THE BROOKINGS INSTITUTION

COMMISSION ON EVIDENCE-BASED POLICYMAKING

KEY CONSIDERATIONS FOR THE COMMISSION
RELATED TO EVALUATION

Washington, D.C.

Friday, November, 4, 2016

PARTICIPANTS:

Call to Order:

KATHARINE G. ABRAHAM, Chair
Director, Maryland Center for Economics and Policy
University of Maryland

RON HASKINS, Co-Chair
Senior Fellow, Economic Studies
Brookings Institution

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

DEMETRA NIGHTINGALE
Chief Evaluation Officer
U.S. Department of Labor

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

KATHARINE O'REGAN
Assistant Secretary for Policy Development and Research
U.S. Department of Housing and Urban Development

EVELYN KAPPELER
Director of the Office of Adolescent Health
U.S. Department of Health and Human Services

MATTHEW KLEIN
Executive Director
New York City Center for Economic Opportunity

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

TANYA BEER
Associate Director
Center for Evaluation Innovation

JAMES SULLIVAN
Co-Founder, Lab for Economic Opportunities
University of Notre Dame

ADAM GAMORAN
President
William T. Grant Foundation

KELLY FITZSIMMONS
Director of Innovation and Policy Planning
The Edna McConnell Clark Foundation

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

NAOMI GOLDSTEIN
Co-Chair of the Interagency Working Group on Evidence Policy
Deputy Assistant Secretary for Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

* * * * *

P R O C E E D I N G S

MS. ABRAHAM: I think we can go ahead and get started. I'd like to welcome everyone and thank you for joining us at the third meeting of the Commission on Evidence-Based Policymaking.

We have had two full meetings of the Commission so far. Our first meeting in the summer was a chance for us to gather for the first time and get introduced to the Commission's charge. Our second meeting was devoted entirely to thinking about issues related to privacy and the importance of privacy considerations in developing our findings and recommendations.

We are now turning to the field of program evaluation, which is obviously a critically important community for us to hear from in terms of what needs to happen to generate the relevant and rigorous evidence needed to support policymaking.

We have several goals for today. We want to develop a better understanding of some of the barriers faced by evaluators in terms of generating evidence, and also to understand some of the opportunities for how better access to data and the evaluation that could result from that can support and improve programs that serve the public good.

We are really looking forward to hearing from our nine speakers. I don't want to talk too long because we are really here to hear from them. We are certainly looking forward to leaving today with a better understanding of where the gaps of access to data are and what might be done about that.

I might mention that in addition to the formal meetings of the Commission that have been held, two weeks ago we held the first of three planned public hearings to hear from members of the public and others who wanted to speak to us.

We heard at that public hearing from 16 witnesses that represented a variety of organizations, including representatives from the American Evaluation Association, the Pew-MacArthur Results First, Results for America, and a number of other organizations.

If you're interested in that hearing, the written statements that were presented and the video recording of the hearing are available on our Web site, which is CEP.gov.

I was surprised, I will say, to learn that more than 800 people had already viewed the hearing. I didn't realize this was such an exciting topic for members of the public to come in and learn about.

I do want to thank you all for joining us and for what I'm sure will be a productive discussion this morning.

Ron?

MR. HASKINS: Thank you. We have already collected a ton of information about programs that could be better, have more information for policymakers about, so they can improve government programs, and also for researchers. We are going to have plenty more occasions to hear from the public and from professionals and anybody who thinks they have something to say to us.

The Commission's broadest charge is to identify a strategy for ensuring that evidence increasingly informs public policies, so that is why we are so anxious to get views from the public.

For those of you who would like to give us your ideas about how we can achieve this goal, we are accepting input directly through requests for comments that was advertised in the Federal Register. We also, as Katharine mentioned, are going to have two more public hearings. We have had one already. We will have one in Chicago, and

one somewhere on the West Coast. I don't think we have decided yet, have we?

Before we lose control here, something that I'm quite used to, I encourage everyone to provide us with information and be as specific as possible. There are lots of generalities that we hear repeatedly, and specific ideas, specific recommendations, what you think the Commission should report on and what it should consider. That is really the most important thing that we should hear to help policymakers and researchers do better.

As many of you know, today, just a few blocks away, the Association of Public Policy Analysis and Management is hosting its annual conference. There are several of us that are involved in that conference, like our first witness. We are giving her an award today at APPAM, well deserved. We are going to be going back and forth a little bit. We are going to try not to be disruptive when we do that, but we hope you will forgive us if we do.

Turning to our speakers, we would ask you limit your presentation to 12 minutes or less. We like less, to allow us time for questions. We think the question period is important. Please be sure to stay within the limit, but just in case you have any trouble, right over there somewhere is a little device. At two minutes, it will turn yellow. At 30 seconds, it will turn red. At zero seconds, it says "end," and there is a fine spray that will leap across, and the spray is deadly. (Laughter) I am sure none of you will want to violate the time limits.

Our first witness is Demetra Nightingale, whom I just mentioned. She is in an extremely important position that I hope replicates in other agencies to increase their evaluation capacity.

She is the Chief Evaluation Officer of the U.S. Department of Labor. I have talked with her on several occasions, including once at lunch, and heard a lot about

what she is doing. This is an ideal thing. This is really a good idea. In fact, I'm going to suggest we have something like this in our report, to have someone in each agency that does a lot of research that is in charge of evaluation and above all, that has a budget.

Demetra has done a tremendous job. We are so happy to have you here, Demetra.

MS. NIGHTINGALE: Time, okay. Thank you.

MR. HASKINS: By the way, these mikes, these are great mikes.

MS. NIGHTINGALE: Stop the timer. (Laughter)

MR. HASKINS: Keep it going, she's losing time every second.

MS. NIGHTINGALE: I've already lost five minutes.

MR. HASKINS: Okay, listen. This is really important. These mikes are really good but if they are away from you like this, they don't work at all. You have to bring them in close and push the button so it turns red, and don't get confused, that doesn't mean stop. That means speak.

MS. NIGHTINGALE: Okay. Thank you. I am Demetra Nightingale from the U.S. Department of Labor. I'm going to tell you a little bit about what we do, the kinds of evaluations we do, how we get data, and some of the issues around accessing information that we need and use, and then I will end with some suggestions that would help us to do evaluations better, which I think fall under the Commission's purview.

First, for those who don't know, the Department of Labor's mission is to improve the well-being, welfare, and employment related activities for wage earners, job seekers, and retirees of the U.S. We also work in other countries as well.

There are over a dozen operating agencies in the Department, many of them have sub-agencies, lots of programs, and over 17,000 or 18,000 staff around the country.

We have responsibility for worker protection and labor standards. That includes agencies like OSHA, Occupational Safety and Health, but also the Mine Safety and Health Administration, the Wage and Hour Division, which enforces minimum wage, overtime, and child labor laws. The Workers' Compensation program.

We also have employment services and job training and employment training, the Employment Training Administration, and then the Veterans employment and training service, and unemployment insurance.

We also have policy advocacy organizations, through the Office of Disability Employment Policy, the Women's Bureau, and the International Labor Affairs Bureau, that works with other countries doing grant making and evaluations of those grants around worker protection.

Many of the agencies in the Department have research analysis and/or evaluation offices.

Our office, the Chief Evaluation Office, is a departmental evaluation office that compliments but doesn't centralize all evaluation activities. We have an important role as being the hub of evaluation and data issues for evaluation and analysis. We have an evaluation budget and authority that continues to grow each year, over the past five years.

Our job is to raise the quality of evaluations that are done, the awareness and understanding of what evaluation means, and importantly related to this Commission is to improve the quality of, use of data, and we have a Data Analytics Unit that does consultative analysis with all of the agencies, particularly around using the administrative data within the Department, both on the employment and training side and on the worker protection side. We coordinate very closely with the Bureau of Labor Statistics and the Performance Management Center on data and analysis issues.

This chart gives you a little sense. People always ask how do we come up with what evaluations we are going to do. This chart lays it out pretty simply, I think, which is we look at the strategic plan for the Department and the priority goals and objectives that are in the Department's strategic priorities, strategic plan. Congressional requirements. Often Congress has statutory requirements for evaluations. OMB guidance around evaluations.

We take all of those pieces and the statutory mission for each of the agencies and develop what we call learning agendas. Each agency in the Department develops a learning agenda in coordination with my office, and comes up with a priority list of evaluation and research and analysis, data analysis topics and projects that would be of use.

On this chart at the bottom on the left gives you a sense of the kinds of capacity activities we are involved in, one of which is data quality and access, both for good evaluations and for data analysis, both internally and through modeling. On the right, it gives you a sense of some of the kinds of data analysis that our Data Analytics Unit does.

The types of evaluations we do are standard that many in this room know about. We have formal evaluations using experimental design random control trials and treatment and control groups to measure and estimate formal net impacts between a treatment group and a control group.

We also have many rapid cycle behavioral tests, sort of the nudge test. We have about a dozen of those going on. All those typically use administrative data to test different process changes.

We also do quasi-experimental analysis with created comparison groups and statistical matching techniques, so we have a treatment group and comparison

groups, and various levels of analysis, at the national level, state level, local level, grantee level, program level, demonstrations, strategies and models.

The outcome analysis that is done uses mostly econometric statistical modeling, micro simulation work, again using big databases. We do a lot of survey analysis often in collaboration with BLS and statistical analysis with BLS and Census data.

I just want to put in one little plug for implementation and management evaluations, which is very important and uses both quantitative and qualitative data, usually at the program administrative level and often with surveys.

What are some of the priority data issues that we need? We need to have the appropriate outcome variables, usually earnings and wages and income, but also the information that might be in evaluations of some of our enforcement agencies, which would be compliance rates and activities related to the Fair Labor Standards Act compliance, workplace injuries and fatalities for some of the worker protection agency evaluations we do.

We need appropriate independent variables. This is critical for evaluations. We need as much independent variable information that we can get, both about the labor market conditions, the characteristics of the individuals and families, but also business characteristics on the firm side as well, and you can see some of the other variables that we typically need.

The time frame has to be aligned to the evaluation goals, whatever that might be, and each evaluation is different. What is important is we need pre-program earnings as well as post-program earnings to do a good evaluation. Pre-program may include information on education, earnings, but then on the firm side, could mean business functions, business characteristics, and enforcement history.

We always want micro level data, as micro as we can get it, aggregate data doesn't help much with evaluations. We need longitudinal data, where the information can be linked together, and related to that, we need an agile way to merge data longitudinally. We need to be able to link data from different sources together, and using unique identifiers, and often for individuals that is Social Security numbers.

We are working across Federal agencies led by OMB to think about firm and establishment unique identifiers that could be more readily used, but we absolutely need to be able to link the data that exists, and critically important for evaluations, we need to be able to link other evaluation data that are being collected over time, whether it is from surveys or program data that may be from the field, from states or from non-profit organizations/service providers.

We need a way we can continuously add to the research data files that we are using.

The priority data systems and issues for evaluations, we also need to make sure the data that we are accessing have the adequate data infrastructure that we can use at the Federal level, whereas the staff analysts are using it, but also by third party researchers who may be in universities or in contractor organizations.

For validation, we need to know the underlying data, the source data, and that usually means before it has been cleaned and scrubbed.

We need again to have timely access. Security. We have very strict security and informed consent procedures that we need, and cost efficient. It is not unusual when we have a national evaluation that it could cost over $1 million to get the earnings data that we need from states. We always get what we need, but sometimes it costs a lot and takes a long time.

So, the wish list. To improve the data for evaluations. For many of our

employment and training evaluations, earnings data is what we need. We need a direct and less costly access to earnings records, particularly those that are produced already by the Department of Labor's own state employment security agency partners, and particularly the national directory of new hire's and LAHD both get quarterly earnings data from our state agencies, but there are statutory limitations that mean we cannot access that data at the Department of Labor. I hope that will be changed.

We need firm identifiers, as I've already mentioned. I would hope that the Commission will also look at reforming the Paperwork Reduction Act to make it a more streamlined, less costly process so we don't have to halt or stop evaluations because we are still waiting for PRA approval.

We need a more streamlined process for sharing data across agencies. We all, I think, want to work more across agencies, but the cumbersome IAA process right now is very limiting in terms of being able to data share and match at the Federal level.

Again, having some more clarity and consistency in procedures across Federal agencies by Federal agencies about sharing data and about privacy and security information.

I will stop there. I have more that I could say, but here on the last slide is how you can reach me. We have our Web site that would have the evaluation policy statement about data integrity and independence, and our clearinghouse, which has evaluation guidelines and standards.

Thank you.

MR. HASKINS: Let me make a recommendation first, personally, you are free to ignore it if you like. We have heard a lot about interagency issues with data. I think it would be really helpful -- I realize it is the end of the Administration, and we are

still going to be seriously engaged until spring, I think it will be some time in the spring before we start writing our report -- if there were an interagency group of some kind that could send us the five most important things that we should recommend that would make it easier for agencies to cooperate with each other, including data sharing.

I think that would be really helpful. Maybe OMB could help facilitate that. It's not the kind of thing that would take six months, in one or two meetings, I think you could come up with a great list.

MS. NIGHTINGALE: We already have a list. We will write it up. That's great.

MR. HASKINS: What I am thinking is some of our recommendations are going to require legislative action. You have mentioned several. If we can say in our report that five agencies agreed and so forth, I think it will provide a little political backing.

I want to ask you one question. Especially after work on Moving to Opportunity, I think a lot of people already knew this, the further we can follow samples, not only the more we are going to find out, but sometimes there are big surprises, like Moving to Opportunity was really a surprise. We would never have known if we didn't follow up 10-15 years later.

With the data that you have, especially on wages, employment, and income, that is exactly the kind of thing we need to know in the long term, have you thought about this? Have you tried to figure out ways that studies that you fund or states or other agencies can follow people for longer periods of time, and do you have recommendations about things that we could recommend that would facilitate that process?

MS. NIGHTINGALE: Yes. What we try to do, for example, right now, we probably have three evaluations, live scale, formal evaluations, where we have an

additional contract, where we are extending follow up for several more years, which would give us 10 years or longer for some of these evaluations. That's important.

What can be done, I think, is again some kind of statutory or regulatory changes that have to do with the length of time that a contract can be allowed, because right now, we have limitations on how long a Federal contract can exist.

It would help, for example, if we had no year funding rather than funding that had to be obligated or expended within a certain period of time to allow the evaluations to go out over a longer period of time.

We can include those kinds of suggestions also in material that we can send to you.

MR. HASKINS: Other questions from members of the Commission? Hilary?

MS. HOYNES: Thank you for your presentation. It was a great way to start out the morning. I was wondering if you could talk a little bit more -- I very much wanted to second what Ron had to say and get input as to what the major impediments are.

I wonder if you could talk a little bit about your view about the Paperwork Reduction Act and how that sort of affects things on the ground. I wanted to ask one other thing on a different topic, and you can answer them however you like.

You just talked about extending contracts in order to have longer term follow ups. What is also happening in the research world that I live in is that people are very -- outside researchers are very motivated and interested in engaging in these kinds of activities as well.

I wonder in terms of the issues around interagency sharing and sharing more broadly what your thoughts are and impediments to that.

MS. NIGHTINGALE: On the PRA, you might want to have a whole session devoted to the PRA issues. I think the PRA serves a critical important function to make sure that we are not overburdening the respondents in data collection, and that we are cognizant of data security and data privacy.

What happens is that over time, there has been a little bit, more than a little bit of mission creep to include requiring PRA submissions to OMB and clearances for even minimal, non-invasive, non-sensitive evaluation information, rather than just focusing on program data collection that affects the general public.

The PRA requirement adds costs to each of the evaluations because the contractors -- most of our evaluations are done by independent outside contractors for transparency and independence, and it adds extensive costs and time to the contract because the contractor has to develop -- sometimes it may take four or five months to develop the PRA package, and then it could take -- some of our PRA packages have taken longer than a year before we hear back from OMB, which means the cash register is clicking on the contract side while you are waiting for that clearance.

We can get you more information on that. We thought about options that could minimize it, for example, agencies like ours that have a very sophisticated evaluation office with our own intensive peer review, technical review, security policies, and technical guidelines, maybe there is a way that OMB could delegate some of the PRA functions to those agencies that already have the functions in-house.

The data sharing, I think that is an issue that the Commission should look at, For many of the data items that might be of interest, there is either a legal or regulatory restriction on re-releasing data, so that even if the researches might be able to get it, they can't release to another third party, which means it is harder for academics, for example, to analyze some of the same data without going again through, for example,

if you are trying to get LAHD data from Census, it could easily take over a year to get approval.

Depending on what data are being accessed, you may not be able to re-disclose that data to a third party that is a contractor that has a grant to do the evaluations.

The data access is a critical issue, and I think we all understand the importance of having the protections for privacy and security, but right now, it limits the ability to do comprehensive evaluations, and it limits the amount of time that we can have for doing it, and it also limits the access for academics to engage in collaborative research with the agencies.

MR. HASKINS: Other questions? Allison?

MS. ORRIS: Thank you. Thank you, Demetra, for your presentation today and really being a leader in this work. I just wanted to take a minute on the PRA as the OMB representative to acknowledge that we understand there are delays, but to also kind of say we have a statute that binds us, so to the extent that there is a collection that impacts 10 or more people, we take a look, but absolutely would be open to having conversations and hearing from all of you about ways we can improve the process.

Over the years, OMB has tried to apply the PRA as flexibly as we can, but don't want to minimize the challenges that all of you are facing, so just endorsing the idea of having some further conversation about that, and thinking about the way we can balance public transparency and the purposes for which the PRA was established.

MR. HASKINS: Robert Shea. On this point?

MR. SHEA: Yes.

MR. HASKINS: Okay, Robert Shea.

MR. SHEA: Demetra, thanks for being here. Can you talk about the

impact of your work on policy? To what extent is there interest in your work from the policymakers of the Department, Congress, OMB, and what changes have resulted from your work?

MS. NIGHTINGALE: There is increasing interest in evaluations because of the White House, OMB's, and Congress' focus on evidence. That is really good. Especially, I have to commend the guidance and the direction from OMB around budget submissions. The budget instructions now clearly indicate that if a department or an agency or a program's budget submission request is asking for additional funds, there must be an evidence justification for that, which that in itself has increased the use of and the value of evaluation and research and performance information throughout the Department.

We spent a lot of time, and I think we are getting better at it, we are continuously working on better ways to communicate the findings that come out of evaluations for non-researchers, non-evaluators, non-techy people to understand it, so drawing those implications, it is really an increasing emphasis on, for example, having practice briefs and program briefs in order to do that.

One thing, I often tell this little two second story. Ron, I won't take much time. When Secretary Perez first came to the Department, he was a huge supporter of evaluation. The second day that he was on the job, he said to me, so, what works in job training. I have been doing research on job training for 40 years. The next day I gave him a two-page memo based on the accumulation of knowledge summarized down to a level that can be useful for someone at his level.

I think job training is an area where we continue to need more research, but it's an area with huge interest. Other areas where research is contributing, for example, would be some positive results we have around registered apprenticeship,

which has led to a tripling of the White House targets, goals, and funding around registered apprenticeships. That is really good.

We have an accumulating body of research around effective strategies for improving outcomes for formerly incarcerated individuals that is leading toward revising and improving the programs there.

MS. HASKINS: Ken, can you ask a brief question, and Demetra, can you give a brief answer?

MR. TROSKE: I can try. Two questions. I was involved in a DOL study eight/nine years ago on the WIOA program. I was working with the contractor and we had a grant, thank you very much. A letter went out from DOL asking all 50 states to provide us their UI wage record data. We got UI wage record data from 11 in a timely fashion. Twelve tried to but couldn't pull it off.

You seem to suggest that has improved, and perhaps it has. I was kind of surprised that a letter from DOL telling the states to give us data for their project was so roundly ignored. I'd like maybe a little more detail on has it improved or if it hasn't, what we can do to improve it.

My second question is along the lines of Bob's question. How do we get evaluation studies done in a timely fashion so that it can actually affect program funding? For example, the experimental study of WIOA, which was supposed to be done, I think, in 2015, has still not been released, a year and a half after WIOA has already been re-funded.

Presumably, at this point, and in release of it, people can say oh, we completely changed the program, which of course, they didn't, but how do we get studies done so they can actually have an impact on the next funding round, because far too often it seems to me the valuation studies come out after Congress has taken action to

change the program funding.

MS. NIGHTINGALE: I'll give you highlights of that, but I'm happy to talk to you more on both of those. On accessing earnings data, we are doing better.

Right now, for example, and again it changes with different Administrations, eight years ago or nine years ago

-- not eight, nine years ago, it was very difficult to get access to NDNH, the National Directory of New Hires, which really is the best, most comprehensive source of earnings data, again, from our agencies at the state level, but we cannot easily access that at the national level.

HHS has approved requests from the Labor Department for over 12 evaluations where we are getting NDNH data, so that would be nationwide coverage, which is great. In that way, it is much better.

The WIOA evaluation is a good example of this where you may need a national sample according to the statutory requirements to evaluate that program. You need about 27 states in order to get a nationally representative sample, and our evaluations got all 27 to give the UI earnings data, but it cost over $1 million to get the data needed.

So, we are doing better. We also have a policy in place for non-WIOA -- WIOA has a special provision that does not allow for a national database, which makes it difficult to do linking because you don't have a unique identifier.

For other programs, like unemployment insurance, the Trade Adjustment Act, the Secretary has the authority to require states to provide that data.

MR. HASKINS: Thank you. We are now going to go to the next panel. Unfortunately, you lost half your time because the Commissioners asked too many questions. We will bear under the pressure.

Katharine O'Regan is Assistant Secretary for Policy Development and Research, U.S. Department of Housing and Urban Development. Evelyn Kappeler is Director of the Office of Adolescent Health, U.S. Department of Health and Human Services, who just released a spectacular evaluation. I hope you tell us something about that.

Matthew Klein, Executive Director of the New York City Center for Economic Opportunity.

Katharine, you are first.

MS. O'REGAN: Thank you. Good morning, and thank the Commissioners for this invitation.

Collecting, analyzing, and using data and research to inform HUD's programs and policies is the core function of the Office of Policy Development and Research at HUD, which I direct.

Like many Federal agencies that you are hearing from, HUD is focused on learning more from the data we collect, and more from the research we conduct. We believe that to make the most of the government data and fully leverage how the data can inform policy, we do 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.

Linking is particularly key to expanding cross domain work. Housing has big impacts on non-housing outcomes, so for HUD, this has been one of the areas we are very focused on, and also by ensuring capacity of governmental agencies to leverage the linked data, to conduct in-house research, which is part of my answer to your question about getting answers faster.

The Commission's charge is kind of the invitation to make that leap now.

I'm going to talk about some of our efforts to give examples of lessons that we think would be really useful for moving forward.

We have been focusing a lot on cross agency data linking, and I want to talk about two efforts here. The first is an effort that we actually had this year, which is to sign an MOU with the Department of Education that allows HUD to provide the Department of Ed with data on assisted tenants for matching to Federal student aid data on application and receipt of Federal student aid.

This now gives HUD the ability for the first time to have national data on one proxy, to measure college attendance for those we serve, the matched data has enabled us, for the first time, to do a large scale rapid cycle in-house experiment, because we have the matched data right in-house.

We are also able to provide aggregated outcome data to public housing agencies, which some are going to use to evaluate their programs. They don't have those data, we can do it in aggregated forms. It goes immediately into program evaluation.

Some lessons from this experiment, the linked data are not located in a secure research center. We get an aggregated batch back. We can process that and do some things with it. We can't delve in, go deeper, and go inside a center to ask different questions.

The infrastructure is lacking to make the most out of this effort.

Our second effort we have been working on for quite some time. With the National Center for Health Statistics, we have an MOU, and we have linked hard longitudinal data to 14 years of cross-sectional health data from two national surveys.

These linked data mean that researchers can look at the effect of housing on a collection of health utilization and health outcome data, and we have done

benchmarking. We have now re-upped and extended the MOU so that we have a three-year cycle going forward, and we already have research coming out of this.

The CDC Research Center provides the infrastructure was missing from our previous MOU that we don't have with the Department of Ed. They have the ability to do linking in-house, and they have the ability to secure privacy.

The MOU that we signed means the CDC will do that linking free. HUD geocodes all their data in our geocoding center free, so we actually have a structure that is financially feasible, and one thing to be thinking about as we move forward on data linking, and we are running into this on other efforts, is designing a sustainable financial model for data sharing. It needs to be built into the work.

The work that we did, we encountered amazing legal barriers due to the privacy concerns on both sides. It took many years to negotiate the MOU, and to be honest, the project only moved forward because we had some dedicated career staff who kept pushing it as their pet project and looking for somebody who would kind of lean in on it.

I would say that across both of these experiences, they feel a bit like one off time consuming interagency matches and that the most efficient way to move forward is going to be something more centralized that is easier to sustain.

That comes to my third example, which is something that we are doing just now, and is a pilot. HUD has entered into two agreements with Census that we think will greatly expand access to HUD data and the matching with non-HUD data, and for the approval of internal and external researchers to get access to the data.

The first one is with the Center for Administrative Records Research and Applications, CARRA. We have an IAA with CARRA that already links HUD's tenant data to other data that are inside the Census research centers, and select randomized trials of

our data are being put inside CARRA so that outside researchers will be able to do more with them.

In this IAA, the first dataset that we are sending over are the MTO data, so the MTO data that everyone is familiar with from the TEDDY work, so that researchers can write proposals and get access to it through a secured site.

The benefits here are large. I think most of you know, and I know the Commission will be hearing from Census o its role in supporting your work. This IAA that we have written is the first Federal pilot that's kind of proof of concept, how a Federal agency can do this. It is both on the data matching and on the evaluation, which I think is what you want to keep your eye on here.

To jump start this, HUD is creating its own RFP. We are going to create a small grants program to promote knowledge of this to researchers so that they will come in and use the data.

What I would say a lesson here is this is a pilot. It is a proof of concept in which we are ponying up money for doing this. There is an annual cost both to each dataset going over and then there is a marginal cost for researchers. Every dataset then would have an annual cost.

I have scoped out what this would mean for our trials. It would eat up my entire research budget within a short period of time. The pilot structure we are doing is a proof of concept, but the group needs to figure out a sustainable financial model, if you want Federal agencies to participate in this.

The second example I want to use with the MOU that we have with Census is a joint statistical project agreement. This part of what we are doing is to increase access to linked data for internal researchers. CARRA is great for getting external researchers. One of the things we want to make sure we do is we get the linked

data so that it is accessible to Federal researchers who know the programs really well and can move program and evaluation data immediately into policy.

With this IAA, we are committing to doing four projects with Census data and joint projects over the next year. It is a very important thing to be thinking about when we do this, to figure out how to make sure that Federal research staff get access to the linked data.

We actually think this is also going to be really helpful with recruiting and retaining some of our highest capacity staff and researchers. We are a great place to come to do cutting edge research on programs.

One of the things we have already learned from the matched data agreements that we have with CARRA is how it is playing into our research agenda. Right now, we are creating our next four-year plan.

We look at a collection of research questions and there are questions we probably would not have prioritized, big questions, that we couldn't see feasible or cost effective strategies for getting done that we are now putting in our CARRA bucket. Oh, we can see a way to do this with linked data across and within CARRA. We have a collection of things we are going to get to prioritize now that should be done that might not have made that cut.

I want to make sure I have time to talk about two ways in which we think we could make the most of this kind of data linking with CARRA.

The first is thinking about some restrictions that we have with the way the IRB guidance is currently done. The MTO data that we have been talking about was created at a time for 4,600 families where the consent was not time limited, and the consents for matching to types of administrative data were general, so you were able to take the data 10 years later, 20 years later, and match it to data.

The family option study, the second set of data that we are sending over, I think it is our new MTO in terms of what we are going to learn, in terms of long-term outcome, it is a randomized trial of homeless families with children. We have a five-year consent. They are going to be put over in CARRA, but if we want to learn about adult outcomes or later teen outcomes, we have to go find those families and do new consent forms.

If we want to make the most out of our data, we need to be doing something a bit different here, and one idea is that IRBs get some guidance from this Commission on consent forms where the data would then be put inside Census and be under Title 13, and think about guidance that we could then incorporate into the way we do IRBs with our randomized trials.

A second thing to talk about, we have already touched on, is thinking about the non-Federal data that are so crucial for getting evaluation work done. All of our demonstrations take place in locations. We rely heavily on state and local data.

For the MTO data, we ran into varied experiences on the ease of getting those earning data. Worked really well in some places. We still have one state we are trying to get data from, so we can look at outcomes.

There is a second pilot going on at CARRA, which has to do with Chapin Hall. They are submitting state and local data from a limited number of places. We could expand on that, and we should think about the data list beyond Federal data.

I'm going to leave this last slide up as the timer ends. Doubling down on PRA. For HUD, this is critical. Our programs are implemented by non-Federal staff. We can't learn about what is happening with our programs without going through PRA. We frequently don't bother. We use anecdotes because we don't have time to learn about something that is going on the ground. We can talk to fewer than 10.

We also think a session just on that would be useful. I will stop before it dings.

MR. HASKINS: Very wise decision. Thank you very much. (Laughter) Evelyn? Thank you for coming.

MS. KAPPELER: Good morning and thank you for inviting us and the opportunity to present on our efforts to strengthen capacity for implementing evidence-based programs, evaluating teenage pregnancy prevention programs with rigor, and using data to improve our work, all of which are key priorities for the Office of Adolescent Health.

At the Office of Adolescent Health, we administer the national teen pregnancy prevention program which is a tiered evidence-based grant program. Between 2010 and 2014, we funded 102 grants across the country, each at between $400,000 and $4 million a year for a five year grant period.

We awarded a total of $100 million in competitive funding to school districts, community-based organizations, universities, health departments, and other eligible organizations.

This funding supported cooperative agreements that included $75 million for replicating one or more evidence-based programs that were proven effective through rigorous evaluation to reduce teen pregnancy, behavioral risk factors underlying teen pregnancy, or other associated risk factors.

This, we called tier one.

It also included $25 million for research and demonstration grants to develop, replicate, refine, and test additional models and innovations. This, we referred to as tier two.

In addition, Congress provided program support funding, and also

earmarked specific funding to carry out evaluations of teen pregnancy prevention approaches.

To help identify evidence-based programs that were eligible for replication through this program, HHS conducted an independent and ongoing systematic review of the teen pregnancy prevention research literature. The evidence review examined research for study quality and assessed whether program models demonstrated positive impacts on sexual risk behavior and sexual health outcomes.

The results from our first cohort were just released this past summer, and we are now moving forward with a second five-year cohort of 84 grants across the country.

Our first cohort used a multi-pronged evaluation approach to address the questions of whether replicated evidence-based programs and new innovative strategies were effective. The approach included the collection of performance measures by all of our grantees, two Federal led evaluations, and a series of grantee led evaluations.

A uniform set of performance measures were collected and reported twice a year by all of our grantees which helped us assess the program's process in achieving goals and grantees were also encouraged to use the data to assess and improve their own program operations. The measures looked at things like reach, fidelity and quality, partnerships, training, and dissemination.

Two Federal led contracts supported individual evaluations of funded program sites. The first in tier one examined whether the three program models previously shown to be effective in a single study continued to demonstrate effectiveness when implemented with fidelity in different settings and with different populations.

A second Federal evaluation in tier two had two components, an in-depth implementation analysis of five selected programs to examine context and delivery, and a

rigorously designed impact study of each individual program.

Additionally, grantee led evaluations were required of all tier one grantees that were funded at $1 million or more per year, and all tier two grantees were required to rigorously evaluate a new or innovative approach. We required that all of the grantee evaluations be conducted by an independent third party evaluator, and that the evaluations be designed and implemented to meet the research quality standards set in the HHS teen pregnancy prevention evidence review.

We also found that there is a specific set of conditions that were necessary to ensure that grantee level evaluations maintain a high level of rigor. These include sharing detailed criteria for what is considered a rigorous evaluation, including a detailed description in the funding opportunity announcement about the expectations for conducting a rigorous evaluation, involving an independent evaluator, someone who is independent from the grantee organization, and independent from the developer that developed the original program model.

It is also important to allow for sufficient resources so that the evaluation can ensure power and the ability to maintain rigor. We required grantees to allocate between 20 and 25 percent of their overall budget for their evaluation.

We also placed a condition on their grants to stipulate that continued funding was contingent on meeting the standard sets for evaluation rigor, and we had an ongoing internal commitment to hold grantees accountable to those stated evaluation standards.

In addition, we provided intensive programmatic and evaluation technical assistance to grantees to both ensure high quality program implementation and rigorous evaluation.

At the office, we use the evaluation results and the lessons learned from

our first cohort to inform our funding and evaluation framework for the second cohort. In addition, it assisted our new grantees in selecting which evidence-based programs they would be implementing this next five years.

In cohort two, we are implementing a holistic approach to replicating evidence-based programs in order to have a much larger sustained impact. Using the first cohort's evaluation data and results, our current set of grantees have had more information on where, when, and with whom specific program models have been most effective.

This enabled them to better identify and implement the evidence-based programs that are a good fit and likely to have the greatest impact in their own communities.

We also recognize that not all communities are ready to implement evidence-based programs. We set aside some tier one funding specifically to provide capacity building assistance to those who are interested but not yet ready. We have eight grantees who are now providing tailored community building assistance to youth serving organizations in their communities to implement, evaluate, and sustain evidence-based teen pregnancy prevention programs.

We took a similar approach in our tier two funding. We continue to support grantees to rigorously evaluate new and innovative approaches, and specifically those designed to reduce disparities and fill gaps in our knowledge base.

We are also now funding two grants to foster and support early innovations to prevent teen pregnancy. These are promising programs or technology innovations that need more support for development before they are ready for rigorous evaluation, and the intention here is to bring good ideas to a point where they can be evaluated in a rigorous way.

In our new cohort, we continue to use a multi-pronged evaluation approach. All grantees continue to be required to collect and report annual performance measures, and grantees are required to conduct independent evaluations.

What is new in this round is that at the end of this grant period, all grantees will be expected to archive their data in a designated data archive that will be accessible to future researchers and may eventually be able to be linked with other data sources.

In addition, several new Federal evaluations are underway, and the second generation of evaluation activities addressing a more targeted set of research questions to fill the gaps in the current evidence space.

This includes a multi-site evaluation study to identify and test replications of commonly used evidence-based programs, to identify and test significant or meaningful adaptations to existing evidence-based approaches, and to identify and test selected core components, key activities, and implementation strategies of common programs.

Another study is looking at the feasibility and impact of scaling up evidence-based programs, and the mechanisms through which grantees aim to reach the highest risk populations. This project will also identify a smaller number of projects to implement rigorous evaluations, as well as design a rigorous cross grantee study to assess the effectiveness of this new holistic approach.

A third study is an implementation and impact evaluation of commonly implemented but under studied teen pregnancy prevention programs.

Finally, to extract the maximum value from previously funded evaluation efforts, we are also funding two additional studies, a secondary data analysis to enhance understanding of interventions designed to reduce teen pregnancy, and programs

implemented with middle school youth, and then a meta analysis synthesizing the evidence from across previously funded Federal grant evaluations.

This study will examine what program or contextual elements made a difference to youth in their outcomes based on reports of individual level data.

I think our experience brings to light several important issues regarding the capacity of community-based organizations to conduct rigorous evaluations, and many of these challenges and lessons learned may have applications for others.

We recognize the importance of providing adequate support in preparation for program implementation. Therefore, we required all of our grantees to spend the first 6 to 12 months of their grant engaging in a planning, piloting, and readiness period. This gave them time to identify and make necessary adjustments to staffing and program models before they went to full implementation.

We also know that communities need the capacity to implement evidence-based programs with fidelity and with high quality. We found it critical that grantees set up a system for monitoring, analyzing, and using fidelity data for continuous program improvements.

We also had some challenges in incorporating evaluation effectively once program implementation had begun. Recruiting sites was complicated by the fact that for most of them, their interventions were not designed to be implemented with the requirements of a rigorous evaluation.

We also found it difficult in some circumstances to ensure a strong contract between treatment and control, given that very large Federal grant programs span 39 states and the District of Columbia.

We are also learning a lot about building capacity and an established evidence review provides a standard yet rigorous set of expectations for future

evaluations if a program model aims to be an evidence-based program.

Clear communication about evaluation expectations along with training and technical assistance helped to build the capacity of our grantees to meet the research standards of the HHS teen pregnancy prevention evidence review.

We also found it critical for us and our grantees to collect data on the implementation of the program and to use that data on a regular basis for continuous program improvements.

We think it is vital to disseminate all of our evaluation results in a transparent manner. We know that no single program is going to work for everyone every time, and that mixed findings, which are to be expected, are critical for continued learning about which programs and approaches are most effective with different populations in different settings.

This July we released the results of 41 rigorous evaluations, and half of those evaluations have been featured in a supplement in the American Journal of Public Health, and all of them are now contained in reports held in a collection at the National Library of Medicine.

You can find more detailed information about our program implementation and our evaluation results at these links on our Web site.

Thank you for the opportunity to present.

MR. HASKINS: Thank you very much, well timed. Matthew Klein.

MR. KLEIN: Thanks. I appreciate the opportunity to be here. I direct the Center for Economic Opportunity in New York City. It is a unit of government inside the Mayor's Office, and we use data, evidence, and design to help the city reduce poverty. We assess the city's existing anti-poverty approaches, develop new interventions, facilitate data sharing across city agencies, and rigorously evaluate the impact of key

initiatives.

I'm pleased to offer a local government's perspective on the opportunity here, since obviously there are big implications for the Federal Government's work in this area, and in particular, the opportunity to make it easier for cities to integrate information from different sources so that we can deliver more effective, holistic, personalized services to our residents.

I have the privilege of participating as a local government fellow in the Results for America's program, and through the knowledge sharing that they offer, I trade notes regularly with counterparts in different cities, and the issue of data sharing and navigating data issues is top of the agenda for all of us. I really appreciate the work of this Commission.

My recommendations are going to focus on two areas. First, the need for the Federal Government to continue to model, champion, and support evidence-based approaches, which I know we will find a ready audience here, and second, the benefit of amending key Federal laws and regulations to allow local governments to share data across our own city agencies without consent for the discrete purposes of providing services, benefits, and care, for program planning and implementation and evaluation, and for using computer matching technology to facilitate these activities.

Cities are in a unique space when it comes to service delivery and engaging with residents. We are the closest level of government to programs on the ground, and importantly, we receive direct candid feedback from our residents.

We know that a key challenge is to engage with residents holistically rather than through the silos of particular departments who are charged with specific programmatic areas.

Our residents, and particularly our most vulnerable residents, engage

with the city in multiple ways, and we can be most effective when we are aware of an individual's or a family's touch points, and coordinate how our various programs work together to support them.

Cities are both creators and users of evidence. We are in a position both to build evidence that contributes to the national strategies, and also to draw on this kind of work.

In this context, I want to talk a little bit about some of the efforts we have made at the Center for Economic Opportunity where we focused on advancing the use of evidence and data in program and policy design, service delivery, and budget decisions.

For example, drawing on federally collected data but supplementing it with New York City specific sources, CEO established a local poverty measure, one that provides a more accurate description than the Federal standard rate. The CEO measure reflects the local cost of New York City, particularly of housing, while the more recent Federal supplemental poverty rate includes the value of forms of assistance, such as tax credits, supplement and nutrition program benefits and the like. This measure gives the city government a clearer picture of poverty in New York City and helps guide our anti-poverty efforts.

Our evaluation of those efforts focuses both on new programmatic innovations, and also on large city-wide efforts. As this Commission knows, evaluations shouldn't be seen as simply a formal effort to issue a thumb's up or a thumb's down verdict on whether or not a program works.

Different evaluation activities serve different purposes, and that has been the case for CEO. For example, in the course of rolling out New York City's new universal pre-K initiative, we employed qualitative studies to help us understand implementation issues and identify best practices that can be shared across sites.

With the support of the Federal social innovative fund, we are currently launching a new program and outcome study to assess the impact of integrating mental health and social services using a comparison group approach. We have also employed more rigorous impact studies using randomized designs and cost studies that help us understand the return on investment of an intervention.

I would like to cite a few of these that have been instrumental in shaping policy in New York City. The first is a model created with the City University of New York called the accelerated study and associates program. It aims to increase the college persistence of low income students seeking associate degrees.

A five-year experimental design led by our MDRC research partner found that ASAP students with developmental needs graduate at more than double the rates of non-ASAP students. As a result, New York City is growing a program from its initial test of 1,000 students to 25,000 students, and the model is now being replicated to other cities, and being tried at four year colleges.

Another example that we have also heard about already is in the area of workforce, and in particular, for CEO, our efforts around sector focused workforce development programs. We took note of research which found that sector based strategies benefit both workers and employers, and based on this finding, New York City launched a one stop sector focused career center at the time using a blend of city investment and WIA funds.

In 2013, a two-year study conducted by our research partner found that compared to standard centers, sector focused center participants are 12 percentage points more likely to be employed in the year after the program, earn 50 percent more on average than matched career center participants, and that industry specific training increased earnings by 82 percent.

These results were instrumental in the city's adoption of a career pathways workforce development strategy that is influencing the way that we are investing city-wide.

We have subsequently launched and tested a work advance program which has increased employment for low wage workers in target sectors and increased earnings by an average of 14 percent or nearly $2,000 annually.

Finally, another example showing the interplay between Federal and local evidence building is Jobs-Plus, which is something we have done with HUD. Jobs-Plus is a place-based employment and training program focused on supporting residents of public housing. It combines workforce, services, with resident led community-based support for work, and work incentives to help public housing residents obtain jobs or advance in careers.

This model was first implemented from 1998 to 2003 through a demonstration funded by HUD with an RTC evaluation by MDRC. It found Jobs-Plus increases public housing residents' average earnings by 16 percent when fully implemented, and these findings persist over time.

CEO piloted the first NYC Jobs-Plus site in 2009, expanded it again with support from the Federal social innovation fund in both New York City and San Antonio, and lessons from this implementation and evaluation are now informing HUD's more recent iteration of the federally funded Jobs-Plus program.

I also want to talk a little bit about the intersection between evaluation and administrative data that is used on a more real time basis. Since this group knows getting good results involves more than implementing a model that was shown to work at one time, we need ongoing real time feedback from data, and administrative data can be crucial on this front especially when we have it readily accessible.

Among the major steps that we have been taking, we are launching what we are calling a city-wide data integration initiative, which is making data accessible with the appropriate permissions and protections through a platform that draws from our agencies' own administrative data, so very much along the lines of what you have heard recommended on the Federal front.

An overarching purpose of this initiative is to facilitate a one city approach to program planning, development, and service delivery. The ability to capture data longitudinally as you have heard, linked across sources, with clear documentation is critical for this to work.

I want to particularly point out for those thinking about how to organize this at the Federal level, that at the city level, the Mayor's Office of Operations has proven critical in leading the effort, notably because of its historic role as a performance arm of the city. Locating a data integration centrally with the body that is charged with performance management for the city, I think, has been critical in helping us advance the initiative.

Also, of course, the technology platforms have to be secure. We need to comply with regulations of HIPAA and others that demand confidentiality and privacy, so working with all of New York City's technology partners. That is something we have been focused on and committed to doing.

We have other tools in place that allow case workers to see integrated case files of clients drawn from a number of different sources. It is different than a longitudinal warehouse of data, but very useful for on the ground implementation.

Finally, under our econometrics initiative, we have launched/engaged on something that I know you have heard about also, which is to define outcomes similarly across different agencies. That allows us to make better relative comparisons,

comparing apples to apples, when looking at things like workforce development where things like job and placement can mean different things depending on the funding stream.

Finally, let me quickly skip to a recommendation beyond just championing the furtherance of this evidence-based approach and focusing on econometrics, to say that Federal laws should be amended to give local governments more freedom to share information among its own divisions and employees.

Many of the laws that were designed to protect individuals from having their privacy compromised, and that is appropriate, but Federal laws and regulations should draw a distinction between a privacy invasion and use of data by a city internally across its departments without client consent for, like I said, the purposes of providing services, benefits, and care, for program planning and evaluation, and for using computer matching technology to facilitate these activities.

Among the most valuable changes would be amending regulations under HIPAA to create new legal exceptions to authorize city agencies to share information with each other when providing or coordinating benefits, services, and care.

This would give New York City and other cities more flexibility and clearer legal footing to meet specific needs of our vulnerable populations.

For example, under existing regulations, it's not clear if our city's hospital system can legally share identifying patient information with the Mayor's Office of Criminal Justice, a non-HIPAA covered entity, for a data match that we want to conduct to identify homeless individuals with a history of incarceration who are in need of supportive housing placement and could be placed in limited supportive housing slots.

Similar clarity in exceptions in FERPA would allow us to do things like better target youngsters for inclusion in services that could benefit them, and finally in a wide array of areas, allowing us to conduct computer based matchings without labeling

that match as a disclosure would facilitate faster, better use of administrative data by implementing local agencies.

With that, I will stop, and again, appreciate the opportunity to be here.

MR. HASKINS: Thank you very much. Commissioners, questions? Bob, and then Katharine.

MR. SHEA: I am going to follow up on Matthew's comments, a couple of things about the data integration. I guess three things, sorry. One, you have heard your prior presenters talk about difficulties of cross agency work. I would like for you to reflect on that a bit.

Secondly, I'm a little confused in your uses whether you are making a distinction between statistical uses of data and program uses of data, and by that I mean the first being completely uninterested in an individual record, and the second being intensely interested only in individual records.

Could you comment on that?

MR. KLEIN: Sure. I'd echo the challenges that I think I've heard expressed about doing cross agency data integration. The challenges come in three forms. One, just programmatic, what questions are the most important questions to ask. Second, technical. The technology, as everyone knows, is advancing rapidly. Data is exploding. Many of our systems are not keeping up with modern advances.

So, being able to take advantage of new data sharing technologies in conjunction with legacy data systems is extremely complex. Part, I think, of the solution going forward is to think about on the technical side what the appropriate requirements should be for new build's in governments securing of technology. I know there is a lot of work being done on that front.

The last thing is privacy and legal challenges, which add to the

complexity of the other two, because in order to integrate data from multiple sources consistent with privacy laws, which are sometimes explicit and sometimes ambiguous, it is both a process to negotiate the legal interpretation of those privacy restrictions, and then to build the technology that can accommodate what you can and can't do and who can see it and who can't.

To that point, I'm speaking mainly about the individually PII data, the individually identifiable information, the personally identifiable information, that really can help a city do sort of the ideal, connecting the right program to the right person at the right time. To do that, we want to know, again, consistent with protections and responsibly, how our residents are engaging with the city, and how we can better coordinate those agencies that may not know they are working with the same family or to send an alert from one agency to another when someone enters a particular system.

There is just a lot of innovation work we can do around the delivery of services if we could draw on the totality of the data, and of course, measure it and see what works and so forth.

MR. HASKINS: Katharine?

MS. ABRAHAM: I had a question to follow up on some of what Katharine was talking about. If I understand correctly what you were describing with one of your examples, HUD is now planning to send some of its data over to CARRA, where they would reside and be generally available for research and evaluation uses.

One of the reasons why assembling data from different sources is so difficult is because the places where the -- the organizations where the data are originated are unwilling to give up control over who is going to get to see them and how they are going to be used.

If I understood what you said, you have sort of jumped over that hurdle. I

am wondering if you could talk a little bit more about that.

MS. O'REGAN: Sure. One way to jump over the hurdle is we have not totally given up control. Our data, and the MTO data being the first, we have made the MTO data widely available to researchers but have not had an infrastructure for doing that, so it has been piecemeal. By having those data over at CARRA, researchers can then apply with projects, but we would be able to approve or disapprove a project.

I can tell you it is a lot easier to get this signed and done where there is a process where it wouldn't just be at least at the starting -- we have something that is called data licenses, which is we make our data available to researchers through a process where they send us their information on the data they would like.

You will see lots of studies where micro data on voucher households have been matched to other data. They don't have PII, but they have an address, they do something. We do it under a data license. This is something we could think about doing through the infrastructure at CARRA.

They do apply and have to tell us what they are going to do. A main part of what we focus on is that they are researchers, so the privacy component. I don't expect that the fact that the process of going through us will lead to research projects by researchers being denied access to the data.

We are putting some money on the table to cover the cost of access to CARRA for some subset of them, but we are expecting we will get a bunch of projects that are all approved, three of whom we can fund and the rest of you can go get some foundation or just pay for it yourself, but click, you have access to the data.

MS. ABRAHAM: Just a follow up, could you imagine a scenario where you would lay out a set of criteria for what projects were approvable and delegate that to CARRA, or do you think that is just beyond what you could think about at this point?

MS. O'REGAN: I actually think it might be faster for us to get it done. Nothing against the Census, but we have been doing this kind of piecemeal with data licenses, we take the work off our plate for the infrastructure and the data pieces, and what CARRA does is all of the special sworn status in addition, so anybody that is going to be accessing the Federal statistical data goes through that process. They work on that part. We just look at the proposals to make sure this is actually a research proposal for which the data are appropriate.

They do not have to go through the type of -- if the key data they are getting access to are our data, they do not have to do the type of Census proposal and go through the review committees for Census. We can just sign off very quickly on the proposal. They do the special sworn status and the researchers gets in.

Given the experience some researchers have had with having to go through the special sworn status part through Census, we believe we are going to be able to portray this as a faster process by the way we are handling it. It is a pilot. We could imagine very much going in a different way after we learn from this round.

MR. HASKINS: Bruce, and then Paul.

MR. MEYER: I'll be very brief. I just want to thank Katharine for submitting the six-page supplement to the slides that answered my questions about the PRA.

MS. O'REGAN: Yes, the testimony is longer than the slides.

MR. HASKINS: Paul?

MR. OHM: My question will be much longer than Bruce's. I am a privacy law scholar. This was very new and informative to me, so thank you all. The one question I have been having for a few meetings now is -- I am also a computer programmer and computer scientist.

It seems like a lot of the work you are talking about, and please correct me if I'm wrong, is ordinary kind of old-fashioned multi-variant regression and other econometric tools.

I haven't really heard a witness yet talk much about around the corner machine learning techniques. One question is that something that is beginning to become part of the evaluation landscape.

Number two, for our purposes, does that raise new and different data management and data access questions that this Commission ought to be considering, or should be instead just focus on rather orderly tables that get linked in rather traditional ways as opposed to kind of I need all the data and I need it now sort of mindsets that machine learning tends to engender?

MS. O'REGAN: I can speak very quickly, and then I will be at the outer bounds of my knowledge on it and will pass it on to Matt. Not within the data matching component of what we are talking about here with our own administrative data inside HUD, which comes from arcane systems with different variables and a whole collection of stuff.

We are moving in that direction, it sounds like a data lake, where we would be able to be doing machine learning, but we are at the forefront of doing that, and working now on designing that. That would be just for our data systems at this point, which are massive enough that it would be very helpful for our program work.

MR. KLEIN: I'd echo and I'm at the same boundary, but I'll give you an example. I think we are at just the beginning of the machine learning, and predictive analytics and all things that we care a lot about, and are trying to build the capacity which requires some of these more modern data techniques.

To give you an example of one problem that we are engaged potentially

with outside researchers to think about, using wage records plus data that we have about folks' work histories to help newly unemployed make decisions about whether or not to take offers as they get them, by learning from the experience of the tens and hundreds and thousands and millions of folks who have come before with similar profiles, and how long it has taken them to find work. That could help inform the kind of decision-making that someone might want to make.

There are enormous implications, of course, to all the algorithm based decision making, but I think there is broad excitement about utilizing it. It does speak, I think, to the implications of allowing things like matches and updating privacy laws that were created before that, the machine learning kind of techniques, which you would know much more about.

MR. HASKINS: (Inaudible)

COMMISSIONER: Thank you very much for those presentations. They were very helpful. Just as a follow up to what Paul mentioned, I see machine learning as just another tool. It's the same data we would all want to leverage with machine learning technique rather than a different kind of evaluation technique.

I wanted to follow up on one thing that Katharine talked about near the end of your presentation, mentioning something about differences in IRBs that were the MTO IRB that had an open-ended consent that allowed for this long term evaluation without having to go back to the population, but then you talked about a new project you are working on with a five-year consent.

Can you talk a little bit about that, and what is guiding those different outcomes?

MS. O'REGAN: For the evaluations that we do as we work with outside organizations, everybody has an IRB Board, and the recommendations and constraints

that come in are set by their IRB policies.

There has been a movement over time for privacy and a variety of other reasons for the IRB Boards to have more restrictive -- place greater restrictions on consent forms.

Where we are now, the randomized trial that we are doing with the homeless families, the restrictions are a five-year restriction, and that is common across our current randomized trials. It is different than not just MTO, but we have gone back and looked at some of our older randomized trials, we would be able to do something like MTO, we would be able to take them in and look forward beyond for quite some time.

It would be good to hear from other Federal agencies, and maybe we will get it on the list, the five things list. What we have from this Commission was a discussion of the benefits for evaluation of looking at long term outcomes, that we have now learned this, and if the Title 3 Census structure and protections are going to be over the data for that purpose, would it be possible to extend the time period or the breadth of administrative data under that rubric.

We would be able to incorporate that kind of language into future IRB consent forms, working with our outside evaluators. That is what I had in mind.

MR. HASKINS: Yes?

COMMISSIONER: It will be quick. Thank you, Katharine, Evelyn, and Matthew for your presentations. I was going to ask Demetra but we ran out of time. In your evaluations, do you also do a cost-benefit analysis, and do you have a formalized process? That is part of this Commission, to look at cost-benefit analysis as well.

MS. O'REGAN: I can speak on behalf of us. Earlier, Matt, I think you were talking about the fact that different types of evaluations have different roles, so it wouldn't be true for every evaluation you would want to do a cost-benefit analysis

because that is not the type.

As an example, the family option study, looking at outcomes, an important piece of that work was it also looked at the different costs of different types of interventions. It is not at the level of a formal cost-benefit since it is early stage. We are at the point now where have three year follow up data, where we have information about the housing stability and some decreases in adult stresses and mobility across schools for kids, we have some good information about behavioral problems with kids.

You could do a back of the envelope cost-benefit from the costs that we have seen, but you would have to play out those benefits, which are different than the housing benefits that we have looked at, but that is a key piece of the impact studies that we are doing, to make sure we get the costs to be able to do that type of assessment.

MS. KAPPELER: In the teen pregnancy prevention program, we have started to look at cost issues. We have seen in the implementation of evidence-based programs great variation in grantee costs for implementation.

We have done some internal work to look at why does it cost one grantee a certain amount to implement the same program versus another grantee. We haven't gone as far as cost-benefit analysis, but it is part of our early work in terms of understanding the costs of the actual program implementation.

MR. HASKINS: Matthew, I think we have time for about a two-word answer here. It's a very broad question. You talked about, and several other people I know have done the same thing, I'm not sure before the Commission, about an ability at least within the agency to share personal data without any sort of permission form would greatly facilitate things, which I am sure everybody would agree with.

Do you have any experience with the politics of that? Is it public, that you want to do this? We had a witness that came to our open hearing who said he didn't

think we should even have individual data, education researchers had a near death experience with FERPA a couple of years ago. This is really concerning that we would broaden access without getting permission from individuals.

Have people complained about that, have any City Council had a fit, anything like that?

MR. KLEIN: Yes. I don't know if the City Council has had a fit. I know the question of student data privacy is significant in New York and New York City and elsewhere, and in particular, that can be tied to the fact that educational technology products are private companies that are utilizing public data.

I think the purposes for which you would share data without consent need to be clearly articulated, but there are exceptions to FERPA and HIPAA that allow one to do that. In particular, when you think about outreach campaigns, this person is qualifying for this program, but we don't think they are getting that program, let's let them know about it, so we do campaigns including calling folks about that. I would tell you that the folks who are on the other end of the call are grateful to be told about the opportunity.

I think there are legitimate concerns, and I think we will grapple with them, but the purpose and being explicit that the purpose is allowed is a real advance that I think this Commission could produce.

MR. HASKINS: With that, let's take a break. We are on a break for 10 minutes. We will start again promptly at 10:10. Thank you. Thanks to our witnesses, great testimony and answers to questions.

(Recess)

MS. ABRAHAM: Okay. I'd like to move on and welcome our next panel. We have four members of this panel: Tanya Beer from the Center for Evaluation Innovation; Jim Sullivan, from the Lab for Economic Opportunities at Notre Dame; Adam

Gamoran from the William T. Grant Foundation; and Kelly Fitzsimmons from the Edna McConnell Clark Foundation. And I would start with you, Tanya.

MS. BEER: Excellent. Thank you. Hi, everybody. I'm Tanya Beer, from the Center for Evaluation Innovation.

MR. HASKINS: Could you bring your mic closer?

MS. BEER: Oh, yes, closer. Thanks for having me. The Center is a small non-profit that focuses on building the field of evaluation in areas that are difficult to measure. But we also run a network of foundation, Directors of Evaluation, among the 40 or 50 largest foundations in the country, who get together regularly to struggle with the sort of art and science of evaluation and evaluation use in the philanthropic sector. So, I'll be speaking from that perspective today.

And I think that my task is primarily to offer kind of a big-picture context about trends in the social sector around evaluation, and to sort of set the stage or the other speakers this morning. So, just to make sure we are all, kind of, on the same page about -- I'm not sure where to point -- There we go.

All right, so just to set the stage, just so we are all on the same page about this sort of size and scope of resources that we are talking about, there are about 86,000 foundations in the U.S. with $715-rougly billion in assets that give away about $52 billion a year. If you zoom on the thousand largest of those, they account for about half of that giving, and about two-thirds of that go to these sort of top three categories here, health, education and human services.

That dollar amount that we recognize is pretty paltry, relative to public sector investments, but it does represent some capital that can be used for, theoretically, innovative approaches, developing innovative approaches -- Relative, right, to the public scope of resources. So, it lets us though innovate in program delivery and systems

change efforts, but also, it leaves a little more space for innovation and evaluation, and evaluation approaches and use as well.

So, I would like to talk -- we'll go to our slides, and I'm sorry to say this is one of those animated things, so you can just go ahead and get them all on there, and we'll skip to the end. There you go.

I just want to talk about kind of five core trends that we are seeing in philanthropy over the last 15 years or so. We've seen a very distinct evolution in the model of giving from a kind of traditional, charitable giving approach that was just finding and writing checks for good causes to a very deep attention to effectiveness and outcomes that started about a decade-and-a-half ago, to really kind of catch hold among more than just the few biggest foundations.

The implications of that have been that we've also seen a really big increase in the interest in and demand for evaluation that is manifest in the form of increased investment and in-house evaluation staff at foundations, the kind of cultivation of a whole field of evaluators who primarily work in the philanthropic and non-profit sector, rather than for public sector evaluation.

And also what it looks like is a real change in the breadth of purposes that evaluation is used in the philanthropic sector. People are increasingly interested in not just retrospective conclusions about the effectiveness of individual programs, but also increasingly, how do you get real-time data and information to information to improve decision-making along the way.

The second trend we've seen also, not unlike the public sector, is a real increase in the focus of complex problem-solving, and sort of systems-level change, and the implications of that have further the evaluation universe, that means that we've really got a growing demand for systems-level data. So, just as a few examples, more and

more foundations are engaging, and in fact the large foundations, are engaging in investing on the local level in efforts to do things like improve early childhood systems coordination, or allow for health care systems to experiment with patients and patient-centered care models that bring together oral, physical and mental health.

What that means in the philanthropic sector is that they have sort of bumped up their attention and evaluation from the individual grant and non-profit level to a systems level, but they are continually sort of meeting up against the wall of access to data at that systems level, which usually lives in the public domain. So all of the same barriers we are talking about before to accessing that data are exponentially more difficult in the social sector where there are far fewer resources and kind of negotiating power to get access to that data, and to client-level data. So, there's a little bit of a black box of effectiveness that I think a lot of philanthropic organizations are facing when it comes to the kinds of systems change work that they are supporting.

The third trend of course that should be familiar to everyone is that we are seeing a real interest in scaling these smaller scale pilot efforts in partnership with public sector; the public sector and sometimes the private sector. So, there is both an opportunity here for much more cross-sector collaboration on evaluation, but we actually don't see it very often, and I have a huge amount of empathy for the non-profit organizations who serve public sector grantee, or funders, and several non-profit philanthropic funders, all of whom have different evaluation demands. And these are non-profits who often have very limited budgets for evaluation, but the lack of coordination between public and philanthropic actors in this space has really created a strain, I think, in the non-profit sector.

The fourth one, which is maybe more my opinion, just observing how the field has evolved rather than a database observation, is that we've also seen an

explosion in what I think is actually a disproportionate attention to performance metrics which serves a really important accountability function in the sector, but what it does actually, is distract from our -- sometimes -- distract from our ability to understand cause and evidence. So, our ability to build evidence performance metrics don't actually reveal much about causation, nor do they help us build a body of evidence.

But we've seen a huge take-up among philanthropic organizations who are looking for performance dashboards, and sort of rollup of indicators that suck up all of the evaluation capacity of non-profits to just produce those numbers, and leave very little space, except among the biggest non-profits, for actually generating data and insights from true evaluative work.

And then the last one is that we actually have witnessed remarkably limited capacity to use evaluation to make better decisions. Over the last decade the philanthropic sector has started to pare its evaluation function with an explicit learning function, right; where they have learning agendas which an earlier speaker spoke about today. And what we are finding is that despite the increased attention to data and outcomes and effectiveness, and the availability of data and evaluations that say something about effectiveness, it turns out it's really difficult to use that data in decision-making.

Most of the strategies for ameliorating that problem have focused on the evaluation side of the equation. How do we get data in a more timely manner, that matches with those decision-making opportunities, for example, but what we have now come to realize in philanthropy, I think, is that there also has to be a commensurate attention to capacity and experience on the side of decision-makers to be able to actually make sense of that data, and understand its implications for how to move forward.

And that is also that is not just about kind of capacity to absorb data, but

it's also about how the work itself is structured. How the timing is set up, how staff members actually incorporate or expected in their own performance assessments to incorporate data and evaluation into their own decision-making. So, as a result, we are seeing a lot of the philanthropic sector begin to experiment with actually, what kind of capacity do you need on the strategic side, to be able to incorporate data and evidence more effectively into the decision-making.

Just a few more points before I hand over the floor to other folks to get into some specific examples. We still see a really big capacity challenge with evaluation in the social sector, despite all of the kind of rhetoric about it, I think. About 50 percent of the largest funders, only the largest funders, have dedicated evaluation staff in in-house capacity, and 8 percent of non-profits have dedicated evaluation staff. And we suspect that number itself is inflated because people confuse the difference between performance metrics and evaluation.

So, we have just a sheer skill set and bandwidth problem in the sector to be able to really produce and use evidence well. But we also have a budgetary problem of course, not only do the vast proportion of non-profits use tiny portions of their own budgets understandably, when they are, sort of, often just cobbling together their regular operating budgets to do their evaluation work. We also found that fewer -- most foundations fund individual grantees at the rate of less than 10 percent to do evaluations of their own work.

So, while there is a lot of thinking, and experimentation going on in philanthropy that I think could lend some lessons, to evidence-based decision-making, there are a lot of opportunities and risks, too, that we've experienced where we actually think that the better collaboration between the public sector and philanthropic sector evaluation efforts could really speed up knowledge generation considerably in the field.

So that one of the things we wanted to ask the Commission to consider, is a sort of two-fold challenges, one, how is it that we can do better coordination between public sector and philanthropic sector evaluation efforts, and to please consider deeply how it is you build the capacity for evaluation use among strategists and program deliverers in addition to addressing the challenges on the data and evaluation side. Thank you.

MR. HASKINS: Before you go on. Ten percent of the projects that are funded by foundations get any evaluation?

MS. BEER: Get any evaluation dollars from foundations. Ten percent of foundations provide dollars for grantees to evaluate --

MR. HASKINS: Oh, 10 percent of foundations?

MS. BEER: Yeah. So, what we find and that means is that most non-profits have to pull money to do any data collection work from whatever their general reserve funds are which very few foundations actually support. And so there's a big gap in capacity on that front.

MS. ABRAHAM: Great. Jim?

MR. SULLIVAN: Unfortunately, I also have animated slides, so we'll see if we can work with this. So, thank you, Katharine and Ron, and Commissioners for inviting me to talk with you all today. And in preparing my comments for this, I thought about the -- I wanted to think about what the motivation for this Commission was in the first place, and a good source for that motivation was, one of the sponsors of the legislation that created the Commission, Speaker Paul Ryan; he says right now Washington measures success by how much it spends, not by how much it helps people. The Commission will help change government's old way of thinking and make better use of data we get so that we can make more of a difference in people's lives.

And oftentimes when Paul Ryan talks about this, he cites statistics like we spend a lot of money on social programs, and we don't have a lot of evidence on that. If we can advance the slides; I have some statistics on that. So, we spend about 800 billion a year on social programs in the federal government, and yet there's very little evidence to back this, and one statistics that's out there, in terms of fraction of this that's backed by evidence is it's about 1 percent of all programs that are backed by evidence.

And so the main point I want to make to you all today, is that if you wait to till the government is spending $800 billion to evaluate it, it's really too late. If we want to design effective programs that are backed by evidence in Washington, we need to know what's working outside of Washington, and so that brings us to the human services sector.

It takes us to the human services sector, and when I'm talking the human services sector, I want you to be thinking about local job training programs, youth development programs, your local homeless shelter, all sorts of essential services provided to the community. When you add all of those up, it comes to a not so paltry $200 billion, and those resources have a tremendous potential to inform evidence-based programs and policies at the federal level.

So the model that I have in mind is the following. You design a program to address local needs, and that program is launched at the local level, and along with that program launch you build in impact evaluation to establish the evidence of the difference that that program is making, and that evidence is really the springboard for scaling the program up to a national level so that the program can have broad, national impact. And there are a number of examples of this, as to actually how it works in practice, one example that is commonly mentioned is the Nurse Family Partnership, started as a small program in Central New York, and today serves 32,000 families in 42

states.

I do want to point out that it doesn't have to be a program that's replicated. Evidence at the local level oftentimes informs policy at the federal level. Perry Preschool played a big role in terms of informing early childhood education. The Nurse Family Partnership, even when it was replicated, it informed the investment of millions of dollars of federal funds in terms of home-visiting programs.

What I do want to point out is that this tends to be more of the exception, this process. A better way to perhaps describe the typical scenario of how this works; is you have an innovative local leader who identifies a need at the local level, and then that innovative leader designs a program to address that need and launches that program. Now there's not evidence that's built into evaluating that program, but this dynamic, innovative local leader can also raise the resources to scale up that program, so that program is scaled up.

There are lots of anecdotal stories that explain how this program is successful, and that it's scaled up at a national level, and we have unknown impact on what that program is doing. And the big and important problem here is that once a program is scaled up, it's much harder to use evidence to reallocate resources towards other programs if that program is not working. It's much easier to prevent a program from being scaled up based on evidence, and to reallocate resources after it's scaled up.

There are a number of examples, I'm not going to pick on particular programs here, but I do want to kind of give you a sense of where I'm getting the sense of that typical scenario. And so at Notre Dame we launched the Wilson Sheehan Lab for Economic Opportunities, which is a research center where we identify innovative, effective and scalable programs that move people out of poverty. We work directly with local service providers to build in impact evaluations, so that they can establish the body

of evidence that their programs work, and use that evidence to inform other providers all across the country, and policymakers at all levels of government.

We have established a portfolio of projects all across the country, and I'm not going to go through all of them, Ron wouldn't let me, but what I will do is highlight one program as an example of kind of how this works. And so, if you are on the brink of homelessness, in the City of Chicago, you call 311, and 311 routes you to the Homeless Prevention Center, the call center screens you for eligibility, and then refers you for financial assistance, to pay your one month's rent so you don't become homeless.

Unfortunately, some days funds are available, and some days funds are not available. And so what we did to evaluate the impact of this program, is link those data to homeless shelter data, and see, are you less likely to check into a homeless shelter if you call on a day when funds are available, as compared to a day when it's not? And we show that you are significantly less likely to show up in the homeless shelter data if you called on a day when funding was available.

And I should point out, that the availability of funds is unpredictable, sporadic and functionally random. And the point I want to make with this study is that we would never have been able to study this if we weren't able to link the call center program data to administrative from HMIS, on homeless shelter entry. In fact, a point we make with this study in motivated, is that there's far too little research looking at the impact of programs on homelessness because it's simply very difficult to get information on homelessness.

The Family Option Study is another example. They also use HMIS data, but the access to administrative data made this study possible, we would not have been able to get access to the administrative data if it wasn't for the Homeless Prevention Call Center folks, who referred us to the program that houses the HMIS data so we were able

to make that link. Unfortunately, this is more the exception than the norm.

In our work with local providers, I want to highlight why it isn't the norm. So, what are the obstacles that these local providers are facing that prevents them from doing this sort of impact evaluation? In the first it is a very practical one on the point of funding, and this is something that Tanya has already highlighted. So these budgets for these local providers are very stretched, the majority of those budgets are targeted towards the services to serve clients, they actually collect information on what fraction of the budgets go towards their clients. It's very challenging for these providers to raise resources that are targeted explicitly for impact evaluation, so it's difficult for them to roll those kinds of evaluations out.

Even if you have the funding, it's oftentimes very difficult to get the data necessary to do the kind of rigorous impact evaluation that we would need. Service providers are very good at collecting information on outcomes for clients they are currently serving, it's more challenging to get information on outcomes for clients that they are no longer serving, it's even more challenging to get information on the clients that they never served. The comparison group we need to do these kinds of impact evaluations, all of that information is available in administrative data.

The other point that I want t make, that it's challenging for service providers to implement evidence-based programs for a number of reasons; one is that there's a lot of information out there, but it's challenging to synthesize all that information what really are best practices. And then once you have the report on an effective program, how is the service provider going to take that research report and translate it into implementing something at the local level? Replication is very challenging.

As a result of these challenges there are a number of programs that are launched, and persist that lack evidence, and I'm just going to highlight one to give you a

flavor for what I'm talking about; and this is a national program, a refugee resettlement program. The Office of Refugee Resettlement in the U.S. resettles tens of thousands of refugees each year, they did 70,000 in 2014, and the way this works is they contract with local service providers, to help refugees transition to the local communities so that they can be successful.

This is a program where the costs exceed 215 million annually. And the key problem here is that we have virtually no information about what are the economic outcomes for refugees more than six months beyond the time that they are resettled. We don't know their earnings, we don't know employment rates, we don't know their usage of government benefits, and all of that information is publicly available. We've worked with a number of these service providers to try and get access to that administrative data to be able to answer questions about the effectiveness of the programs. Should we be putting refugees in English Language classes? Should we be giving them job training? We don't know the answer to the effective of those programs because we can't access the administrative data.

So let me just conclude by talking about what I think the Commission could do to help on these margins. And this is going to resonate with a lot of what you've heard already, and in terms of what would help cities, but also what would help the foundation sector as well. So, in each of these, and I'm going to go back to the three points of the obstacles, because I think there's ways to address these obstacles.

The first of the obstacles that I mentioned was that of funding, and that part of the problem is that the funding doesn't incentivize the providers to allocate resources towards impact evaluations, and so, you know, obviously more funding would be good, but there needs to be tied to this additional funding explicit incentives for the funding to be used for impact evaluation. There's lots of good examples of this, i3 Social

Innovation Fund, et cetera, but the problem is that's not the norm, and I think it needs to become more of the norm if we want to see more impact evaluations at the local level.

There are also opportunities to do this through legislation, the 2014 Farm Bill had Section 22 had a provision for pilot experiments to evaluate changes in the Food Stamp Program. Included in that provision was the requirement that those pilots were evaluated in a rigorous way, and also require that the states provide administrative data. And so those kinds of incentives can go long ways in terms of incentivizing providers and giving them the resources they need to do impact evaluations.

The obvious request is to make data available. I want to kind of go a step further and say we need to make data accessible, so that providers can use it, and there, what I'm talking about, is what kind of personal identifying information should all providers collect so that when they go to some clearinghouse that makes data available, they know they can link the data appropriately.

What kind of consents should the providers be collecting initially so that they can get the approval to link the data? And so the last comment that I'll make in terms of recommendation is to help with this facilitation process. I know you are thinking, no, no, not another What Works Clearinghouse. Really what I'm thinking are two things. One is to have better synthesizing of that information, and the other is to help them use that information to replicate, not just, here is some research, but how do I use this research to actually replicate and implement the program. Thank you.

MS. ABRAHAM: Great. Thank you very much. Adam?

MR. GAMORAN: On the list of recommendations, because my list is longer than Jim's. So that's for Commissioners and Staff, it's also the last page of my written statement which will be accessible to everyone. Thank you, for the opportunity to speak today about one of our nation's most vexing challenges: serving the public good by

making smart policy decisions using data our government already collects.

My aim is to provide a nongovernmental perspective on two key issues facing the Commission. First, what data are needed to support public good activities, and second, what steps to build the capacity to use data and create the evidence required for better policy decisions. The William T. Grant Foundation, which I lead, support social science research to improve the lives of young people, ages 5 to 25 in the United States.

Within this overall mission, we have two areas of focus. The first is to support research on reducing inequality among young people, and the second is to support research on improving the use of research evidence in policy and practice. These priorities would be greatly aided if the Commission were able to accomplish its goals. This Commission is uniquely positioned to consider not only how to achieve linkages among administrative datasets, but how the federal government can use data to create the evidence required to achieve our policy aims, as well as how to create the infrastructure to support the use of evidence in policymaking.

A close reading of your charge fully supports this broad and essential mandate. So, if I have one message for you today, it is this. Be the evidence-commission, not just the data commission. Now here are some thoughts about how to make that happen. To inform policymaking data must be deployed and researched in evaluation to create research evidence. By research evidence I mean evidence derived from applying systematic methods and analyses to address a predefined question or hypothesis.

Policy insights can emerge from a variety of types of research studies, and consequently liked to administrative data has multiple purposes. Among these rigorous evaluations of program impact, both experimental and quasi-experimental program improvement efforts, including performance management, and descriptive

studies that contribute to policy formation. Now the Commissioners already heard much today and in past meetings about rigorous evaluations, the program impact, so I'll mention that topic only briefly.

Suffice to say that administrative data can be a powerful tool when attached to evaluations the program impact both for long-term and short-cycle studies, and I've provided more details in my written statement including examples from research my Foundation is funding.

Experimental and quasi-experimental studies often focus on the overall impact of programs, but their findings have limited value for policy if they do not say what works for whom and under what circumstances. This is because evaluations are most useful when they point the way towards improvement. That, however, requires information on context, and on implementation, including conditions that facilitate or impede success. Often, such contextual information is available in administrative records.

The Commission could provide guidelines for federal grant makers that would strengthen researchers' access to contextual and implementation data. Performance management is a tool used for program improvement, or can be used for program improvement that relies substantially on administrative data. To assess outcomes of the government program, however, a federal agency typically requires administrative data that are gathered elsewhere within the government.

While current federal law requires agencies to set performance goals, it does not require them to share the data that are needed by other agencies to access progress towards these goals. The Commission can boost federal program improvement efforts by recommending policy changes that will support goal sharing with data-sharing requirements.

A third purpose of linked administrative data is to understand the nature of challenges facing our nation, and to identify possible programs and policies that may address that. For instance, President Obama has declared that reducing inequality is the defining challenge of our time. Data that our government already collects, but which are not currently well used, could help us form more effective policies to respond to this challenge.

I'll use two cases to illustrate this point. The first case is about foreign policies to reduce education and equality by linking data from state education records, and national education surveys. During the Obama administration officials at our National Center for Education Statistics negotiated with 10 stated education agencies to prepare to link a longitudinal cohort study of high school students with state education records. NCES did its part by oversampling students in each of the 10 states, but ultimately not a single state provided its data for linkage with the national survey.

Despite agreement at the political level, NCES and its state counterparts were unable to resolve the bureaucratic barriers to linking state and federal data. This promise is ironic because state longitudinal data sets were largely built with federal grants to states, yet the federal government failed to use its leverage to compel or even encourage states to make their data available for linkage. The Commission could address this challenge by recommending legislative or administrative language that help states understand the value for their own decision-making.

The second example is about and exciting and innovate proposal to use survey and administrative data we already collect to create the American opportunity study, a major new resource for the social sciences. The last major U.S. survey of social mobility occurred over 40 years ago. By social mobility we mean the chances that persons born into disadvantage can rise above their circumstances of origin to achieve

educational and occupational success as adults.

Today, a standalone survey is no longer needed to gage patterns of mobility across generations. In fact, we can do so even more effectively than in the past by linking data from U.S. Census to federal administrative data, and to existing national surveys, that are already being conducted. Key components of the AOS include Decennial Census records, Federal Tax and Employment records, state data from federal programs such as food stamps and unemployment insurance, and national surveys that are already being done, such as American Community Survey, and the survey of Income and Program Participation and others.

The AOS would carry substantial benefits for those wishing to assess mobility trends, and form and assess policies that aim to strengthen equal opportunity processes. It would allow the formulation of evidence-based policies and when linked to research study samples, would also support program evaluation. Despite the promise of this approach from meeting the inequality challenge that President Obama has called out, substantial barriers remain. Most important, much of the data about program participation resides at the state level, even for programs supported by federal funding. If we wish to know whether our policies are working, we need to know who is participating and what their long-term and even inter-generational trajectories are. We can accomplish this by linking program participation data to census and employment records, but only if states cooperate with federal agencies and data sharing.

The federal government ought to have leverage in this regard, because the programs are federally funded. The Commission could address this challenge by recommending legislative language that would require states to share data on program participation for purposes of research on policy formation, and evaluation. More broadly -- whoops, I got behind a little it.

More broadly, the Commission should aim to standardize and streamline procedures to facilitate linkages and encourage a bureaucratic and legal environment in which such linkages are viewed as assets. To accomplish this goal, the Commission should provide a framework that articulates the technical requirements, legal standing and accepted procedures for linking and sharing data across federal agencies and with willing state partners.

The framework should specify the types of data that may be linked including administrative experimental data, and he different purposes of such linkages, including program continuation decisions, program improvement plans, and policy formation. The Commission might further recommend a pilot that would begin with specific agencies, based on the investigations the Commission seem best prepared to implement a standardized procedure, and streamlined procedure for linking data and making data accessible to researchers inside and outside of government.

As I noted at the outset, the Commission is uniquely positioned to advance the capacity to support the use of evidence in policymaking. By capacity I mean the technological structures, human capital, organizational arrangements and fiscal investments needed to create and use evidence. These considerations are often under-resourced, and demand attention at the outset. Based on my experiences outside government and my interactions with those in government I encourage the Commission to recommend specific steps to build an infrastructure inside government to use administrative data to answer policy questions, and to establish partnerships between researchers and policymakers that will increase the likelihood that research evidence permeates the policy process.

A stronger infrastructure within the federal government could help ensure that the production and use of evidence remains a high priority. It could increase the

capacity of individual agencies to produce and use evidence and align efforts across agencies. Examples of such infrastructure elements that already exist in some parts of government but which could be promoted more widely, include leadership positions focused on evidence use, interagency collaborative bodies, a dedicated office for research and evaluation within each agency, a codified set of principles and practices. Details of these examples are provided in my written statement.

Since 2009, the William T. Grant Foundation has supported studies of conditions that promote the use of research evidence in policy and practice. Partnerships between researchers and decision-makers emerge as a key finding from the scholarship as a mechanism for getting evidence produced and used. A research practice partnership is a sustained structure for facilitating relations of trust and shared goals among practitioners and policymakers and university-based researchers.

The partnership may lie between outside researchers and government officials, as is common in education, and increasingly so in child welfare. Or, it may involve researchers and policymakers within government, even within the same agency. The partnership carries two essential benefits; first, because the research agenda within the partnership is co-constructed, the questions pursued are important to the practitioners.

Second, the partnership creates a culture of evidence such that look into evidence before making decisions becomes normative practice. Fundamental to the partnership is the sharing of administrative data that allows researchers to address policy questions of interest to decision-makers. The Commission can improve the use of research evidence by identifying effective partnership models, and promoting them across the federal government, many models of partnership exists, check out our website.

The recommendations I provided are ambitious, but the Commission is well positioned to drive this work forward, especially if it focuses on the full continuum of activities from linking and sharing data to using those data to create research evidence, to using that evidence to inform policymaking. By serving as the Evidence Commission, you can help bring about smarter policies and greater opportunity for all. Thank you.

MS. ABRAHAM: Thank you. We are inspired. Kelly?

MS. FITZSIMMONS: Adam is certainly a hard act to follow. But I'm Kelly Fitzsimmons with the Edna McConnell Clark Foundation. And our Foundation has spent over a-decade-and-a-half helping non-profits use data, build evidence of their programs' effectiveness, and improve and expand those programs with the goal of helping to lift young people out of circumstances of deep disadvantage. During that time, we've worked with over 40 non-profit organizations, some of those include Nurse Family Partnership, Youth Villages, communities and schools, Harlem Children's Zone, and we've also funded organizations who play an important role in this dialogue, including results for America, MDRC, we have some MDRC friends here, Child Trends, and others.

The work that I'm going to share and then the thoughts and comments about what the Evidence Commission might consider, really come through the lens of that work. We fund our grantees directly to build evidence, and grow their programs, and so that’s really the space, those sets of experiences that I'm seeking to share with you this morning.

About 10 years ago, our grant-making model shifted from a sort of one foundation, one grant at a time, to incorporate co-investment with others. And since that time we've leveraged 155 million of our own dollars along with about 487 million of other foundation and government dollars to help support this work of growing programs and building evidence. Our grantees and the communities, families and young people that

they serve have benefited from this work -- have benefited from the increased emphasis at OMB, federal agencies in Congress, has placed on evidence-based programs. I think, indeed, I we feel we are in a golden age of evidence, and folks here have shared.

And we are proud to have partnered with the federal government in the Social Innovate Fund as an intermediary, which gave us an opportunity to bring out own tiered-evidence-base model to bear with public funding and learn about the challenges of comingling public and private funds, but learn about the challenges of building evidence in a policy-relevant, high scrutiny context. Many of our grantees also participated in the Invest and innovation Fund with the Department of Education i3, and many of our grantees have been at the frontlines in the Pay for Success Movement.

So, because of their track record, because of many of these leaders who, quite frankly, were the motivated, perhaps those that were more at the leading edge of seeking to build the evidence, they were more ready to participate in some of these grant arrangements. But there are hard-fought lessons and not everything is rosy, so I'm seeking to share, I think, bright spots here as well as challenges for consideration.

Our grantees and the communities they serve have absolutely benefited from rigorous studies that have helped them learn. In some cases, we've actually had really strong impacts and there are bright spots to hold up for sure, but in most cases our rigorous studies have produced next to no findings. Some of that is definitely due to programmatic challenges, or where programs are coming up short. Some of that is also due to methodological challenges, challenges with data, many of which have been described today. In spite of the perhaps imbalance with challenging, with challenges I think much of this is addressable, and we have an opportunity to take this on.

I think if we see ourselves as an emerging marketplace, and we are in a landscape that is shifting dramatically right now, the emphasis and focus here is a credit

to this, but I think it's also very important that we understand that much of this work is about innovation and improvement and we, together, are in an innovation and improvement life cycle. There are many constraints that providers face when building evidence, and we've experienced this again directly through our work, but more recently we've had an opportunity to step back and visit with 80 leaders in the field, and ask eh question, what more could we do to help our grantees in the field move the needle on evidence building.

We funded evidence generously. We've helped to build the capacity in grantees to do this work. We've worked with other funders and co-investors to talk about the importance of using evidence, and we've tried to promote the use and uptake in evidence, yet still we are confronting challenges. So, I want to share four buckets of challenges that, perhaps, we might consider together, how we can get under the skin.

First and foremost, and we've heard a lot about that this morning, is that the funding for evidence building, the support of capacities particularly around technology, data and policy, is limited for non-profits. This is not talent that you find in ample supply, it is hard to attract this talent to work in the social sector, it is hard to fund this work. This work is considered administrative overhead, and with the pressure to keep these costs down for many non-profits, it constricts our budget availability.

I believe most non-profits, sort of at a max probably are spending 2 percent of their operating budgets on this work. It's not sufficient to what the work is required. And I think as a result non-profits are forced to take a pay-as-you-go approach to building evidence, one study at a time, rather than think about this as continuous business process. They don't take a product lifecycle approach, they don’t think about this as an ongoing function. And I think that the role of competitions has been helpful in creating more demand, but the reality is there's not a lot of thought given to, what

happens after the competition is over? What happens after the study is over? What do we do next with the results? How do we prepare to keep on innovating, apply the results, improve implementation? Funding can definitely play a role in shifting that, by increasing the supply of resource for this capacity building. But I think, also, stimulating efforts to think differently on the part of researchers and TA providers to build in more of an ongoing capability at the level of the non-profit.

Secondly, and I think we've observed that currently the roles and incentives for evaluation researchers are not always well aligned with the practitioners needs or their local contest. And this is just a difficulty that I think we can face together, but often the timelines, the expense of evaluation, the pressure to publish that many research entities confront, are not conducive to innovation, to needing to experiment, to needing to fail in order to learn which is a core part of building evidence, we wouldn't hold most for-profit companies hostile to the publication requirements at every turn, yet non-profits feel that that is the construct that they have to operate under, even though they want to get better, they want a little bit more safe space for taking on that work.

Finally, the timeline for evidence building, and we've heard a lot about this, this morning, is not always conducive to getting information back to providers on time where they can make changes. So, we have a recent instance of a randomized control trial that took six years to get the final results from, and the non-profits needed the results sort of two years after they started the study, and of course they are going to go ahead and make changes using other sources of data. So it dilutes the power of the evaluation, and the more that we can do to think about aligning, I think these incentives, we might increase the supply and accelerate the production of knowledge in all --

MS. ABRAHAM: I just want to make sure, Kelly, that you've got an eye on the clock.

MS. FITZSIMMONS: I do. Next, I think that policymakers' demand for programs that work right now, are running ahead of the available supply of non-profits, and service providers to delivery. We do need a greater supply of organizations and leaders equipped to build and use evidence. And I think that is something that we can think about incentivizing more of the talent to come into this work; that the building of evidence as well as the delivering of evidence at the provider-level is critical.

And finally, well, policymakers, government and private philanthropy funders are demanding evidence, we seldom understand the full operational data and funding realities that non-profits face, and we don't use evidence consistently to set policy or award contracts. We might get excited about what Brain Science Research is telling us, and lose sight of the fact that most folks that work in preschools, earning $10 an hour, don't have the capability necessarily right away to adapt and apply this learning.

A quick analysis of our grantees tell us for their public contracts, again, with states and localities, they are compensated 30 cents to 70 cents on the dollar for what it takes to deliver the programs. So they are facing pressure to deliver more quality, delivery outcomes, but they can't afford to do so. I appreciate the call earlier to include cost benefit analysis into evaluation work, and research studies, because we need to make things work at scale, which means we need to be able to deliver outcomes cost-effectively.

I'm going to jump forward to quick potential solutions the Commission might consider. One, given the breadth of the providers' network in this country, the $800 billion in spending they represent, the Commission might consider crafting a platform for engaging providers directly, and the learning, the knowledge, the way that you do your work so they can learn from that.

Secondly, further incentivizing the use of evidence in procurement and

supporting efforts of program providers, not only to build their evidence but to use it and shift TA dollars from compliance to support outcomes building. The White House Office of Social Innovation just held an interesting conference on outcomes focused TA, might be something to consider there.

Finally, setting clear and shared standards for evidence, definitions and data standards and requirements, we've heard a lot about that this morning, that can actually be shared not only by government but by the private sector. So we are not sending mixed messages of what counts as good evidence.

And finally, removing barriers to accessing federal, state and local data, particularly in education where we've heard from grantees that struggled most with being able to get access to district data. I will stop there; my comments area in the write up. But thank you for this opportunity. Thank you for the work that you are doing on behalf of non-profits.

MS. ABRAHAM: Great! Thank you, all, very much. We have some time for questions from members of the Commission? Ken?

MR. TROSKE: I have a couple of questions, and I guess first for Tanya and Jim. Both of you, I think, and particularly Tanya, discussed the fact that performance metrics aren't valuations, and I tend to -- I guess I agree with that statement. And what should we think about here? I mean, I guess in my mind, you know, performance metrics, primarily focus on data that a program would be collecting already, and doesn't need permission to get access, whereas evaluation, rigorous evaluation more likely to need those linkages. Am I echoing what you have in mind? Or should we be focusing more on evaluation? Or, how important is it that we think about providing data for just, you know, measures of the performance, and measure the performance of the program?

MS. BEER: So, I'll give a short at that, but then ask that somebody else

does as well. So, I think you are right, you are putting your finger on, there's a sort of reach of the evaluative lens, that goes beyond what a program can just sort of collect in the daily curse of doing business. When you are talking about evaluation versus performance metrics, and I don't want to say that I don't think performance metrics important.

MR. TROSKE: And I didn't want to imply that either.

MS. BEER: Yes. Yes. But I do believe that there is such a pull, there's such a gravitational pull towards performance metrics, and frankly they get sort of reduced to what's just most easily accountable, that I feel like we are at risk of pushing the field towards more attention to performance metrics at the cost of actually making some determinations about the outcomes, the effects of that performance on the target audiences.

So for me it's a cautionary tale, about just trying to stay focused on making sure that -- What we are not emphasizing here is just the sort of counts of outputs and activities that, non-profits in particular, that get sort of caught up in. I'm not sure if that addresses your question but --

MR. TROSKE: Yeah. I think it does.

MR. SULLIVAN: And just from the service provider perspective, that I've had the -- you know, part of the reason we see so much emphasis on performance metrics is that's what a lot of funders ask them to collect information on and that was part of my point, I think that there needs to be more emphasis on thinking about evaluation from the funder perspective. But even in terms of, you know, when we think about performance metrics, you know, it was shocking to me when I first started talking to Homeless Prevention Call Center, that they couldn't tell us anything about the likelihood of their callers every becoming homeless.

And that seems like a simple outcome that might not be necessarily a full-brown evaluation that they simply didn't have access to, so I certainly think that access to administrative data is going to help with performance metrics, although the missing link here is that they don't have outcomes after they are served, and they don't have outcomes on the people they don't serve. And that's where administrative data fills a huge hole.

MR. TROSKE: I guess I'd like to ask all of you, too. Many of you mentioned that the results from evaluations are often difficult to implement in a timely fashion, and also a timely fashion to occur. I think a lot of it, in my mind, is focused on the fact that we seem to view random clinical trial is sort of -- that we call the gold standard, and they are expensive and hard to implement, and take a long time to create. I also think that there's a large focus on the impact of the program on the average person treated, which I don't think is a particularly useful metric.

I mean, that's a decision about whether we open or close the program, and not a metric that allows you say which parts of the program are working well, or others. So, could you help? What would you like? What are some things that we could encourage people to do in terms of evaluation research that would be more helpful for programs? Yes, go ahead, Kelly.

MS. FITZSIMMONS: Just two quick examples from some of our grantee work. Good Shepherd Services in New York wrestled with how they could undertake rigorous evaluation. They couldn't create a control group, because ethically and existentially they were not going to deny services. They were able to work, eventually over time with CIDI in New York to access administrative data, run rigorously for $140,000 that could have taken them 16 months to run, it took them another additional year-and-a-half to get through the data challenges. So, this is a clear cut case of, can we

cut the data challenges.

Youth Guidance's to Becoming a Man Program, in Chicago, that has received a lot of attention for its effects, was able to tap Chicago public schools, Chicago Police Department, to better target their program for who they needed to serve when the city was seeing spikes in certain gang activity. Being able to promote and fund that type of quick, more precision-based, rigorous evaluation, when you have policymakers, program providers and the data folks in alignment around what problem we are trying to solve, promoting incentivizing that, I think can get us more rigorous results faster, and hopefully cut through some of the cost and data constraints that are in the way. It doesn't serve all purposes, but we certainly can do more to increase the supply of that.

MS. ABRAHAM: Ron, you had a question.

MR. HASKINS: Kelly referred to age evidence-based policy, and I think I know a way a way that we can make that age quite short. And that is to constantly find that programs, don't work, and that's what we find. Do you have advice about how we can handle that problem, and what should we recommend the federal agencies that we know that 60, 70, 80 percent of evaluations and this -- in all findings that the program doesn't work, and we need to do something to makes sure the program operators don't say, the hell with you, we are not going to take your money, we don't want to evaluate because everything always turns out wrong.

MS. FITZSIMMONS: I don't have the entire answer to this massive issue, but I think there are a set of things that we can do to make sure we are using evaluation in the correct way, to ask the right questions at the right time for where an organization is in its lifecycle. I think too often we've moved, in our experience, to randomized control trials too quickly, when a series building evidence incrementally over time, with smaller scale trials learning, adjusting along the way.

Again, this is the product life cycle idea versus the big study at a time might improve the odds. I think leadership is another critical part of the answer. We look at leaders like Youth Villages in Memphis who, years ago, came to terms with the fact that things were not getting better for the target population they were trying to serve, and they got rid of the program. They replaced it after studying and looking around, finding MST, finding evidence-based solution, promoting more leadership to use early entrance reviews before they commit to evidence-based programs using the tools of What Works Clearinghouses, doing that type of evidence due diligence and promoting the adoption of the things that we do, we know work, and holding those leaders up, and giving them the appropriate recognition and incentive, I think could go a long way to supporting that flywheel along with smarter smaller-scale, targeted studies to accumulate a body of evidence.

MS. BEER: Can I also add something to that. Just very quickly, I want to pull and add something that Adam said. I think that we have a habit of talking --

MS. ABRAHAM: If you could, pull your mic up.

MS. BEER: Oh, sorry. Yes. We have a habit of talking about what works and what doesn't, and there is often, once it gets translated into the decision-making, a great deal of nuance is missed in that story, right. That it is about what works for whom under what conditions, and when evaluations aren't designed to really understand those distinctions, and communicate those distinctions, that's when we make gross generalization about what works and what doesn't, sort of, allowing for people to grapple with those nuances and understand, am I in the context where we know this works? In what ways is my context different so that I can make the necessary adaptations of these kinds of things? I think there's also kind of a framing issue in the way we talk about determinations about what works. It's really much more nuanced than

that.

MR. SULLIVAN: Actually, can I add something quickly to that, I want to reiterate that point as really important, in that there's a fair amount of trepidation about impact evaluation in the service provider community, and they've heard the comment that you made, Ron, and that perception scares them. So, they are going under the microscope to get a thumbs up or thumbs down, is really intimidating. And some in going to go back to Ken's question on that, is that instead of thinking about this as a thumbs up thumbs down, it's what works for whom, and at what time, and oftentimes the impact evaluations can lead us to certain mechanisms within the intervention or for certain subgroups of the intervention for which this has impact. And also, it's not, this doesn't work let's shut it down. If it doesn't work, why doesn't it work, and the impact evaluation can shed light on that.

MR. GAMORAN: Yeah. I'll build on that by reminding Commissioners that a big part of the purpose of evaluation is program improvement. And, so, if you can answer the questions that Jim just raised, then you can figure out, not just whether the program should be continued or not, but how to improve it. It shouldn't be surprising, and we should despair when 80 percent of impact evaluations find no results. This is the way the world works, it's hard to improve programs to do better than the status quo, and it takes a long time, and we need to be thoughtful.

But getting back to the Commission's focus, linking data, rich data about context, about what's going on inside the program, as well as the outcome is essential for that, and I think the commission can help in the in that aspect of moving from black box studies to studies that inform program improvement.

MS. ABRAHAM: One more, super quickly, and I don't know if anyone is here Department of Education, but my understanding is that the Tier 1 designs in the i3

were reviewed by a different evaluation firm than the firms that executed the evaluation, and as a result there was a higher level of stronger designs in those Tier 1 studies. There might be innovations of that nature the federal government could consider to strengthen the design and execution of rigorous studies where we are investing larger dollars, different than promoting some of the smaller-scale innovative studies. Bob?

MR. GROVES: This is an impossible question but I can't think of better people to answer it. It's motivated by privacy concerns in a world that radically beefs up evidence-based policy, so what's your insight with regard to the non-profit sector, building these programs for social good with regard to what the participants think that data that are being collected on them? How that will be used? So, is there a culture of informed consent with regard to that or what's your overall view of that landscape?

MS. FITZSIMMONS: Two quick examples. In those organizations that have tighter connections to community and are trusted in the community and stand for equity, we've seen them be able to be in relationship in a way that breeds more participants and data sharing. And Paul, you asked a question about machine learning earlier, there is a foster care organization in California that has been told they can't rigorously evaluate because of control group issues. They are working with the kids in their program to build a database of their experiences and using machine learning, combining with others so they can amass datasets and evaluate over time with a group of other partners in the area, because it's important that they are able to tell local government what is happening with the target population.

So, we've seen bright spots of this, but it really does come down to the strength of the relationship at the community level, that’s able to overcome trust. One other experience in a community setting, when evaluators actually went and talked to residence and a social program, a supported housing program and the different levels of

evaluation, and they said they were going to do a lower-level evaluation, residents pushed back, they wanted the top shelf, they wanted the RCT because they didn't feel that anybody else in the absence of that rigorous information would believe they could change. And I think there's a strong opportunity to draw constituent voice to these data in privacy discussions through community-based providers.

MR. SULLIVAN: I'll just give you a quick sense of what we've learned working with local providers. I've actually I've been reasonably impressed that informed consent is standard practice in the industry. They often get permission to link the data from their agency with partner agencies that they want to get information on so they can do a more comprehensive, holistic approach to treatments, so that's fairly standard.

What's unfortunate is I feel like the consents don't quite go far enough. A simple line like, for research purposes we may be able to link this to other administrative data. That's much less common in these kinds of informed consents, and a line like that would enable us to link these to all sorts of important outcomes, and that's in part what I was referring to in terms of standardizing this, knowing what would be the key lines that should be in informed consents so we can use local data to link to administrative data.

MR. GAMORAN: Well, I think the Commission faces a big challenge; it's absolutely the right question to ask. We have a weird climate in our country where for commercial purposes it seems fine to share data across platforms very freely, it happens all the time, and a few people object, but most people go along with it happily. But when it comes to public good purposes then we get frightened, you know, about who is looking over our shoulder.

The Commission, despite my ambitions for the Commission, I don't think you can change that, but you can, I think, make a statement or take a stance that expresses the value of data linkages for public good purposes to at least begin to stake

out that territory and explain that this is in public interest.

MS. HOYNES: Thank you very much for this very helpful conversation that we are having here. I wanted to sort of -- This is a little bit more of a comment than a question, but maybe there might be something that you could respond to. It's sort of tying a little bit, the earlier question about cost-benefit analysis, Ron's question about not passing good impact evaluations, and so on; and yet the need to incentivize and support infrastructure and legislative framework around doing more evidence-based policymaking.

And just one, I think, inherent tradeoff and challenge that I'm sure everyone in this room knows of but it's worth saying, is that typically in these settings, the costs are easy to measure, they are up front, and the benefits often accrue over a lifetime. And that's not to say that therefore we throw up our hands and don't do any evaluation, but I think it's a useful thing to keep in mind, and there are some kinds of evaluations we do where we've got upfront metrics that we know or will allow us to say something super informative in the short run. Do they show up at the homeless shelters? Or, you know, going back to the 1990s welfare reform experiments that did have evaluation required in the law, there were metrics around employment and welfare participation, and so on.

But oftentimes a lot of the benefits of the kinds of policies that we are evaluating we really don't expect to really accrue until further down the line, which I think just creates challenges for how to make good use of the evaluation. And that's not to say that I of course think it's important, but I think it needs to be said, about the limitations of what we can learn in the short run.

MR. GAMORAN: I think that's well said, and I think that also amplifies the importance of this Commission's work because gathering those long-term information

is most likely to occur if you can link it to long-term administrative datasets, such as in the Moving to Opportunity example, and other examples that the Commissioners have heard about.

MR. OHM: So, briefly, and again it's a comment more than a question. So, thank you for the comments. Adam, I just wanted to not let the record demonstrate what you just said. And I think you were speaking a little casually as in challenged. So, the idea that most Americans are happy with the kind of state of corporate information sharing, I think is quite falsifiable. I think there's a lot of people who are deeply unsettled by that.

And in the last week, this is way beyond the scope of this Commission, but I was part of an effort that convinced the Federal Communications Commission to put out a pretty strong rule that changes the default rule, when it comes to what your Internet service provider can do when they track you. And so I think there, survey after survey where people feel kind of helpless in the wake of all this commercial sharing, and I don't think it's right that they are just (inaudible) about it, or even happy about it. And so I just wanted to, like, put my two cents out there.

MS. ABRAHAM: Unless there's further conversation on this point, I actually had a question for you, Adam. You mentioned in your remarks the importance of state data for many evaluations, and you suggested that the federal government might play a role in requiring -- creating incentives for states to share data-related to federally funded programs, with the federal government to have some kind of clearinghouse where the data could be shared and linked. And you suggested further that we might consider a pilot. So, if you were going to design a pilot what would it be?

MR. GAMORAN: Yes. I'm so glad you asked that question. So, the Laura and John Arnold Foundation working Chapin Hall at the University of Chicago, has

created a series of pilots for doing these kind of data-linking. And so as I mentioned in written statement I will encourage the Commission to take a look at that example, and see what they are learning, and see how it might be built out as a pilot that the Commission could recommend. My own experience is in the field of education, and we had a fantastic failed pilot that I described where the Department of Education was unable to link with the state data.

And so another thing I might do is go back to the U.S. Department of Education and say, okay, what's one state that really has the will to do this. Because all the data are there, it's not a complex task from the data standpoint, and it was not objectionable from the political or policy standpoint. It was a matter of getting the lawyers to sign off on it is how it was described to me, as when I was a Member of National Board for Education Sciences.

So, going back to the Department of Education saying, okay, what's one state that is most ready to make this linkage, and let's connect with this national survey, which is now already in, but it's not out of date, the data still exist. And then demonstrate what that state learned from having a rich survey of students' background and schooling experiences attached to the state's administrative data on student test scores, attendance, course selection; and show that the state has learned a lot more than other states because they had this, and use that as a pilot to try to create this incentive structure.

MS. ABRAHAM: Thank you. Other questions? I thought you were going to argue -- raise the question about the performance measurement data?

SPEAKER: No --

MS. ABRAHAM: Other questions that people want to raise? If not, I'll -- I was just looking around making sure I'm not cutting anyone off. If not, thank you very

much, all very much for being here. We have one last speaker this morning, and why don't we turn to you, Naomi; Naomi Goldstein from the Administration for Children and Families.

MS. GOLDSTEIN: Thank you. So, I oversee research and evaluation at the HHS Administration for Children and Families, and I'd like to tell you a bit about how we do our work, highlighting implications for your deliberations and drawing on ACF's evaluation policy which establishes five principles; rigor, relevance, transparency, independence and ethics, as listed on this slide.

I'd like to make seven main points listed on the next and final slide (laughter) -- It's too bad that Ron had to leave to go do his Presidential duties at APEM, because he would have loved that. My comments echo many already made, but I won't let that stop me.

So one: data are necessary but not sufficient for evidence that can inform policy and practice. It is no small thing to collect data that are valid, reliable relevant and representative, but data alone are not evidence. To turn data into evidence we must examine them using a sound analysis plan. In addition, for many questions, especially questions about the impacts of policies or programs an evaluation must be designed and put into place before the data are collected. A randomized control trial is an obvious example where the study must assign participants to treatment or control groups before providing services and collecting outcome data.

This point is related to the principle of rigor in ACF's evaluation policy. Rigor means getting s close as we can to the truth. It is not an optional gold seal, rather without rigor we may be generating answers that are just plain wrong. Rigor is not code for randomized control trials, and it is not restricted to impact evaluations, it is necessary in all types of evaluations.

Two: specialized data collected for the purposes of specific evaluations, are often necessary for learning about and improving program effectiveness. This will continue to be true even as administrative data and ongoing surveys become higher quality and more accessible, there are several reasons for this, for example, some types of information may not be included in ongoing surveys or administrative data. In addition, critical administrative data may not be available for comparison groups that don't participate in programs under study.

Three: notwithstanding my previous comments, easier accessibility and improvements in administrative data would greatly streamline many evaluation activities. We don't typically use the availability of data as a starting point for deciding which questions to pursue, but we do use existing data when we can; for example, using the National Directory of New Hires, or State Unemployment Insurance records to measure the employment impacts of an intervention. However, when we need to gather data from multiple jurisdictions, the cost, time and effort of developing agreements in carrying out data exchanges can be comparable to the cost of collecting individual survey data. I'm confident that your recommendations will help.

Four: relevance is just as important as rigor and implementation and descriptive studies are just as important and impact evaluations. Pursuing rigor without attending to relevance can mean producing work that is elegant but useless. Integrating implementation and impact studies, is an especially powerful way to generate useful evidence, helping us not only learn about a program's impacts, but also what aspects of design or implementation may enhance or inhibit those impacts.

It's also important as other have said, to study the characteristics of populations and the context, so we can learn which approaches work best for different populations, or in different settings. This information is essential for scaling up or

replicating evidence-based practices and for improving effectiveness.

Five: in order to conduct evaluation, we must have at a minimum, both statutory authority and funding. This is an obvious but important point that's already been made. Several large human services programs at ACF lack authority or resources for evaluation, including refugee resettlement program that Jim Sullivan mentioned earlier. Further, most of our evaluation funding is linked to specific programs with little available for crosscutting studies.

Although the people we serve have complex needs that do not map neatly onto our programmatic structure. While some agencies such as the Department of Labor have broad authority to set aside program funds for evaluation, many agencies, including my own, do not. In addition, we need a skilled federal workforce and a robust private sector to compete for contracts and grants to carry out the work. Our evaluation policy states that we will recruit and maintain a workforce with training and experience appropriate for planning and overseeing a rigorous evaluation portfolio.

The federal evaluation and statistical workforce is populated by experts whose skills, commitment and integrity knock my socks off daily. They are worthy of respect and support. ACF's evaluation staff, whom I know the best; are a national treasure. Private organizations typically research firms and universities also play an essential role in carrying out federal research and evaluation. We rely on them, in part, for practical reasons. For one thing, Congress often provides funds that can be used for grants and contracts but not for Federal staff.

In addition, evaluations require specialized staff of different types at different times, a study may need a large staff of survey interviewers one year, and a small staff of data analysts the next. It's more practical for us to contract for these tasks than to carry them out internally. In addition to these practical reasons for conducing

work externally, this practice adds an important element of independence, another principle that I'll discuss in a moment.

Six: several bureaucratic challenges pose substantial barriers to federal evaluation in areas including procurement, information technology and security and collection of information. In the area of contracting, some federal rules are severely constricting. For example, it's difficult to gain approval for a contract longer than five years, but many evaluation questions require more than five years to answer. Even more problematic are prohibitions on incremental funding of non-severable contracts. And I'm sorry that I don't have enough time to translate that sentence into English. (Laughter)

MS. ABRAHAM: Wait, wait. Actually, seriously, would you mind saying just a few words more about that?

MS. GOLDSTEIN: Sure. Does that count against my time? Okay. Appropriations law requires that funds appropriated annually must be spent on bona fide needs of the current year. That means that we can fully fund a five-year contract with one award. However, we rarely have enough money to fully fund a contract in the year of award, or we would be able to start very few things at any given time. We are allowed to kind of string together service components of a contract and fund them at different times over the life of a contract.

But that means we have to divide a contract up into different pieces and it multiplies the complexity, so that a great deal of time and effort by my staff, who are the technical contract overseers, by the contract staff in the federal government and by the staff employed by the contractors, goes to managing those complexities, which are not necessary in some limited circumstances. such as when no year funding is provided. I would be happy to provide -- Pardon?

SPEAKER: (off mic)?

MS. GOLDSTEIN: Yes. Okay. Second area of bureaucratic challenges, in the area of ITN information security, there is ambiguity about the application of several laws to the evaluation context. Whether or not there is clarity in concept in practice there is often confusion about how to apply the Privacy Act, the Federal IT Acquisition Reform Act, and the Federal Information Security Management Act to evaluation projects. This confusion leads to delays and expense.

In the area of information collection, the same Paperwork Reduction Act rules apply to evaluation as to any other federal information collection. Required public notice periods and review by the OMB Office of Information and Regulatory Affairs, take a minimum of four months, and often eight months, or even more. This timeline can make it impossible to evaluate grants that are funded for only a few years, may prohibit gathering any baseline information on program participants or program implementation, and may prevent the timely provision of evidence to inform decision-making.

OMB has developed streamlined mechanisms for some limited situations but for the most part the requirements do not differ depending on the scope of the information collection. They are the same for a set of interviews with state officials imposing a total burden of 50 hours, and for a major of survey of thousands of people.

My seventh and last point is that The Federal Evaluation Enterprise lacks many elements of the government-wide infrastructure that supports and protects the strength and integrity of the Federal Statistical System. As one example, data collected by federal statistical agencies are covered by the Confidential Information Protection and Statistical Efficiency Act, whereas data collected by evaluation offices are not.

Second, the Federal Statistical System benefits from a formal structure for exchange of information and ideas across agencies, as well as from a statutorily-mandated leadership function at OMB. To support evaluation and the use of evidence,

OMB has created an evidence team and an Interagency Council on evaluation policy which I co-chair. However, these activities remain limited and largely informal.

Third, while some agencies have established evaluation policies there is no government-wide statement of principles for federal evaluation as there is for statistical activity. ACF's evaluation policy aims to protect the transparency and independence of evaluation, insulating evaluation functions from undue influence, and from both the appearance and the reality of bias.

However, without a national infrastructure these protections are somewhat fragile. While some agencies, notably the Institute for Education Sciences have statutory protections for independence, this is not typical. ACF leadership established our evaluation policy, and future leadership could choose to eliminate it. It would be a shame to lose safeguards such as requirements in our policy to publish study plans in advance, and to release comprehensive results regardless of what they show.

The policy also gives authority to the person in my position, which is the career position, to approve the design of evaluation projects and analysis plans and to approve, release and disseminate evaluation reports.

In closing, I note that the mission of my agency is to foster health and wellbeing through the compassionate and effective delivery of human services. The importance of these goals demands that we evaluate our activities in order to continually improve. I know that this is true for other agencies as well. You have the opportunity to enhance this work. I especially hope that you'll aim your recommendations at furthering the use of data evidence and evaluation for learning and improvement rather than primarily for answering yes or questions that are often of surprising little use for these purposes. I look forward to your report. Thank you.

MS. ABRAHAM: Thank you very much. Questions, comments?

Allison?

MS. ORRIS: Thank you, Naomi. That was really helpful and a quick overview. I wanted to just focus on your seventh point a little bit, and as the Commission things about how we can institutionalize the capacity for evidence-building and the evaluation work that you some of your colleagues we heard from earlier. Can you talk a little bit more about whether you think statutory changes are where the Commission should look at? Is it more guidance from OMB, more support from OMB? What other tools and strategies, money, just kind of wanted to give you a chance to develop on number seven?

MS. GOLDSTEIN: Sure. I don't know that I have a fixed opinion on which approaches are the best, every approach that you mentioned to accept the money has pros and cons so, you know, I don't think that it is generally helpful for statutes to include a lot of detail on evaluation activities. On the other hand, broad statements in statute about using rigorous or state-of-the-art approaches, broad statements about learning, about authority for using resources for learning and evaluation are extremely helpful.

I think that OMB's role with respect to the statistical system and OMB's emerging role with respect to evidence and evaluation have been very constructive. Obviously my comments about the Paperwork Reduction Act show that I'm somewhat skeptical of centralization and central oversight. So, I hope that I don't regret five or 10years down the line, advocating for, and I do advocate for a greater role for OMB or some other entity although I don't know what it would be, in providing leadership for federal evaluation.

I do think that statements of principle are very important, I've been surprised at how much traction ACF's evaluation policy has gotten. It's not something I

set out to do. Our Agency leadership pushed me to do it, but having done it, I find that I'm asked to speak about it frequently, other federal agencies have adopted it, so your Commission has a big megaphone that you can use to make statements of principle, and I think that goes a long way too.

So, I had some thoughts about some of the questions that were asked earlier, if people don't have other questions. (Laughter)

MS. ABRAHAM: Sure.

MS. GOLDSTEIN: Okay. So, one of the Commissioners made a point about randomized control trials being expensive, hard to implement and slow, and also tending to focus on the impact on the average person treated, and I just would like to say that I don't think any of those are inherent features of randomization. Other types of studies can also be expensive, hard to implement and show, and can focus on the average person, and RCTs can be implemented quickly, cheaply, and relatively easily, and also can look at variations. So, I think sometimes RCTs are kind of a lightning rod for broader challenges in evaluation.

MR. TROSKE: I should have said typical, because you are 100 percent correct.

MS. GOLDSTEIN: And then Ron Haskins asked how we cannot constantly find that programs don't work. So, I have three answers for that, one is to design better programs, and I don't mean that facetiously, I think that is the point of this entire enterprise. So, you know, learning what doesn't work is a way to get better at designing things better. I also think rapid cycle improvement studies are a promising and emerging strategy that are not focused on big yes or no questions but smaller, quicker questions about how we can improve practice. I think there are a lot of benefits there. I think the big-picture questions and the longer-term questions are still important, but some

of those smaller questions are also really important to answer and may be a little bit easier for local programs to swallow. So, I wanted to offer that.

MR. SHEA: Thank you, Naomi. Can you talk about a little bit more about your human capital challenges?

MS. GOLDSTEIN: Sure.

MR. SHEA: What's the availability of talent for you? How does the system enable you to get it? And a third question, I guess, is on the technology question. Is your technology challenge exclusively related to privacy and security? Or are there availability of technology procurement challenges in that area as well?

MS. GOLDSTEIN: So, starting with human capital. I think there is a deep pool of talent out there. I think that graduate education in related disciplines has advanced enormously since my own time, I learned very little about evaluation in graduate school. I've learned a great deal since then, and I am constantly impressed at the quality of the candidates that we are able to attract. Actually hiring them is extremely challenging in the federal context, and I imagine this Commission doesn't want to weigh on the challenges of the Federal HR system. I can only tell you that they eclipse the contracting challenges, and that's saying something. (Laughter)

So, the next question; I'm not sure I quite understood it.

MR. SHEA: On the IT question; or your number six suggests bureaucratic challenges related to information technology and security. And in your narrative you suggest privacy and security challenges. Are there other technology challenges to getting your work done? Do you have the tools to get your work done? Or could they improve?

MS. GOLDSTEIN: I am not sure of the answer to that question, so let me pivot it a little bit. Something that I didn't emphasize was increasing capacity and

focus on designing information systems so that they can talk to each other. So there certainly are privacy issues and bureaucratic issues but there are also technological issues. And I think that's an area with tremendous potential that is increasingly coming to the forefront.

MS. ABRAHAM: A comment and a question. The comment is, I was very struck, having spent a lot of time in the federal statistical system by the parallels you were drawing between the way that the infrastructure that's set up to protect the independence of the federal statistical system, doesn't exist with respect to the evaluation function within government, and just speaking for myself, I think that that's something that we'll want to think more about for sure.

My question relates to sort of models for building the evaluation capacity within the federal government, our first speaker this morning was Demetra Nightingale who you obviously know, and my Co-Chair, Ron, was effusive in his comments about the Chief Evaluation Officer model that's been established at the Department of Labor. Do you have a thought or a view about what -- the right way to structure the evaluation function within federal agencies is?

MS. GOLDSTEIN: Kind of. I have to acknowledge that federal agencies differ quite a lot, so how to structure the evaluation function may differ, in some degree, depending on the context. Just to go back a little bit in history, and I don't know how much Demetra may have addressed this. The current, robust evaluation function at the Department of Labor did not always exist, and did not come about by accident. There were some unfortunate incidents in the past where the function and -- you know, was weakened and reports were suppressed.

There was a series of GAO reports about that, which I referred extensively in putting together ACF's Evaluation Policy. And as I understand it, there's

both law and strong encouragement from OMB to develop the current structure that exists. I do think some degree of centralization at some level is important, so at the Administration for Children and Families, we are the primary, but not the only place where evaluation is housed. And over the last 10 or 15 years there's been an increasing move towards centralization within ACF, so that there was a separate office for early childhood evaluation, there was a -- Sorry, I'm losing track of how my empire has grown (laughter). So, some of that has been centralized.

SPEAKER: You left your tiara at home.

MS. GOLDSTEIN: I did. There definitely are benefits, you know, that we have kind of critical mass of really talented people who can bounce ideas off each other. But people who have been in evaluation offices that are located within programs will tell you there are benefits to that as well. So, you know, independence and rigor perhaps are better supported by some degree of centralization; relevance, may be better supported by some degree of decentralization.

So I think these structural questions are interesting and important, I also know that our Interagency Council on Evaluation Policy is working to kind of gather information on different models, and sort of put it out there so people will have something to refer t when working with through those questions.

MS. HOYNES: I have a little bit of a follow up to that, and that is your thoughts about, one of the things that we've heard throughout today and at other times, is that the lack of resources, not just for the evaluation but for sort of assembling, cleaning and documenting the administrative data that agencies are creating as part of their business-as-usual model. And so my question is, do you think that that capacity would naturally be built into an evaluation structure that you were just having a conversation with Katharine about? Or would that better belong in the statistical framework, or

somewhere else?

MS. GOLDSTEIN: Well, I don't know the answer. I will say that my agency, the Administration for Children and Families does not have a statistical structure. So, administrative data are collected within specific programs, we don't have a statistical office. We have recently established a new division, a division of Data and Improvement that we intend to work with programs to help improve the quality, usefulness, availability and analysis of administrative data, so I think we'll be learning a lot about the kind of question that you're asking.

You know, there are a lot of related activities. There are statistical activities, there's performance measurement, there's strategic planning, there is evaluation. I think that can be configured in a lot of different ways. I think that it's difficult to integrate them, and I don't think that there's one right way to structure it.

MR. TROSKE: So, thank you for coming. Your comments have been quite helpful and thought-provoking. So, I'm going to put you on the spot, since you've been so helpful.

MS. GOLDSTEIN: Thanks.

MR. TROSKE: Yeah. No problem. So, if you were tasked with designing a structure that overcame -- I mean, you point 5, 6 and 7 and just, here are all the limitations we have, lack of a skilled federal workforce, a robust, bureaucratic challenges, and the evaluation that lacks many elements of the infrastructure that supports and protects the federal statistical system, what would you do?

MS. GOLDSTEIN: Do you mean if I were you, or if I were ZAR?

MR. TROSKE: If you were ZAR. Because I think we get to be ZAR, I mean, in some sense. At some point we get to be ZAR and then maybe we have to be realistic.

MS. GOLDSTEIN: Okay. Well, I would streamline Paperwork Reduction Act requirements for evaluation. I would establish a -- and this is maybe actually, I'm consistent, but I would establish an entity that was a central government entity for providing coordination and leadership for federal evaluation activities. To what degree that entity would have authority and oversight, I'm not sure, but certainly the coordination and leadership are important. I would like to reform the entire Federal HR system, but I --

MR. TROSKE: You already indicated that we were not allowed to do that, and I think I'm going to agree with you there.

MS. GOLDSTEIN: But I might try to create some specialized Pathways programs for evaluation, and I will allow incremental funding of non-severable contracts.

MR. TROSKE: Thank you.

MS. ORRIS: Thank you. I guess one additional question, and I don't know if it would go on your wish list, but we've heard a little bit this morning about the challenges that states have in sharing their data, accessing data, and I imagine that, especially at ACF and HHS where so much of the work happens at the state level, that's a challenge, so I just wanted to give you a chance to elaborate on anything you would add, or just to think about from the state perspective.

MS. GOLDSTEIN: Sure. So I did once work in a state, but it was a long time ago, so I'm probably not the best person to answer this question. You know, I think, as I'm sure has been said at this meeting, and other meetings of the Commission. There's a complex web of laws and rules at the state, local, and federal levels. It's very difficult to navigate through. I think, and hope that there will be efforts to develop, you know, standard procedures for Memorandum of Understanding, data exchange, I think it will likely have to be an incremental and difficult process, but that's what I see as the way to try to move forward.

MS. ABRAHAM: Great. I'd like to thank everyone who came to speak with us this morning. We've certainly heard a lot to think about, and we maybe will want to follow up with many of you with further questions. So thank you so much for taking time to come and talk with us.

We will adjourn the meeting at this point. This will not be our last occasion to talk about evaluation. We look forward to future exchanges.

* * * * *

CERTIFICATE OF NOTARY PUBLIC

I, Carleton J. Anderson, III do hereby certify that the forgoing electronic file when originally transmitted was reduced to text at my direction; that said transcript is a true record of the proceedings therein referenced; that I am neither counsel for, related to, nor employed by any of the parties to the action in which these proceedings were taken; and, furthermore, that I am neither a relative or employee of any attorney or counsel employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

Carleton J. Anderson, III

(Signature and Seal on File)

Notary Public in and for the Commonwealth of Virginia

Commission No. 351998

Expires: November 30, 2016