Perez had asked about what works in job training on his first day and she gave him a 2-page memo summarizing 40 years of research. They've also found positive results for registered apprenticeships and are accumulating a new body of research around effective strategies for formerly incarcerated individuals.
- Commissioner Troske spoke about a study he worked on in which a letter from DOL asked states for unemployment insurance wage data. He was surprised that only 11 states gave permission and wanted to know how that could be improved. He also asked how evaluation studies could be completed in a timely fashion so that they can affect reauthorization and program funding rather than coming out after Congressional action. Ms. Nightingale replied

that DOL is doing better in getting access to wage data. The Department of Health and Human Services (HHS) has approved a request from DOL for access to the National Directory of New Hires for 12 evaluations. She said the ban on a national database for the Workforce Investment Opportunity Act (WIOA) has made it hard to link to outcomes data for evaluation.

Panel 2: Policy-Driven Demand for Government Evaluation: Data and Capacity Needs

Presentation #1: ***Katherine O'Regan, U.S. Department of Housing and Urban Development***

- Ms. O'Regan discussed how collecting, analyzing, and using data to improve policy is a core function at the Department of Housing and Urban Development (HUD). She noted that HUD is focused on learning more from the data it already collects. To fully leverage existing data, it needs to make administrative data more available and linkable while protecting privacy. She said that housing has a big impact on non-housing outcomes, so linking data is key to expanding cross-domain work.
- Ms. O'Regan provided some examples of cross-agency data linkages and discussed lessons learned. First, she spoke about an effort to link HUD to Federal Student Aid data. The linked microdata were not located in a Research Data Center (RDC) and HUD and Department of Education (ED) staff could not get access except to aggregate results. The infrastructure was not sufficient to make the most out of the data linkage and analysis.
- She also described a project to link 14 years of HUD data to cross-sectional health survey data. The National Center for Health Statistics provided data linkage and access services through an RDC. The agreement included cost-free linkage and free RDC access for HUD researchers but took years to negotiate because of privacy concerns and perceived legal barriers. Ms. O'Regan noted that both of these projects resulted in very time-consuming one-off matches and said that the best way to move forward is with processes that can be sustained over time.
- Two new agreements with the Census Bureau are designed to expand access to HUD data matched to non-HUD data. The first expands the data available to external researchers; HUD created a small pilot grant program to promote use of its data. The second aims to increase access to linked data for government researchers. This will help HUD recruit and retain high quality staff who want to do cutting-edge research.
- Ms. O'Regan indicated that there are two ways to make the most of data linkages at the Census Bureau's Center for Administrative Records Research and Applications (CARRA). First, the Commission could recommend that data being put in CARRA would be covered by Title 13 and therefore excepted from Privacy Act consent requirements. This would solve the problem of getting consent to measure long-term outcomes. Second, the linkages could include non-Federal (state and local) data, which are crucial to evaluation.

Presentation #2: ***Evelyn Kappeler, U.S. Department of Health and Human Services***

- Ms. Kappeler described the tiered evidence program for teen pregnancy prevention in the Office of Adolescent Health. It has provided 102 competitive grants totaling over $100 million to eligible organizations.
- Most of the funds ($75 million annually) go to Tier 1 projects designed to replicate evidence-based programs. Another $25 million annually goes to Tier 2 projects to develop and test

new and innovative approaches to teen pregnancy prevention. Some funds are also used for program support and evaluation.

- Grantees were required to use a uniform set of performance measures and were encouraged to use those measures for their own program operations. All grants required an evaluation by an independent third party.
- HHS established conditions for grantee evaluations to maintain high levels of rigor and included those conditions in grant announcements. It was important for grantees to allocate sufficient resources to ensure statistical power and rigor in evaluation, so grantees were required to reserve a portion of their of funds for evaluation. Continued funding was contingent on the grantee meeting evaluation standards. At the end of the evaluation, grantees were required to archive the data so they would be available to future researchers.
- Ms. Kappeler spoke about the need for capacity to promote rigorous program evaluation. She indicated that it is important to provide support and a 6- to 12-month period of planning, piloting, and preparing to conduct an evaluation. She said communities need the capacity to implement programs with fidelity and quality. She spoke about the difficulty in incorporating evaluation once a program has begun, especially for recruitment. Finally, she said that without prior planning it is difficult to ensure a strong contrast between the treatment and control groups.
- As a result of the experience fostering evaluation of teen pregnancy prevention programs, Ms. Kappeler offered four suggestions about building capacity. First, establishing and promoting uniform evidence standards is important. Second, there should be clear communication about expectations for evaluation along with training and technical assistance to meet the standards. Third, programs should continuously collect data on implementation and use it to make quality improvements. Finally, all evaluation results should be disseminated transparently, including null findings.

Presentation #3: ***Matthew Klein, New York City Center for Economic Opportunity***

- Mr. Klein described the Center for Economic Opportunity in the New York City Mayor's office, which he directs. His office seeks to help the city reduce poverty by advancing the use of evidence and data in program and policy design, service delivery, and budget decisions. The office looks for opportunities to make it easier for the city to use data to give better service. As a Results for America fellow, Mr. Klein trades notes with his counterparts in other cities about data sharing and use.
- Cities are unique in that they are the closest level of government to residents and receive direct and candid feedback from them. A challenge is to engage residents holistically rather than through silos. Cities are creators and users of evidence. They provide funding and real-world data for research and use evaluation to inform policy making and ensure accountability.
- Mr. Klein described how New York City is using administrative data to improve services and outcomes for residents. He gave examples of studies on college persistence, workforce development, and job training for public housing residents. He also discussed new efforts for citywide data integration and common data metrics.
- Mr. Klein made three recommendation to the Commission. First, he suggested that the Federal government should allocate more funding to evidence-based strategies. Second, he encouraged the Federal government to continue developing common outcome metrics.

Third, he suggested the Federal government amend laws such as HIPAA and FERPA so local governments can share data across their own agencies.

Discussion and Questions:

- Commissioner Groves asked Mr. Klein about the difficulties of cross-agency work and about the distinction between statistical uses and program uses of data. Mr. Klein responded that one challenge of cross-agency work is agreeing on the most important questions to answer. He also spoke about the need to take advantage of new data sharing technologies and to address privacy and legal challenges.
- Chair Abraham asked Ms. O'Regan about HUD giving its evaluation data to CARRA. Ms. O'Regan replied that HUD retained the right to approve or disapprove projects using its data. She also indicated that it might actually be easier and faster to have HUD approve the projects and have CARRA qualify the researcher.
- Commissioner Meyer thanked Ms. O'Regan for her written testimony on the PRA.
- Commissioner Ohm asked about machine learning techniques and whether it raises new and different data management and access questions. Ms. O'Regan said that HUD is now moving in that direction and working to modernize its legacy computer systems. Mr. Klein said that New York City is just beginning to explore machine learning and descriptive analytics.
- Commissioner Hoynes followed up and said that machine learning uses the same data with a different tool. She then asked Ms. O'Regan about the role of Institutional Review Boards (IRBs) and consent in HUD evaluations. Ms. O'Regan replied that all of the third party independent evaluations go through IRBs and there is a movement towards greater restrictions on consent.
- Commissioner Wallin asked if there is a formal cost/benefit analysis in these evaluations. Ms. O'Regan replied that different types of evaluations have different goals. The family options study, for example, looked at both outcomes and costs. Ms. Kappeler indicated that her office is starting to look at the differences in costs for grantees to implement the same program.
- Co-chair Haskins asked the panel about giving agencies the ability to share personal data without explicit consent. Mr. Klein responded that the issue of student privacy is significant and a legitimate concern to grapple with. He said that government should be explicit that evaluation purposes are allowed.

10 minute break at 10:00 AM.

Panel 3: Non-Governmental Demand for Evaluation: Capacity to Support Public Good Activities

Presenter #1: *Tanya Beer, Center for Evaluation Innovation*

- Ms. Beer spoke about the Center for Evaluation Innovation, a small nonprofit organization that focuses on building evaluations in fields that are hard to measure. It works with a network of foundations that seek innovation in program delivery and evaluation.
- Ms. Beer discussed five trends in philanthropic and nonprofit work and how those trends relate to evaluation and data collection: 1) nonprofits are moving from traditional charity giving to deep attention to effectiveness and outcomes; they are also increasingly interested

in real time results to improve decision making; 2) nonprofits are increasingly focusing on complex problem solving and systems change, so evaluation needs to include system as well as program level outcomes; 3) nonprofits seek to scale pilots in partnership with public and private sectors but are stymied by the different sectors' priorities and demands; 4) nonprofits pay disproportionate attention to performance metrics which suck up evaluation capacity and don't generate useful evidence; 5) nonprofits have limited capacity to use evaluations to make better decisions and are now experimenting with strategies to do this better.

- Ms. Beer indicated that capacity is still a big challenge; it is a bandwidth and budgetary constraint.

Presenter #2: ***Jim Sullivan, University of Notre Dame***

- Mr. Sullivan said that the Federal government spends about $800 billion a year on social programs and only about 1 percent of them are backed by evidence. His main point was the need to look at what is happening outside of Washington, D.C. to learn how to design and evaluate programs.
- The human services sector consists of programs on topics such as job training, youth development, and homelessness and has about $200 billion in funding. These programs and resources could inform evidence-based programs and policies.
- Programs can be designed at the local level to meet local needs and have impact evaluation built in. The evidence generated from those evaluations can be a spring board for scaling programs to the national level for national impact. What typically happens is that an innovative local leader identifies a need, designs a program, and launches it. There is no money for evaluation but the program is scaled up with no evidence. Once it's scaled up, it is much harder to use evidence to reallocate resources.
- Obstacles to evidence-based programs include funding, getting the outcomes data necessary to do impact evaluation, and the difficulty of replicating evidence-based programs.
- Mr. Sullivan suggested that the Commission could help by providing incentives for programs to conduct impact evaluations, making administrative data more accessible, and facilitating the dissemination of best practices.

Presentation #3: ***Adam Gamoran, William T. Grant Foundation***

- Mr. Gamoran focused on the use of data to serve the public good. He said that two key issues are what data are needed and what steps are needed to build capacity to use them. He described the William T. Grant Foundation and its support of social science research to improve the lives of young people by reducing inequality and improving the use of research evidence.
- Mr. Gamoran said his message is that that this should be the "evidence" commission, not just the "data" commission and asked Commissioners to consider how to make that happen.
- Mr. Gamoran called for research evidence derived from applying systematic methods and analysis to questions or hypotheses. Policy insights can emerge from rigorous experimental and quasi-experimental evaluations of programs for program improvement and policy formation. Mr. Gamoran said such studies must consider program context, implementation, and conditions that facilitate or impede success. He suggested that the Commission could provide guidelines for Federal data managers about data sharing for program improvement.
- Mr. Gamoran gave two examples of projects that require linked data to figure out how to reduce inequality. The first was linking national education surveys to state education records

and the second was the American Opportunity Study. He indicated that it is important to build the capacity to use data to create evidence. In particular, he suggested that the Commission should recommend attention to technology, expertise, organizational arrangements, and fiscal investments; should strengthen the Federal infrastructure for evidence; and should establish partnerships between researchers and policymakers.

Presentation #4: *Kelly Fitzsimmons, The Edna McConnell Clark Foundation*

- Ms. Fitzsimmons gave an overview of the goals of the Edna McConnell Clark Foundation and discussed its focus on funding projects to encourage the use of data to build evidence. She said that the foundation partners with other foundations and government to encourage evidence building and that they have learned a lot about the challenges of blending public and private funds.
- Ms. Fitzsimmons indicated that grantees have benefitted from rigorous impact studies even though in most cases the studies have yielded mild to null findings.
- Ms. Fitzsimmons addressed four constraints for using and building evidence. First, funding and capacity is limited; it's hard to find talent and fund the work out of overhead. Second, the incentives for evaluation are not aligned with practitioner needs; the timelines, expense, and pressure to not publish null findings are not conducive to innovation and risk taking. Third, the timeline for evaluation is not synced with policymaker demands for evidence. Fourth, policymakers and funders demand evidence without understanding what it takes to create it and then don't use the evidence consistently to set policy.
- Ms. Fitzsimmons suggested four potential solutions to these constraints: 1) engage service providers in the process of evidence building; 2) develop incentives for using evidence in procurement; 3) establish clear standards for evidence and data; and 4) remove barriers to accessing Federal, state, and local data, especially in education.

Discussion and Questions

- Commissioner Troske said that he tends to agree that performance metrics are not evaluation. He asked if performance metrics use data the program is already collecting while evaluation is more likely to need linked data. Ms. Beer responded that performance measures are important but there is a gravitational pull that gets them reduced to what is easily countable rather than how program performance affects the target audience. She said this is a cautionary tale about staying focused on results and not just counting participants and activities. Mr. Sullivan added that funders ask grantees to collect that kind of program performance data and that there needs to be more focus on evaluation from the funders' perspective. He said that access to administrative data will help with performance metrics and with studying outcomes for people after program participation.
- Commissioner Troske said that we seem to view randomized control trials as a gold standard, but they are expensive and take time to conduct. There's also a large focus on impact on average person treated which is not always a useful metric for decisions about eliminating or continuing programs. He asked what evaluation can do to get better information on what parts of a program are working. Ms. Fitzsimmons gave an example of a program in New York that couldn't randomize participants but still used administrative data for evaluation and a program in Chicago that used the Current Population Survey from Census and local police records. She said that these kind of projects can cut through cost and data constraints.

- Co-chair Haskins asked how to address the problem that most evaluations have null or negative findings so programs don't want to be evaluated. Ms. Fitzsimmons said that leadership is a critical part of the answer. Leaders should do due diligence and promote the adoption of programs that work. She also said that leaders should have incentives to use evidence and that funders should support smaller scale targeted studies to build evidence. Ms. Beer said not to miss the nuance of what works for whom and under what conditions; evaluations need to communicate context and implementation decisions so other sites can make adaptations as appropriate. Mr. Gamoran said that the point of evaluation is program improvement. He added that linking rich data about context with implementation and outcomes will promote studies that look at program improvement. Ms. Beer added that the evaluation designs for the Tier 1 studies in the i3 program were independently reviewed and suggested that such an innovation could strengthen evaluation design and execution.
- Commissioner Groves asked for insight about how participants in evaluations of foundation-funded programs think their data will be used and whether there is a culture of informed consent. Ms. Beer said that tight connections and trust in the community being served builds relationships that breed participation and data sharing. Mr. Sullivan said that consents do not go far enough and that adding a clause saying these data can be linked for research purposes would help programs access outcomes. Mr. Gamoran said this is the right question to ask and noted that most people freely share their personal data for commercial purposes but when it comes to the public good they are scared about how their data will be used. He suggested that the Commission should focus on the public good. Commissioner Ohm noted that the idea that most Americans are happy sharing their information for commercial purposes is overstated and that surveys show people feel helpless about this.
- Commissioner Hoynes talked about a challenge of cost/benefit analysis in that typically program cost is easy to measure but the benefits accrue over a lifetime. Since there is a limitation on what we can learn in the short run, she asked, how can we make good use of evaluation? Mr. Gamoran said that this challenge amplifies the importance of the Commission's work because of the need to link to data sets with long-term outcomes.
- Chair Abraham asked Mr. Gamoran about the importance of state data for evaluation and how he would design a pilot for the Federal government that creates incentives for states to share data related to Federal programs. Mr. Gamoran replied that the Arnold Foundation has been working with Chapin Hall at the University of Chicago to design a series of pilots. He also suggested that states in these pilots will learn a lot more than other states about their own programs and this should be an incentive to share data.

Presentation 4: Evaluation Capacity Considerations from a Federal Evaluation Office: U.S. Department of Health and Human Services, Administration for Families and Children

Naomi Goldstein, U.S. Department of Health and Human Services

- Ms. Goldstein described 5 principles for evaluation—rigor, relevance, transparency, independence, and ethics. She then made seven main points:
    1. Data are necessary but not sufficient to create evidence. The evaluation design must be in place before data collection. Data must be examined using a sound analysis plan. Without rigor, evaluations may generate wrong answers. Rigor is not a code word for randomized control trials; rigor is needed in all kinds of evaluations.

2. Evaluations need specialized data to learn about program effectiveness. Some information needed for evaluations will not be available from existing administrative or statistical data.
3. The data that are available shouldn't be the starting point for evaluation, but evaluation should use existing data when it can.
4. Relevance is as important as rigor; it doesn't make sense to produce work that is elegant but useless. Information on implementation and context is critical for scaling and for figuring out what aspects of the program design enhance or inhibit results.
5. Federal evaluation needs statutory authority and funding. This is obvious but important. Most funding is linked to specific programs but there is not capacity for cross-program evaluation. The Federal workforce and contractors need expertise in evaluation. Because evaluations have different staff requirements for different stages from data collection to analysis, it's efficient to contract out for expertise. Contracting out for evaluation also supports independence.
6. Bureaucratic challenges and barrier exist that limit evaluation, including related to procurement, information collection, and information technology and security.
7. Data collected by statistical agencies are protected under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) but data collected by evaluation offices are not. Statistical agencies have OMB standards and infrastructure supporting and protecting them, but there is no statement of principles and practices for Federal evaluation. The Institute of Education Sciences in ED has the independence of evaluation in its statute but other evaluation offices do not and Department leaders could compromise independence.

- Ms. Goldstein concluded by saying the mission of the Administration for Children and Families (ACF) is to foster health and wellbeing. ACF needs to evaluate its activities to continually improve service delivery. The Commission has an opportunity to enhance this work by recommending further use of data to create evidence for learning rather than just answering yes or no questions.

Discussion and Questions

- Commissioner Orris asked about institutionalizing Federal capacity for evidence building and whether the Commission should look at statutory changes, more support from OMB, more money, etc. Ms. Goldstein replied that all of those approaches (except more money) have pros and cons. She said it is not helpful for statutes to have too much detail but it useful to have broad statements about learning agendas, rigorous approaches, and authority to use resources for evaluation. OMB's role with the statistical system and emerging coordination of evaluation has been helpful, but she is skeptical of centralization. OMB or another entity should play a leadership role in Federal evaluation. The ACF Principles and Practices for evaluation have gotten traction and are being used. The Commission could make statements of principle that would help. Ms. Goldstein added that randomized control trial evaluations are expensive, hard to implement, and slow. She said that other kinds of studies have flaws too.
- Co-chair Haskins asked how to not always find that programs don't work. Ms. Goldstein replied with two suggestion: implement programs that do work and do rapid cycle improvement studies to answer smaller quicker questions about how to improve practice.
- Commissioner Shea asked about human capital challenges, technology challenges affecting privacy and security, and procurement challenges. Ms. Goldstein answered that there is a

deep pool of talent filled by the increase in graduate education in related disciplines. Actually hiring that talent is extremely challenging. For IT challenges, Ms. Goldstein indicated that it is important to increase capacity and focus on designing information systems so they can talk to one another. Consider both privacy and technology.

- Chair Abraham asked about the right way to structure the evaluation function in the Federal government. Ms. Goldstein said that it may vary by agency. She said the Department of Labor evaluation function did not always exist and did not come about by accident. There were times when the function was weakened and reports suppressed. There was a series of GAO reports that led to encouragement from OMB to develop the current evaluation structure. Ms. Goldstein added that some degree of centralization is important. In ACF they have some centralization, which provides a critical mass of talented people, but some program offices have their own evaluation shops also. The Interagency Council on Evaluation Policy is collecting information on different structures across agencies.
- Commissioner Hoynes asked for thoughts about the lack of resources for assembling, cleaning, and documenting administrative data and whether that capacity could be built into an evaluation structure. Ms. Goldstein replied that ACF has established a new division of data and improvement to work with program offices to improve the quality, usefulness, availability, and analysis of administrative data. There's no one right way to structure it.
- Commissioner Troske asked Ms. Goldstein how she would structure the evaluation function if she were the czar. Ms. Goldstein replied that she would streamline the Paperwork Reduction Act requirements for evaluation, establish an entity with responsibility for evaluation coordination and leadership, create special pathways for hiring evaluation staff, and allow incremental funding of nonseverable contracts.
- Commissioner Orris asked Ms. Goldstein to elaborate on the challenges of states sharing administrative data. Ms. Goldstein replied that there is a complex web of laws and rules at the Federal, state, and local levels. There are efforts to develop standardized agreements but it is a difficult and incremental process.

Chair Abraham adjourned the meeting at 11:55 PM.