

Statement of Naomi Goldstein
Deputy Assistant Secretary for Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
to the Evidence-Based Policy-Making Commission

4 November 2016

Dr. Haskins, Dr. Abraham, and Commission members:

Thank you for inviting me to speak with you today. You have a tremendous opportunity to further the creation and use of evidence in the federal government and beyond.

At the HHS Administration for Children and Families, I serve as Deputy Assistant Secretary for Planning, Research, and Evaluation, overseeing a portfolio of grants and contracts to conduct research and evaluation related to ACF programs. 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 to govern our work. These are: rigor, relevance, transparency, independence, and ethics.

I'd like to discuss seven main points:

1. Data are necessary but not sufficient to create evidence.
2. Administrative data and ongoing surveys are important resources, but specialized data collected for the purposes of specific evaluations will also continue to be important.
3. Easier accessibility of administrative data would greatly streamline evaluation activities.
4. Implementation and descriptive studies are just as important as impact or outcome evaluations, and relevance is just as important as rigor.
5. Prerequisites for federal evaluation include statutory authority, funding, a skilled federal workforce, and a robust private sector.
6. Several bureaucratic challenges pose substantial barriers to federal evaluation efforts, in the areas of procurement, information technology and security, and information collection.
7. The federal evaluation enterprise lacks many elements of the infrastructure that supports and protects the federal statistical system.

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 *per se* are not evidence. To turn data into evidence, we must examine them using a sound analysis plan tailored to questions of policy and programmatic interest. In addition, for many questions – especially questions about the impacts of policies or programs – data must be used in the context of an evaluation 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 as 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. ACF's evaluation policy states that we are committed to using the most rigorous methods that are appropriate to the evaluation questions and feasible within budget and other constraints. Rigor is not restricted to impact evaluations, and it is not code for randomized control trials. It is necessary in all types of evaluations, including implementation or process evaluations, descriptive studies, outcome evaluations, and formative evaluations; and in both qualitative and quantitative approaches. Rigor requires ensuring that inferences about cause and effect are well founded (internal validity); requires clarity about the populations, settings, or circumstances to which results can be generalized (external validity); and requires the use of measures that accurately capture the intended information (measurement reliability and validity).

In assessing the effects of programs or services, ACF's policy is to use methods that isolate to the greatest extent possible the impacts of the programs or services from other influences such as trends over time, geographic variation, or pre-existing differences between participants and non-participants. For such causal questions, experimental approaches are preferred. When experimental approaches are not feasible, high-quality quasi-experiments offer an alternative.

Two: Administrative data and ongoing surveys are important resources, but specialized data collected for the purposes of specific evaluations will also continue to be necessary.

Administrative data and data from ongoing surveys are important resources for learning about program effectiveness and informing program improvement. I am confident that the Commission's recommendations will strengthen the value and availability of these types of data. But specialized data collected for the purposes of specific evaluations will also continue to be necessary, for several reasons:

- a. Some types of information may not be included in ongoing surveys or administrative data. For example, consider measures of young children's socio-emotional or cognitive development.
- b. Administrative data may be available to measure experiences or outcomes of program participants, but typically are not available for comparison groups that don't participate in the programs under study.
- c. Administrative data may lack the quality, completeness, and reliability needed for evaluation purposes.
- d. Data from ongoing surveys may not capture samples adequate for specific populations or for addressing specific program- or policy-related questions.
- e. Administrative data and ongoing surveys are sometimes available only after a considerable time lag.

Three: Easier accessibility and improvements in administrative data would greatly streamline evaluation activities.

We don't typically use the availability of data as a starting point for deciding which questions to pursue. But we do take advantage of existing data sources when we can -- for example, using the National

Directory of New Hires or state Unemployment Insurance records to measure the outcomes of an employment-focused intervention. Easier accessibility of administrative data would greatly streamline our efforts. Particularly when we need to gather data from multiple jurisdictions, the cost, time, and effort of developing agreements and carrying out data exchanges can be comparable to the costs of collecting individual survey data.

Although well-integrated administrative data systems suitable for supporting complex evaluations can take years to develop and require broad-based collaboration across many agencies and levels of government, such systems could reduce costs and time to carry out a range of evaluations. Some specific improvements to make administrative data more useful include adoption of common data semantics and incorporating data elements to better allow data linking.

Four: Implementation and descriptive studies are just as important as impact or outcome evaluations, and relevance is just as important as rigor.

Implementation and impact studies together are especially powerful, allowing us not only to understand whether a program had desired impacts, but also to explore what aspects of the program design or implementation may have enhanced or inhibited those impacts. It is also important to understand the characteristics of relevant populations, and the context, so we can learn which types of services work best for different populations or in different settings. All of these types of work provide critical information for scale-up or replication of evidence-based practices, and for efforts to improve the effectiveness of policies and programs.

Pursuing rigor without attending to relevance can mean producing work that is elegant, but useless. To ensure relevance, we take into account many influences in setting evaluation priorities: statutory requirements and Congressional interests; the interests and needs of executive branch staff and leadership; partners such as states, territories, tribes, and local grantees; the populations served; researchers; and other stakeholders. ACF's evaluation policy calls for planning evaluations in concert with the planning of a program or initiative, rather than as an afterthought. To make sure that evaluations are relevant, we aim to build strong partnerships among evaluation staff, program staff, policy-makers and service providers.

Five: Prerequisites for federal evaluation include statutory authority, funding, a skilled federal workforce, and a robust private sector.

In order to conduct evaluation to inform policy and program decision-making, we must have at a minimum both statutory authority and funding. In addition, we need a skilled federal workforce, and a robust private sector to compete for contracts and grants to carry out the work.

We conduct evaluations in those areas where Congress has provided resources and authority to do so. Several large human services programs at ACF lack both authority and resources for this type of work. Further, most of ACF's funding for evaluation is linked to specific programs, with little funding available for cross-cutting studies, although the individuals and families we serve have complex needs that do not

map neatly onto ACF's programmatic structure. While some agencies (such as the Department of Labor) have broad authority to set aside program funds for evaluation, many agencies do not.

Our evaluation policy states that ACF will recruit and maintain an evaluation workforce with training and experience appropriate for planning and overseeing a rigorous evaluation portfolio. To accomplish this, ACF will recruit staff with advanced degrees and experience in a range of relevant disciplines. ACF will provide professional development opportunities so that staff can keep their skills current. The federal evaluation and statistical workforce is populated by experts whose skills, commitment, and integrity impress me every day.

Private organizations, typically research firms and universities, 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 appropriates 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. An evaluation may need a large complement of survey interviewers one year, and a small staff of data analysts the next year. It is more practical for the government to contract for these tasks than to carry them out internally. In addition to these practical reasons for conducting work extramurally, this practice adds an important element of independence, another principle of our evaluation policy.

ACF aims to ensure that contractors and grantees conducting evaluations have appropriate expertise through emphasizing the capacity for rigor in requests for proposal and funding opportunity announcements. This emphasis entails specifying expectations in the criteria for the competitive selection of grantees and contractors, and engaging reviewers with evaluation expertise.

Six: Several bureaucratic challenges pose substantial barriers to federal evaluation efforts, including requirements related to procurement, information technology and security, and collection of information.

First, some federal contracting rules are severely constricting. For example, it is difficult to gain approval for awarding 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 for non-severable contracts.

Second, there is ambiguity about the application to evaluation of several laws related to data. While there may or may not be clarity in concept, in practice there is often confusion on 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.

Third, information collection for evaluation purposes is subject to the same requirements as any other federal information collection, under the Paperwork Reduction Act. Required public notice periods and review by the OMB Office of Information and Regulatory Affairs take a minimum of 4 months, and often 8 months, or even more. This timeline can make it impossible to evaluate grants that are funded for only a few years; may prohibit the collection of baseline information on program participants; and may

preclude the timely provision of evidence to inform policy or program decision making. While OMB has developed streamlined mechanisms for some limited situations, for the most part the requirements do not differ depending on the size or 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 a major survey interviewing thousands of people and imposing thousands of burden hours on respondents.

Seven: 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.

1. As one example, data collected by designated federal statistical agencies are covered by the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), whereas data collected by evaluation offices are not.
2. 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 Inter-Agency Council on Evaluation Policy, which I co-chair. However, these activities remain limited and largely informal.
3. Third, while ACF and some other agencies have established evaluation policies, there is no government-wide statement of principles for federal evaluation. Just last week the National Academies of Sciences held a workshop to comment on existing evaluation policies of federal agencies as well as the desirability of a cross-agency or government-wide statement of principles. This might follow the model of the NAS publication, *Principles and Practices for Federal Statistical Agencies*.

ACF's evaluation policy aims to protect the transparency, independence, and objectivity of evaluation, and to insulate 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 Department of Education's 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:

- Make information about planned and ongoing evaluations easily accessible, publish study plans in advance, and release comprehensive evaluation results regardless of the findings.
- Conduct evaluations through the competitive award of grants and contracts to external experts who are free from conflicts of interest.
- Place authority with the career director of the Office of Planning, Research and Evaluation to approve the design of evaluation projects and analysis plans; and to approve, release and disseminate evaluation reports.

Conclusion.

The mission of my agency is to foster health and well-being by providing federal leadership, partnership, and resources for the compassionate and effective delivery of human services. Our vision is children, youth, families, individuals and communities who are resilient, safe, healthy, and economically secure. The importance of these goals demands that we continually innovate and improve, and that we evaluate our activities and those of our partners. This is true for other agencies as well.

You have the opportunity to enhance this work. I particularly hope that you will aim your recommendations at furthering the use of data, evidence and evaluation for learning and improvement, rather than primarily for answering yes/no questions that are often of surprisingly little use for these purposes. I look forward to your report. Thank you.