1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Telephone (202) 484-9220
mathematica-mpr.com

December 14, 2016

Dear Members of the Commission on Evidence-Based Policymaking:

Thank you for giving the research community an opportunity to contribute to the important work of the Commission. For nearly 50 years, Mathematica has been dedicated to delivering evidence-based research, data, and objective analysis of the highest quality. We believe that data used for evidence-based policymaking should be viewed as a treasured resource, one that must be carefully safeguarded. It is clear that the Commission takes safeguarding this resource very seriously, as we do at Mathematica. However, good stewardship of any national treasure must include employing it for the benefit of the nation. We fall woefully short of meeting our stewardship obligations for these data when we fail to realize that the difficulties in accessing them or the restrictions on their use keep much of their value locked away. At Mathematica, we are honored to support the Commission in addressing this difficult challenge.

In response to question numbers 1, 6, 9, 10, 16, 17, 18, and 19 listed on the "Commission on Evidence-Based Policymaking Comments," docket number: USBC-2016-0003, I—along with my colleagues Tamara Barnes, Peter Schochet, Irma Perez-Johnson, and Alexandra Resch—have drafted the following comments. We expect these insights will help inform the work of the Commission and potentially provide discussion topics for future Commission meetings.

Mathematica Response to the Commission's Request for Comments

Question # 1. Are there successful frameworks, policies, practices, and methods to overcome challenges related to evidence-building from state, local, and/or international governments the Commission should consider when developing findings and recommendations regarding Federal evidence-based policymaking? If so, please describe.

Researcher-practitioner partnerships have shown promise as a framework that facilitates continuous program improvement and evidence-building at all levels of government. Numerous partnerships focused on using behavioral insights to improve programs have developed in recent years. The White House Social and Behavioral Sciences Team partnered with several federal agencies to improve programs by simplifying communication to program participants and employing behavioral nudges. Mathematica is working with three agencies in the U. S. Department of Labor (DOL) on Behavioral Interventions for Labor-Related Programs, a project in which we apply behavioral insights to DOL programs. In these three examples, program changes were evaluated rigorously, often using randomized controlled trials (RCTs). We provide examples of similar partnerships with state and local Temporary Assistance for Needy Families (TANF) agencies in our response to question #16.

These partnership approaches combine evaluation and evidence building with a cycle of activities to support better programs and policies that are based on evidence. Activities are adapted for each specific program. Most effective partnerships recognize that different policy questions require different levels of

evidence and can support different levels of program evaluation, allowing us to right-size the evidence-building to the problem at hand.

To achieve an ideal framework:
1. It is important to begin with a clear and well-articulated definition of the problem.
2. Next, efforts to improve policies and programs should be grounded in the available knowledge base about why the problem exists, including what has been done to address it and in what contexts. If available, this knowledge base provides solid grounding and context for further innovation and experimentation.
3. Then we identify untested, promising strategies and decide which to pursue or test given operational, contextual, or other constraints.
4. Once promising strategies are selected, we can move into prototyping, refining the selected strategies with input from relevant users or other stakeholders. We attempt to pilot-test on a large enough scale to assess feasibility and uncover important implementation challenges. Then we continue refining additional iterations until we achieve a practical model.
5. We develop a proof of concept test to yield evidence on whether the promising strategy leads to result(s). If yes, we may refine further based on implementation experiences or move on to more rigorous tests. If not, we may go back to Step 3 to explore other promising approaches.
6. Once we have a promising idea on a pilot scale, we can shift to testing, using the most rigorous method(s) possible.
7. If the program or policy concept is proven at scale, we can move to broader adoption or adaptation in other contexts while maintaining effectiveness.

Evidence-building and testing activities permeate each step and must be right-sized to fit our goals and answer the appropriate questions at each stage of the process. This is the framework that underlies our evidence-based technical assistance (EBTA) and other support activities. We find that when program administrators participate in these activities and the evaluation efforts are focused on program needs, program administrators quickly see the value of evaluation and evidence and are enthusiastic about participating in further evidence-building.

Currently, this approach is rarely implemented sequentially or fully in the way we have described. Rather, evidence-building activities happen at any point in the sequence, with or without the benefit of the evidence from the preceding steps. The end result is a spotty mix of evidence on the wide range of government programs and policies in existence, with evidence for some stages of continuous program improvement, but not all. However, there are promising signs (including the work of this Commission) of a growing interest in and recognition of the value of an evidence-based approach to policy and program development and ongoing improvement.

Question #6: Should a single or multiple clearinghouse(s) for administrative and survey data be established to improve evidence-based policymaking? What benefits or limitations are likely to be encountered in either approach?

The idea of establishing a single clearinghouse for administrative and survey data is very attractive, but is likely to take a long time to achieve, and during the wait for that goal to be achieved, its worth may not be evident to all. An alternative framework that may be more easily implemented would require each agency that houses administrative and/or survey data to establish its own data clearinghouse center that

would coordinate with a centralized body acting as a "virtual" data clearinghouse. This virtual data clearinghouse would be the body through which applications for access to data would be submitted and approved, and would serve as the coordinator between each agency's data clearinghouse center, facilitating access to and linking of data across agencies.

Question #9: What specific administrative or legal barriers exist for accessing survey and administrative data?

In its white papers provided to the Commission, entitled *Using Administrative and Survey Data for Evidence-Building* and *Barriers to Using Administrative Data for Evidence-Building*, the Office of Management and Budget did an exemplary job of laying out virtually all of the administrative and legal barriers that we face daily at Mathematica in our work with administrative and survey data.

What is required above all else is fostering a mindset that administrative and survey data are among our national treasures and should be put to their best and highest use. At Mathematica, we have observed that some federal agencies are slowly shifting their mindset in this direction. Federal legislation that embodies this principle and mandates updating laws and agency regulations in light of this principle would be invaluable in achieving this goal more broadly among all stakeholders, including federal agencies, state and local governments, and federal program grantees.

Gaining access to the administrative data of state and local governments can be particularly challenging, and this is an area where Mathematica confronts the reality of federalism and our republican form of government almost daily. In laying a foundation for cooperation in this area, it would be beneficial to establish a model state act that incorporates the principle of stewardship of administrative and survey data as a state treasure. This model state act could, for example, establish a mandate and processes for providing access to administrative data for purposes that benefit both the state and the nation when appropriate. Such a model state act would be similar in nature to other model state acts, like the model administrative procedures act or model public health act.

With respect to federal program grantees, particularly those that are not state or local government entities, a condition of the grant should be the provision of administrative or survey data by the grantee. Standardized language to include in all grant agreements could be developed as a model for sponsoring agencies. In cases where individual consent is required for the provision of such data, the sponsoring agencies can and should plan to provide standardized consent forms and other tools to help ease the burden of obtaining individual consents.

Finally, with agencies holding the attitude that administrative and survey data are a national treasure, one action that can be initiated right away is the use of System of Record Notices (SORNs), which are mandated by the Privacy Act. SORNs are the instruments agencies use to inform the public of what the agency's permitted uses and disclosure standards are for data under their control. A properly drafted SORN would be particularly helpful with respect to survey data that were collected with individual consent. A properly drafted SORN coupled with an individual consent form that properly references the sponsoring agency's SORN, or any amendments to the SORN, would govern the future use, protection, and disclosure of the collected survey data.

LETTER TO: Commission on Evidence-Based Policymaking
FROM: Mathematica Policy Research
DATE: December 14, 2016
PAGE: 4

Question #10: How should the Commission define "qualified researchers and institutions?" To what extent should administrative and survey data held by government agencies be made available to "qualified researchers and institutions?"

We understand this definition is the basis on which decisions will be made as to who can gain access to administrative and survey data in the data clearinghouse(s). Defining who is "in" and who is "out" is always an important task. Mathematica urges the Commission to view this task as possessing the inherent possibility of shaping the future of evidence-based research by defining who can and cannot conduct such activities with data from the clearinghouse(s). We do not have a crystal ball in which to see the innovations or changes in industry and occupations that will produce "qualified researchers and institutions" decades from now. What we can see is that if you limit this definition to classifications of individuals or institutions as they exist today you may, ab initio, limit the future of the development of public policy research, and we believe such a limit to be a bad thing. We strongly recommend that the Commission use objective standards that are not based on the classification of individuals or entities. Some such objective standards could be, for example, that the proposed use of the data is permitted by law, that relevant ethical and conflict of interest disclosure standards are met, and that the individual or entity can meet data confidentiality and security standards.

Question #16: How can data, stats, results of research, and evaluation findings best be used to improve programs and policies?

Research can be translated into actionable policy changes and program improvements at all levels of government. The examples below highlight just a few of the ways that Mathematica's research results and evaluation findings are being used to improve programs and policies at the federal and state level.

- Our evaluation of Adolescent Pregnancy Prevention Approaches (PPA) is a major federal effort to expand available evidence on effective ways to prevent pregnancy and reduce related sexual risk behaviors among teens in the United States. The eight-year (2008–2016) evaluation documented and tested new and innovative approaches to teen pregnancy prevention in seven sites across the United States. Six of the seven sites received federal funding from the Office of Adolescent Health or Administration on Children, Youth and Families in the U.S. Department of Health and Human Services as part of the federal government's broader effort to invest in evidence-based approaches to teen pregnancy prevention.

- For DOL, the Behavioral Interventions for Labor-Related Programs project involves working with three DOL agencies to examine how behavioral science principles can improve the performance and outcomes of selected programs. For example, Mathematica worked with the Occupational Safety and Health Administration (OSHA) to test whether changes to its citation process could increase employer responsiveness and reduce referrals to the OSHA National Office. Behavioral solutions were found to improve program outcomes and performance, and they can inform the broad adoption of such principles within DOL and many other related programs nationwide. Insights gleaned from this work are contributing to the growing body of evidence on the effectiveness of behavioral strategies.

LETTER TO: Commission on Evidence-Based Policymaking
FROM: Mathematica Policy Research
DATE: December 14, 2016
PAGE: 5

- In work for the Millennium Challenge Corporation, Mathematica's evaluation of "girl-friendly" schools helped identify program changes that could positively impact girls' education in Niger. Valuable data from Niger provided compelling evidence for governments and donors in thinking about how to develop effective programs to provide access to quality education for children across the developing world, especially girls. Mathematica researchers conducted the evaluation in 178 villages across Niger. They assessed efforts to improve educational outcomes for children in targeted communities and took a concerted look at environmental factors related to increasing girls' access to schooling.

- As part of an "employment-focused programming" initiative, sponsored by the Colorado Department of Human Services (CDHS), in 2015–2016, Mathematica worked intensively with three local welfare agencies to assess their readiness to implement employment-focused programming and identify opportunities for innovation and improvement. We identified one of these sites, Larimer County, as the top candidate and formalized a researcher/practitioner partnership with its TANF program administrators. Over the past two years, we have designed and tested changes to communications and other nudges to help TANF recipients better understand program rules, complete required activities, avoid sanctions, and achieve self-sufficiency. As a result of this partnership, Larimer County staff have learned that evaluation can be timely and can directly inform day-to-day policy decisions.

- Another researcher-practitioner partnership with Ramsey County, Minnesota's TANF agency, identified a new service delivery approach—the Lifelong Learning Initiative (LLI). In 2015, Mathematica was contracted to perform three years of technical assistance on the design, refining, and ultimately the full-scale implementation of this concept. During Phase 1 of the project, we facilitated a collaborative learning and design process in which staff at all levels of the organization (administrators, supervisors, frontline staff, and even clients) contributed to an in-depth program assessment and provided input on the intervention's early design. In Phase 2, we created intervention materials based on Phase 1 recommendations and launched three six-week learning cycles in which intervention training was provided to a cohort of staff. After the three learning cycles, we revamped and polished all of the intervention materials and delivered training for all agency staff in August 2016. Now our focus is on enhancing and strengthening the LLI intervention by incorporating rapid cycle testing and other formative evaluation techniques into implementation.

Question #17: To what extent can and should program and policy evaluation be addressed in program designs?

The continuous program improvement framework presented in response to question #1 indicates that evidence can and should be integrated into every stage of program and policy design and development. Research and evaluation activities should be aligned with the goals at any given stage and can be right-sized to meet important constraints or other considerations. Our experience working with local and state agencies suggests that many agencies do not understand their role in program evaluation. In fact, many see research as something that is done to them and often does not closely relate to their day-to-day program concerns. Agencies developing or implementing new programs are eager to learn whether they are effective, but they struggle to find the resources, including staff time and expertise, to conduct evaluations themselves.

To support this work, it would be ideal to set aside a fraction of program funding to support a full range of evidence-based, continuous quality improvement and evaluation, not just one-time program evaluations. Agencies beginning this work also benefit from technical assistance that helps them use evidence and evaluation to answer their own day-to-day policy questions, demonstrates that evaluation can be useful and timely, and develops their capacity to interpret and generate evidence.

Question #18: How can or should program evaluation be incorporated into program designs? What specific examples demonstrate where evaluation has been successfully incorporated in program designs?

It is important that policymakers use evidence-based research to improve their interventions, programs, and policies. Evidence-based research must be rigorous, with proper research questions and study designs to answer the questions a study was meant to answer. Too often, policy is made using subjective judgement or poorly designed research.

To make research most effective, it is critical for it to be ongoing and allow continuous program improvement, instead of being conducted at interim points only. This means that staff at government agencies must be included in the scientific process so they are invested in the research from the start. It is also critical to recognize that research that will be useful to program managers will include a wide range of research methods, not only RCTs. Program managers seeking to refine a current program may use descriptive or non-experimental analyses that help generate hypotheses about what is working and what is not. They may also use machine-learning methods and big data to predict which services might be most effective for each program participant based on his or her characteristics and needs. These managers could then propose changes to their program and evaluate these changes using an RCT or a quasi-experimental design (QED) before deciding to roll out the change widely.

For this to be successful, the nation needs to expand the research community beyond the relatively small number of academics and research organizations that are currently conducting social policy research to investigate the effectiveness of policies and programs. Government agency staff are natural participants to include in research efforts, because (1) they are best informed about the proper interventions to test and how they should be implemented, (2) they can build "rapid-cycle" evaluations into their daily service activities to test incremental changes in their programs, and (3) they may have access to administrative data that can provide outcomes for the evaluations. Perhaps most important, these staff could weigh in from the start about how the evaluation results can be used in practice and be involved in plans for improving the tested interventions.

The examples of researcher-practitioner collaborations provided in response to question #16 demonstrate that agency staff can participate in this work and develop internal capacity for further evaluation and evidence-building. Additional support to agency staff comes in the form of evaluation technical assistance and toolkits to support evaluation. Mathematica has supported these efforts through several contracts, including one with the Administration for Children and Families in the U.S. Department of Health and Human Services and another with the U.S. Department of Education (ED). Through these contracts we provide technical assistance to grantees and local areas conducting impact evaluations, and we produce user-friendly materials (such as how-to briefs, software, and webinars) to support their evaluations. Our recent experience developing an online toolkit for rapid cycle evaluations of educational

technologies for the Office of Educational Technologies at ED suggests that there is high demand for these products. The beta version of our toolkit was released in late October, and we have received inquiries from over 70 school districts that are interested in using the tools and receiving technical assistance. These efforts should be expanded.

The development of evidence at the local level may complement large federal evaluations, but the evaluation of federal initiatives requires evaluations with a broader focus. However, to be most useful to program administrators, large federal evaluations should include careful evaluation of specific program components and variation in program effects. Much federal research has been designed to answer "big" questions such as "What are the average effects of a federal grant program on participants' outcomes?" Although a rigorous RCT for such an evaluation can help us assess whether individuals receiving grant services have better outcomes, on average, than a control group does, it is not always clear how such results can foster continuous program improvement. This is because a complex grant program can have multiple program components and mediating pathways associated with the overall program impacts, and a typical "thumbs-up/thumbs-down" evaluation of a grant program examines the effects of each mediating factor in isolation using non-experimental methods that are prone to sample selection biases. Thus, it may be useful for those who develop research agendas to consider smaller research questions about specific mediators that are varied as part of the research design (for example, different dosages of case management services).

These types of results may be more useful to program managers, but would require a research agenda that could build over time to help us understand which interventions are most effective and for whom, and whether they can be replicated in different settings.

Question #19: To what extent should evaluations specifically with either experimental (sometimes referred to as "randomized control trials") or quasi-experimental designs be institutionalized in programs? What specific examples demonstrate where such institutionalization has been successful and what best practices exist for doing so?

RCTs and QEDs answer questions of impact, which are at the core of assessments of the effectiveness and efficiency of public investments. As such, they should be institutionalized into every public program and policy where appropriate. However, it is most critical to develop evidence of the impact of previously unproven programs. With limited public resources, it seems prudent to focus the effort and cost of rigorous evaluations on the programs that are not yet proven. Tiered evidence grants are an example of institutionalizing evaluation requirements while recognizing that different stages of program development require different types of evidence. These grants support innovation by providing some funding to promising programs that are not yet fully developed, but appropriately require rigorous evaluation using RCTs or QEDs when the program is ready to be scaled up to widespread use. It is important to recognize that other methods may be most appropriate for early program stages and to support program development and refinement before investing in more rigorous, and typically more costly, evaluations.

Focusing on the rigorous evaluation of unproven programs, some laws authorizing funding of public programs require RCT or QED evaluations of the programs (for example, the Workforce Innovation & Opportunity Act). In addition, some RFPs for federal grants to implement programs require that grantees participate in or implement their own evaluations. For example, grant programs including Race to the Top and School Improvement Grants required state cooperation with evaluations and facilitated the research on these programs. These regulations have been an important impetus for rigorous program evaluations that are much less common in Europe and elsewhere outside the United States. Nonetheless, there could be

better enforcement by federal agencies or strengthening of these regulations by making participation in or cooperation with evaluation activities a condition for receipt of federal grant funds.

Mandated participation could improve the efficiency of federal program evaluation efforts by reducing recruiting costs. Recruiting sites for program evaluations when participation is voluntary is often difficult and expensive, because we must overcome sites' resistance to participate in RCT or QED evaluations and also gain cooperation against competing site priorities. Site recruitment can be especially challenging when grants are disbursed at the state level but programs are run locally (for example, education programs). Stronger enforcement and mandates for site participation in evaluation at all the necessary levels can help spur research, reduce evaluation costs, and improve the generalizability of study results by facilitating the recruitment of a more representative set of sites than are typically included in program evaluations.

Thank you again for this opportunity to answer these important questions. Should you have any questions about the contents of this submission, please contact me at PDecker@mathematica-mpr.com, or contact my colleague Carmen Ferro, public affairs manager at Mathematica Policy Research, at (202) 552-6405, or at cferro@mathematica-mpr.com.

Sincerely,

Paul T. Decker
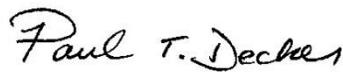
Paul Decker
President and Chief Executive Officer
Mathematica Policy Research

cc:
Peter Schochet, Senior Fellow, Mathematica Policy Research
Tamara Barnes, General Counsel, Mathematica Policy Research
Irma Perez-Johnson, Senior Researcher, Mathematica Policy Research
Alexandra Resch, Senior Researcher, Mathematica Policy Research
Jennifer de Vallance, Director of Public Affairs, Mathematica Policy Research
Shelly Martinez, Executive Director, Commission on Evidence-Based Policymaking
Nick Hart, Policy and Research Director, Commission on Evidence-Based Policymaking