Bipartisan Policy Center

EVIDENCE
WORKS

CASES WHERE EVIDENCE MEANINGFULLY INFORMED POLICY

EDITORS

Nick Hart
  Fellow
  Bipartisan Policy Center

Meron Yohannes
  Research Analyst
  Bipartisan Policy Center

ACKNOWLEDGMENTS

DISCLAIMER

RECOMMENDED CITATION

Nick Hart and Meron Yohannes (eds.) *Evidence Works: Cases Where Evidence Meaningfully Informed Policy.* (Washington, D.C.: Bipartisan Policy Center, 2019).

INTRODUCTION

FOREWORD

Katharine G. Abraham, former chair of the Commission on Evidence-Based Policymaking.

Ron Haskins, former co-chair of the Commission on Evidence-Based Policymaking.

In 2016 and 2017, we served as co-chairs of the bipartisan U.S. Commission on Evidence-Based Policymaking. Along with our fellow commissioners, we spent over a year studying the challenges in using data to produce evidence to serve government decision-makers.

Talking with and hearing from hundreds of people confirmed what many of us had believed for years—that all too often, there are steep barriers to producing the evidence needed to inform critical policy and program decisions. The good news is that there are solutions to the problems we heard about. Many of the recommendations our Commission made to improve the evidence-building process are now being implemented across government. Policymakers should increasingly have access to the information they need to make sound decisions.

On its own, access to information isn't enough. It's also essential that our elected officials and government executives apply available information to their decisions in realistic and meaningful ways. While much of our work focused on the process of producing research and evidence, the desired outcome is ultimately to have this information used.

This is not a partisan issue—Republicans and Democrats alike turn to the evidence that is available to inform their decisions and would welcome better evidence. The case studies in this volume illustrate this point well, providing illuminating examples in which good evidence has had a positive effect on policy debates and decisions.

Evidence Works: Cases Where Evidence Meaningfully Informed Policy is a timely extension of the work of the Commission on Evidence-Based Policymaking. It is our hope that providing policymakers and evidence advocates with a resource for demonstrating the power of rigourous evidence will help with making the case that investments in research matter. The many examples discussed in this volume in which robust evidence was successfully applied to improve decision-making should do just that.

Moving forward, as work to implement our Commission's recommendations continues, we would like to reaffirm the vision articulated in the Commission's final report for "a future in which rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy." Everything we see leads us to believe this vision can be realized.

INTRODUCTION

Why the Use of Evidence Matters

Nick Hart is a fellow at the Bipartisan Policy Center and chief executive officer of the Data Coalition. Previously, he served as the director of BPC's Evidence Project.

Meron Yohannes is a research analyst at the Bipartisan Policy Center.

The use of information in daily life to make decisions is common. Information is perceived to be easy to understand and useful. To purchase or rent a new home, for example, people use web listings to compare different properties and seek out the expertise of realtors for advice; to buy a vehicle, people search the internet to learn about car safety features or to research prices before visiting a dealer.

Decision-making about government policies is no different. It involves making decisions about how to allocate resources to meet goals. Those charged with effectively using taxpayers' dollars for the benefit of society need solid information to make decisions. They need information to fulfill the goals of programs and policies, and to ensure that actions align with expected results.

The concept of evidence-based policymaking has gained attention in the United States in recent years. In short, the idea is to use evidence to inform how policymakers go about reaching decisions. In this context, "evidence" is specifically defined as high-quality information constructed by systematically collecting data, analyzing data with rigorous research methods, then developing conclusions that are valid and reliable about groups of people, households, families, or organizations. This definition does not include hunches or haphazardly compiled information. The term "evidence" here also does not mean the same thing that it does in a courtroom. The data and methods for evidence-based policymaking are built on efforts to understand trends in performance, gain insights about how policies and programs operate, and acquire knowledge about impacts on outcomes and results. The evidence described here typically comes from statistical analysis, policy research, data science, or program evaluation.

Most of the American public will probably respond to the concept by asking, "Don't we already do this today?" or "If we don't do that, what *do* we do today?" The reality is that policymakers use a vast array of information to make decisions—case studies, constituent priorities, electoral consequences, personal values, and evidence. These inputs are important in a democratic society. But today, unfortunately, evidence is not always the priority—or even present in some policy debates.

One goal of the evidence-based policymaking movement is to encourage decision-makers to increasingly use rigorous evidence to guide their actions. Achieving this goal does not mean evidence will exclusively be how decisions are made. In fact, rarely can a portfolio of evidence itself dictate what decisions should be made. But evidence can reduce uncertainty about the effects of a decision, can help policymakers understand the range of benefits for different policy options, and can assure the American public that officials aren't basing decisions on faulty beliefs or misguided theories. In short, evidence can ensure that decisions rely on facts, truth, and reality.

A central tenet of evidence-based policymaking is that evidence should always have a seat at the table of inputs available for decision-makers. Proponents also conclude that it should have a prominent place at that table. It's not that policymakers don't want to make good decisions—they do. But policymakers may have different incentives for making decisions,

may not have access to needed information, or may have to simply interpret what is available to them, recognizing gaps and limitations. Often, policymakers must find evidence, translate existing evidence, and judge the credibility, validity, and reliability of evidence—all under the real-world time constraints of policy decisions and public administration.

In recent years, experts have focused a vast amount of attention on the supply of evidence, including how to reduce barriers to accessing data and strategies for enhancing the capacity to conduct policy research and program evaluation. At the same time, much evidence already exists, and the premise of building a supply is that the evidence will be useful to policymakers and, therefore, used.

But exclusive attention on supply poses challenges for the field. For example, even if we build a vast body of evidence on a given topic and make it available for use, but it is not used, there could conceivably be negative repercussions. One consequence could be that funders do not allocate future resources to building knowledge at all. In practice, this could very well be the greatest threat to the concept of evidence-based policymaking: that authorities perceive existing evidence to be unusable or see that it is going unused—meaning their decisions would come predominantly from other inputs. As a result, future evidence-building activities may be reduced or eliminated. If such a sequence of events took place, it would potentially undermine effective governance across-the-board.

To address this risk, policymakers will need to understand why evidence-based policymaking is worthwhile in the long run. The movement must prove itself, its value, and its usefulness.

The over-attention to supply issues in evidence-based policymaking also means that the strategies for transferring, disseminating, and sharing knowledge most effectively may be underdeveloped. The existence of evidence does not guarantee the usefulness, availability, style, or timeliness that policymakers require. There are many factors that can ultimately affect evidence use.

In fact, evidence-based policymaking itself should be, well, evidence-based. Policymakers can learn to be more effective at using evidence, generating the most useful evidence, and applying new evidence to calibrate, enhance, or improve decisions.

WHY DISCUSS SUCCESS STORIES OF EVIDENCE USE?

Policymaking occurs in an imperfect system. Human beings are fallible. They have limited capacity for knowledge, preexisting value systems, and a range of priorities. Even with perfect information, evidence-based policymaking doesn't guarantee that policymakers will choose the best solution to a problem; regardless, they rarely have the luxury of perfect evidence. But if the availability of evidence can be calibrated to match and fulfill the demand from policymakers who want to make the most of it, the benefits seem clear.

Successful use of evidence can lead to better decisions that are more likely to achieve public policy goals, to produce intended benefits for the American public, and to ensure that taxpayers' resources are being used as effectively as possible.

Does evidence really matter for decision-making? This statement can and should be evaluated.

Evidence Works is a compilation of case studies that demonstrates how evidence really did matter in reaching a policy decision. The text focuses on applying evidence explicitly to policies—as opposed to decision-making more generally, where evidence can inform decisions about individual practices or activities. Policymaking, however, is a distinct type of decision. It can include a group of practices but is ultimately characterized by the application of evidence to a government program as a legal or formalized action, such as an act of Congress, a regulation, or some other formal decision that applies broadly to a program's stakeholders, beneficiaries, or users.

The case studies included in *Evidence Works* were selected because they each tell a story of success—one that program administrators, policymakers, and elected officials can learn from as they apply the information at hand toward better decision-making. The case studies offer unique insights about the challenges that decision-makers face as they grapple with uncertainty, prioritize competing information inputs, and weigh the quality and credibility of available information. Policy decisions are varied and multifaceted—so too are policymakers' informational and evidentiary needs, which the case studies describe in detail. But success does not always mean that the solution preferred by policymakers based on the evidence came to fruition or was ultimately unchanged as further evidence came to light. Success comes in many forms, highlighting the different approaches to using evidence.

While the case studies cover a wide range of policy issues—they all have one common feature: some evidence existed before officials made a decision. The evidence used had to exist prior to the decision in order to inform the decision. In fact, if evidence hadn't existed already, these cases would be instances of policy-based evidence-making rather than evidence-based policymaking. That subtle distinction is what this volume seeks to capture: what was the evidence available to policymakers, and how did they use it to reach a decision? To the extent possible, some of the cases even delve into the issue of why, that is, the motivation for decision-makers in using evidence as a predominant driver of their decision.

As it turns out, these questions are difficult to answer. The government's deliberative decision-making processes (which can be something of a black box) are complex, and details about how decisions are made can be forgotten over time. The documentation for any given decision may be weak or nonexistent. In some cases, those most involved in the decision-making may be limited in what they can say publicly, even years after a decision is reached. Needless to say, there are a great many reasons more case studies are not currently available about the use of evidence in policymaking.

Given the complexities inherent to developing case studies, some of the cases in this volume delve deeper into the subject than others. The authors of each case study offer a summary of

their experience, then they discuss the availability of the evidence and how it was ultimately used. The cases vary in style. Some are academic, bureaucratic, technocratic, and even political. For the reader and potential user of these case studies, the variations in style, prose, and substantive detail also offer some insights and perspective about how different actors in the real-world policy system approach the issue of using evidence. The volume embraces these various styles to help bring the stories to life, many for the first time in the public domain.

The Bipartisan Policy Center did not compensate the authors for their contributions to the volume; each author generously volunteered time, energy, and experience. The authors of each case study participated out of a desire to tell the story and contribute to strengthening the field of evidence-based policymaking in the United States. Notably, some of the authors have conflicts of interest in writing the case studies, but their unique experiences also add to the rich context around evidence-based policymaking. Because these cases are from individualized experiences, they may not align with any specific political or ideological lens. BPC in no way endorses each author's characterizations or narratives, nor does BPC intend to suggest that any one approach explored by the authors is the exclusive way to address a policy issue or interpret the evidence. Similarly, the authors themselves do not necessarily endorse the perspectives of any other case study author. Each case study only reflects the perspective of its author or authors.

We identified cases for the volume through knowledge of the policy community; awareness of recent, relevant, and important evidentiary issues; and discussions with members of the evidence community. While we selected 20 case studies for this volume, we explored many more as possibilities—and with unlimited time and pages, thousands more could have been included. The cases included are not an exhaustive list; they are merely the tip of the iceberg.

Similarly, the project could have made an entire volume about cases where the evidence didn't work, where it wasn't available, and even where it was disregarded, untranslated, or unused. The project also could have mixed and matched successes with failures. Certainly, much can be learned from failure. But in the nascent world of evidence-based policymaking in the United States, the need to facilitate, nurture, and explain successful uses is paramount. Successes offer useful and constructive insights about what enabling conditions facilitated effective and meaningful use. A study of failures, while valuable, simply cannot offer the same insights.

THE EVIDENCE ABOUT THE USE OF RESEARCH EVIDENCE

Evidence Works is also a volume that contributes to the body of academic research about evidence use and what it means. Although there are existing gaps in use when it comes to policymaking, because uses are actually widespread, a great deal is already known about the use of research evidence today. One widely accepted typology to explain evidence use suggests numerous types: instrumental, conceptual, imposed, and political.[1]

Instrumental uses are those that lead to direct action resulting from research, a common expectation in society. For example, researchers or evaluators produce evidence, it directly provides a solution to policymakers' specific questions, and policymakers immediately take the recommended action such as increasing spending or revising a certain program activity. While instrumental uses are intuitively obvious, in practice, they are difficult to observe and realize. These uses are realistically challenged by the presumption that evidence will directly and completely answer policymakers' questions, even if incomplete information. This is not to say instrumental use is irrelevant, but to be useful, policymakers need additional contextual information.

Conceptual uses are those where users apply information to help define problems and outline the contours of potential solution sets or possibilities, recognizing the uncertainties of the public policy process. In practice, conceptual uses emerge when policymakers and decision-makers engage with information to define problems and clarify or frame their goals. This is often based on descriptive information, trend analysis, and other framing information found in implementation studies. Take for example, the unemployment rate in Cass County, Missouri, compared with the national unemployment rate over time. This basic trend analysis offers comparative information that a decision-maker can apply to understand whether a problem is relatively high or low compared with some norm.

While conceptual uses make it easier to understand programs and policies, these types of uses are not completely devoid of individual interpretations about whether an analysis is positive or negative. For example, if the unemployment rate is 4 percent, is that too high or too low? Such a judgment, while key for determining the use of the information, must be based on other inputs, values, and knowledge.

Imposed uses are those that drive actions based on evidence. Approaches to decision-making that use criteria that demonstrate a preference for evidence—even when from an unrelated context—fall within the imposition category. Imposed uses may emerge as mandates about applying some form of knowledge or even as direct methodological preferences in the creation of evidence such as "tiered-evidence" grant structures.

Political, or tactical, uses are those in which policymakers apply evidence to garner support for or opposition to a policy or issue. Practically speaking, political uses are widespread and viewed as legitimate by society, though often such uses are politically charged. Assertions of the misuse of evidence often also fall into this category. Misuse is when policymakers apply information to an argument in a manner that one political faction believes inaccurately reflects the evidence.

With political uses, policymakers can suggest the termination of programs that do not work, though no experimental evaluation alone can objectively justify resource reallocations. Political uses can also help policymakers to garner support, motivate constituencies, elicit buy-in from stakeholders, or even improve an argument for policy change. Political use may even be what drives an issue onto the policy agenda in the first place.

Taken together, each of these types of uses may be transparent and observable—or not. Much government decision-making occurs through deliberative processes that may include decisions not made public or otherwise disclosed. This issue has been one of the historical challenges for the research field examining evidence use—research relies on observable characteristics. When policy decisions are not observable, researchers may feel compelled to draw their own conclusions about the prevailing type of evidence use. This guesswork may even lead to erroneous suggestions about uses that did not occur.

Evidence use may also take the form of a mixture of the categories in the typology, because of the interconnected relationships and nature of decision-making processes. The same information could be political, conceptual, imposed, and instrumental. As a result, we consider the typology of uses only as a framework for understanding, not one to force or dictate a narrow lens about what use does or does not occur. What is a "good use" is still a values-based determination.

Rather than having an expectation of direct action, perhaps "good use" is simply a change in perspective—one that is paired with arguments about policy.

Conditions for research evidence use

Much has been written about evidence use regarding enabling and disabling conditions.[2] The enablers tend to focus on (1) the role of brokers to communicate and engage in knowledge sharing; (2) the capacity and infrastructure for individuals in organizations to prioritize evidence; (3) transparency mechanisms that foster a supply of information and public incentives to encourage accountability; and (4) trusted relationships between producers and users.[3]

Enabling conditions recognize that multiple factors work in concert to facilitate use, which encompasses dissemination, sharing, and system-based approaches.[4] In many ways, enabling conditions reflect ideas that may also be at the heart of effective policymaking, such as explaining concepts clearly, addressing context, maintaining credibility, promoting leadership, offering support, integrating systems, engaging stakeholders, and reflecting on lessons learned to improve.[5]

Sometimes unexpected characteristics can constrain evidence use—or at least add to the perception that the evidence is constrained. Overt partisanship through imposed political decision-making could be one such characteristic; it may appear as a lack of trust between policymakers. Within administrative agencies, litigation risks related to regulatory actions are also a potential reason for some limits to use.[6] There may also be negative incentives or benefits for programs and policies with unknown outcomes—particularly, if there's a perception that a policy is more effective than it actually is. Other barriers include clarity, relevance, reliability, timing, opportunities, costs, or even the skill sets and capabilities of the policymakers themselves.

Policymakers often perceive evidence as going unused. Gaps in the research-policy framework certainly exist. The William T. Grant Foundation, among others, has invested serious time and energy to identifying these gaps and building knowledge about strategies for better evidence use. Yet big questions remain, such as whose responsibility is it to foster use? The policymaker? The researcher? The American public? These questions are even more complicated when it comes to U.S. federated democratic governance, which has multiple policy systems and varying conditions that may affect use. Congressional decision-making processes differ from the executive branch agencies. State and local governments often must consider changes in federal laws and policies, which affect their decisions at the subnational level, often dictating certain limits or policy prescriptions that enable or constrain use.

The policy system is adaptive, interconnected, and highly complex. However, the use of research evidence is not only possible; it also actually happens to create effective public policy.

ORGANIZATION OF THE VOLUME

Evidence Works is a compiled volume of evidence use case studies that readers can either examine independently or in their totality. The goal is to provide a useful resource that informs practitioners and policymakers about how they might go about applying evidence to decisions, the relevance for a range of policy domains, and the prevalence of evidence use today.

The first section of case studies focuses on federal policies. Cases 1 to 4 focus on actions that required the involvement of policymakers in both Congress and the executive branch. Case 1 outlines how information acquired over decades about the Earned Income Tax Credit affected various proposals to expand the policy and offers insights about how this information influenced the 2017 tax reforms. Case 2 describes a major debate about the Social Security Disability Insurance program and how Democrats and Republicans viewed evidence in developing a consensus recommendation for program improvements and sustainability in 2015. Case 3 examines key reforms to the country's workforce and employment training provisions, which have been subject to impact evaluations for decades. Case 4, delves into the intricacies of the Family First Prevention Services Act, which offered the first major national reforms to the country's child welfare system in 40 years, shifting resources and attention from service delivery to preventative services.

Cases 5 to 9 consider major reforms undertaken in the executive branch or initiated by federal agencies based on decisions driven by political appointees or through formal regulatory processes. Case 5 describes an overhaul of the country's child-support enforcement regulations, applying evidence to ensure children not living with both of their parents were able to get the support they needed. Case 6 features a discussion of reforms the Department of Defense put in place to promote early education and childhood development centers at bases around the world for the benefit of families of military personnel. Case 7 outlines the process and decision points for a proposal to transform the national energy assistance subsidies for low-income households. Case 8 examines how the federal housing

department applied insights from a major study about different housing strategies to propose reshaping its budget. Case 9 describes how studying the 2009 health care reforms led to Medicare changes, generating substantial cost savings for American taxpayers.

Cases 10 to 12 involve policy decisions led by career program managers, key decision-makers within government agencies who can often make administrative changes to programs. These decisions occurred with support and backing from agency political appointees, but they were largely driven by managers. Case 10 presents a narrative about how policymakers used a compelling set of descriptive analyses to reform program measurement and ultimately improve the environmental clean-up activity associated with legacy contamination at many closed gas stations around the country. Case 11 covers a data analysis across agencies that led to stronger insights about how effectively public housing serves the American public, including by fostering improved public health outcomes. Case 12 describes how policymakers built evidence on food-assistance programs to ensure that payments to new mothers are most efficiently allocated in order to maximize the number of clients it serves.

The next section of case studies focuses on state and local government decisions. While the previous sections offer valuable lessons that may generalize to other contexts, state and local decision-makers are faced with different constraints and more localized constituencies. Case 13 highlights how Seattle and King County, Washington, developed a strategy to modify food-safety inspections and communications to the public. Case 14 describes the process of adopting an evidence-based policy, and modifying it, to reduce incidences of child neglect in Colorado. Case 15 demonstrates how policymakers applied evidence to improve early literacy programs and help children prepare for academic success in Tennessee. Case 16 explores an innovative approach to policing policies and community interactions in Washington, D.C., where officials assigned body cameras to the police force, highlighting what happens when evaluation findings do not validate preconceived notions.

The final section of case studies offers a completely different mix of stories about evidence use—they each include programs specifically designed to promote evidence use. Even using an evidentiary approach, policymakers still grappled with certain policy reforms. But each case offers a different strategic lens to encourage evidence use in policymaking, so that it ultimately affects practices and program activities. Cases 17 through 19 describe applications of a tiered funding framework to allocate resources based on the level of knowledge about an intervention and how it affects decision-making for the Teen Pregnancy Prevention Program, Nurse Family Partnerships and Home Visiting, and competitive education grants. A final case study delves into another approach that connects the demand for evidence with use, followed by policy action. Case 20 tells the story of how the creation of the U.S. Commission on Evidence-Based Policymaking produced evidence that reformed national policies about the use of data, program evaluation, and the federal statistical infrastructure.

The volume concludes with a brief cross-case analysis, teasing out major themes and lessons relevant when looking across the cases. The aim of the concluding chapter is to present useful information for policymakers, to encourage better use of evidence, and to advance

insights for the research community as it considers strategies and approaches that more effectively enable evidence-based policymaking in the future.

OUR VISION FOR MOVING FORWARD

In the Information Age, providing insights to make better decisions must be recognized as a prudent practice. But does the information make a difference? Does the evidence available for important decisions matter? Does it result in better decisions?

This volume of case studies explores questions about the use of evidence in a salient and accessible way. The cases in *Evidence Works* demonstrate how decision-makers can meaningfully apply and use evidence while addressing important policy questions.

By beginning to better tell successful stories of evidence use across a range of types of rigorous evidence and decision styles, we hope others will come forward to tell their stories as well. As a result, the field will increasingly become better at explaining not just how to go about using evidence to inform decisions, but also why the activity is far more common than many realize. As the knowledge about successful evidence use increases, policymakers will become better positioned to truly evaluate evidence-based policymaking and to determine whether the lofty ideal is achievable and worth the cost. ■

1   Carol H. Weiss. "Have We Learned Anything New About the Use of Evaluation?" *American Journal of Evaluation*, 19(1): 21-33, 1998.

2   D.H. Gitomer and K. Crouse. *Studying the Use of Research Evidence: A Review of Methods.* William T. Grant Foundation, 2019.
     Available at: http://wtgrantfoundation.org/studying-the-use-of-research-evidence-a-review-of-methods.

3   Nick Hart, Sandy Davis, and Tim Shaw. *Evidence Use in Congress: Challenges for Evidence-Based Policymaking.* Bipartisan Policy Center, 2018. Available at: https://bipartisanpolicy.org/library/evidence-in-congress-volume-1/.

4   Annette Boaz, Huw Davies, Alec Fraser, and Sandra Nutley. *What Works Now? Evidence-informed Policy and Practice* (Bristol, United Kingdom: Policy Press, 2019).

5   Ibid.

6   Nicholas Hart. *Evaluation at EPA: Determinants of the U.S. Environmental Protection Agency's Capacity to Supply Program Evaluation.* Dissertation, George Washington University, 2016.

USING
EVIDENCE
FOR FEDERAL
POLICYMAKING
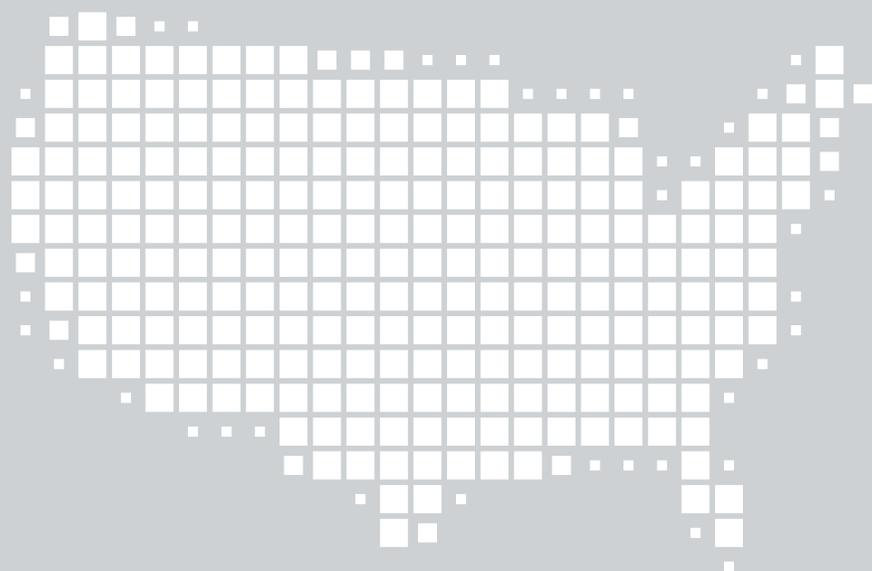
1. EARNED REPUTATION

The Earned Income Tax Credit

Ben Harris is the chief economist and senior advisor at Results for America. He also serves as the executive director of the Kellogg Public-Private Interface at Northwestern University's Kellogg School of Management. Harris was also the former chief economist and economic advisor to Vice President Joe Biden.

Maia Jachimowicz is the vice president for evidence-based policy implementation at Results for America. She previously served as director of policy in the Office of the Mayor for the City of Philadelphia, where she directed the research, design, and execution of public policy across a range of issues.
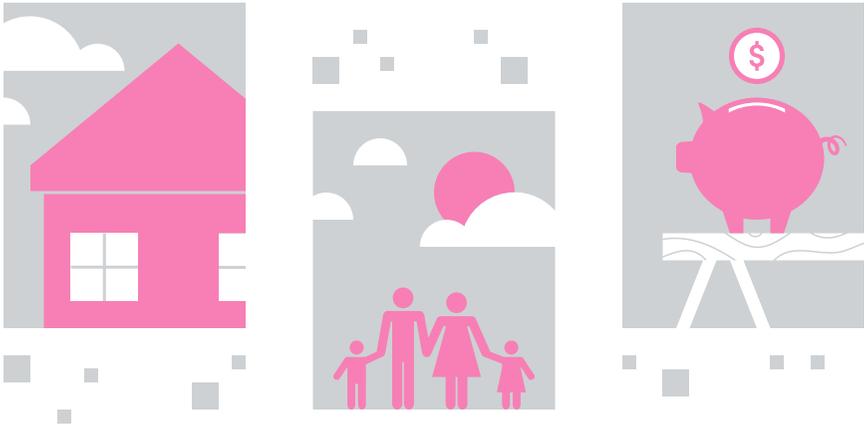
The Earned Income Tax Credit (EITC) is one of the premier pro-work and anti-poverty programs in the United States. Roughly 27 million low-income taxpayers receive the credit annually, with an average benefit of around $2,500. Since its introduction in 1975, dozens—if not hundreds—of researchers have evaluated the EITC, giving it a near-uniform consensus that it boosts labor supply, reduces poverty, and contributes to the long-term well-being of recipient families.

The EITC is a tax-based wage subsidy for low-income workers. Its value is based on a complicated series of factors, including household filing status, earnings, income, and number of children.[1] For each filing status and number of children, the credit features three separate income ranges: (1) a phase-in range, whereby the value of the credit grows with each additional dollar of earnings; (2) a plateau, under which taxpayers receive the maximum credit but don't earn an additional subsidy for more earnings; and (3) a phase-out range, whereby the value of the credit declines with higher earnings.[2]

While the fundamental structure of the EITC is the same for all eligible taxpayers, the particular subsidy rates, income ranges, and maximum credits differ substantially by filing status and number of children. Notably, the credit is worth very little to taxpayers without children, with a maximum credit in 2018 of just $519 for childless households.[3] Meanwhile, the credit to taxpayers with three or more children was as high as $6,431 in 2018.

A few other characteristics of the EITC bear mentioning. It is only available to workers ages 25 to 64 in large part to address concerns about the scope of the recipient pool; college students and Social Security beneficiaries, for example, may have temporarily low-earnings but may not be considered low-income. The credit is also refundable, meaning that taxpayers

with no or little income tax liability will still benefit. Like other tax credits, the EITC pays out annually, with taxpayers generally receiving benefits in late winter or early spring in the form of an income tax refund.

The credit is not without shortcomings. The EITC suffers from error rates of between 22 and 26 percent, which are high relative to other tax expenditures received by low- and middle-income households.[4] Also, the phase-out rate effectively creates an implicit tax on workers, which can have the opposite impact than intended for this group. The EITC is also complicated for many taxpayers, who often chose to use a paid-preparer—diminishing the net value of the credit. Lastly, the EITC's once-a-year delivery has been criticized by those who argue that it would be more effective if taxpayers could claim a portion of the credit throughout the year.

Perhaps due to the positive evaluations, the federal government has periodically expanded the EITC, with major expansions in 1994 and 2009. A host of other expansions have been proposed, including by the Barack Obama administration, Donald Trump's presidential campaign, and think tanks like the American Enterprise Institute, the Brookings Institution, and the Urban Institute. Despite bipartisan support for expansion, the credit was not directly changed in the 2017 tax bill that overhauled the individual income, corporate, and estate tax codes. And because that tax legislation slowed the rate of inflation for tax parameters, the value of the EITC will decline over time. To many observers, the lack of a 2017 EITC expansion was a curious and unfortunate omission given the sizable evidence of its positive impact and the Trump campaign's plan for a substantial EITC expansion.

Notwithstanding congressional actions in 2017, most policymakers recognize the EITC as one of the great applications of evidence to improve economic security. In particular, the EITC is an example of the potential power of evaluating the $1.6 trillion tax expenditure budget—which typically escapes scrutiny within the federal budget system. Indeed, introducing regular and formal evaluations to tax expenditures can markedly improve the application of trillions of dollars in tax expenditure spending to improve the lives of American families.

ISSUE BACKGROUND

In 1975, legislators passed the EITC as a response to a growing concern over the number of families receiving welfare benefits, which had risen from 1.1 million to 3.1 million over a few years. Senator Russell Long, concerned about the disincentive effects of unconditional cash transfers, proposed a "work bonus" plan that would provide a modest subsidy on low levels of earnings. The first iteration of the EITC was a nonrefundable credit equal to 10 percent of a low-income taxpayer's first $4,000 in earnings, with a maximum credit of $400 (or about $1,800 in 2018 dollars). The credit phased out between $4,000 and $8,000 in income. The credit was also justified as a way to help boost consumption in the face of the 1974 recession.[5]

The policy goals of the EITC have historically helped the tax credit to enjoy bipartisan support. The progressive structure of the program—which raises living standards by offsetting taxes based on a taxpayer's ability to pay—tends to appeal to Democrats. Meanwhile, Republicans (as well as many Democrats) tend to value the program's design to encourage and reward work. President Ronald Reagan called the program "the best anti-poverty, the best pro-family, the best job creation measure to come out of Congress."[6] President Barack Obama called for an expansion for childless workers in his 2016 State of the Union address: "America is about giving everybody willing to work a chance, a hand up, and I'd welcome a serious discussion about strategies we can all support, like expanding tax cuts for low-income workers who don't have children."[7] Today, politicians ranging from former House Speaker Paul Ryan to Senator Sherrod Brown support the credit's expansion to achieve certain policy goals.[8]

Policymakers expanded the EITC periodically between its inception and, most recently, in 2009. These expansions became permanent in 2015. In broad strokes, expansions included both increases to the value of the credit and to those eligible to receive it. At times, Congress also implemented eligibility restrictions, such as the requirement that only those with less than $3,500 in investment income could receive the credit. Notable expansions include increases to the maximum credit in 1978 and 1986; increases to the maximum credit overall, and for larger families in particular, in 1990 and 1993; and a higher credit for large families in 2009.

Various expansions have still left holes in the program, the most glaring of which is for childless workers—who receive a maximum benefit of about $500. There are other eligibility concerns, such as a lack of benefit for older and younger workers, and a lack of equity for one-child households. On the latter concern, for example, equalizing the one-child EITC with the two-child EITC in "equivalence scale-adjusted terms" would require boosting the credit for one-child families by about $1,000.[9]

Compliance and administration have also been a persistent concern, as noted above. Some conservative politicians and analysts cite a high rate of fraud in the program, including, for example, providing false income or filing status information.[10] Progressives often worry about the low rate of claiming among eligible households, with about 20 percent of eligible taxpayers choosing to forgo their credit. Policymakers and analysts across the ideological spectrum also cite concerns around complexity, including the high rate of unintentional errors in filing, the widespread need for paid preparers, and the subsequent relatively high levels of enforcement resources the IRS devotes to the EITC. Much of this administrative error is due to the difficulty taxpayers have claiming the correct number of dependents. For example, parents filing separate returns can make errors when indicating the number of children in a given household.[11]

EVIDENCE AVAILABILITY

The EITC is exceptionally well-evaluated. Dozens of high-quality studies in peer-reviewed journals have reached a near-consensus about the positive impacts of the EITC. Researchers and evaluators generally agree that the credit boosts labor supply (especially among single parents), lowers poverty rates, and contributes to the long-run well-being of households. Points of disagreement tend to center around the EITC's net cost in terms of lost revenue, drawbacks related to improper claiming, and, more fundamentally, whether it makes sense to address progressivity concerns through tax-based subsidies.

The plethora of studies devoted to EITC evaluation is due to a host of factors, including differential treatment among similar families (for example, two-child versus three-child families), periodic changes in the program parameters, and state-level variation—all of which allow researchers to identify the causal impacts of the program. In addition, widespread availability of microdata on the labor market and tax returns have enabled researchers to conduct several high-quality studies on the relationship between the EITC and changes in work patterns. In many cases, some of the most convincing and robust studies on the EITC arose from securely linking administratively collected tax records with confidential data from large, representative surveys.

The sheer number of studies makes it difficult in this short brief to summarize the key economic findings.[12] Some of the more important results include findings that the EITC significantly increases labor force participation among unmarried women, but can discourage work among married women.[13,14] The EITC has a positive impact on infant-birth indicators, such as low birth weight and premature birth, and maternal well-being, such as decreased risk factors for strokes and heart disease.[15] Vis-à-vis the long-term impacts of childhood exposure, the EITC significantly boosts high school and college completion rates, as well as the likelihood of employment in adulthood.[16] Also, a recent working paper estimates that 87 percent of the gross costs of the EITC are offset through higher tax receipts and reduced expenditures for public services.[17]

EVIDENCE USE

The EITC has enjoyed consistent support and periodic expansions over the past half-century. This is due to the program's broad, bipartisan appeal and its demonstrated positive and sustained impact on increasing labor force participation and decreasing poverty, among other outcomes of national interest. As noted above, the EITC has enjoyed regular expansions since its inception in the 1970s.

Due to the EITC's positive impact, documented through numerous rigorous evaluations, policy discussions about the EITC most often revolve around how to enhance its reach and impact. For example, an especially popular proposal to enhance EITC suggests addressing

the disparity in the value of credits for childless workers. In recent years, this approach has been endorsed by President Obama, former Speaker Ryan, and a bipartisan panel of experts commissioned by the Brookings Institution and the American Enterprise Institute to propose solutions to lack of opportunity in America. Other notable proposals include a 10 percent increase in the value of the credit, a $1,000 increase for single-child families, and a plan to make older workers eligible for the credit.[18,19,20]

More recently, in the absence of federal action, state and local policymakers found ways to expand the EITC's impact. There are many subnational efforts to provide state and local tax relief through state and local EITCs. To date, 29 states, the District of Columbia, Guam, Puerto Rico, and some municipalities adopted a state or local EITC credit. The first state EITC was enacted in 1986 in Rhode Island.[21] These state or local programs typically use federal eligibility rules, and policymakers structure them as some percentage of the federal credit, with five states offering nonrefundable credits and 24 states and Washington, D.C., offering refundable credits. The generosity of the credit varies substantially among states. For example, Maine provides a nonrefundable 5 percent state EITC in addition to the federal credit, while New Jersey offers a 37 percent refundable credit. In recent years, many states have been expanding their EITC programs, with plans that range from higher credits to increased access, to switching from a nonrefundable to refundable credit.[22]

The connection between the EITC and the use of evidence is not all positive. First, for example, in contrast to direct and mandatory spending, the $1.6 trillion tax expenditure budget—including the EITC—is subject to almost no formal review.[23] Thus, while evaluators have provided extensive evidence on the merits of the credit, there is limited knowledge about how it interacts with other tax expenditures or how it compares with different tax-based approaches for achieving social objectives. Second, the recent Tax Cuts and Jobs Act of 2017 failed to expand the EITC, despite having granted $2 trillion in net cuts to taxpayers—including the expansion of many provisions that have limited or no evaluations. Legislators prioritized many other approaches with few or no evaluations over the EITC, leading many observers and advocates to question why policymakers did not include the credit's expansion—especially given the credit's well-documented impact on labor supply in a time of widespread labor shortages.

In sum, the EITC has enjoyed regular bipartisan support throughout its near half-century of existence. It has been expanded at the federal level in every decade since its creation. It has been adopted by the majority of the states. And several policymakers have proposed expansions in the program to address shortcomings with the U.S. economy. These regular and widespread expansions coincide with decades of positive evaluations of the program's impact and cost-effectiveness. While it is impossible to state how the EITC may have fared in the absence of existing evaluations, it appears to be one of the great success stories in evidence-based policy.

LESSONS

- ***Evidence of impact sustained the program.*** A wide body of rigorous evidence demonstrating the positive impact of the EITC on labor force participation, reducing poverty, improving maternal and infant health, and enhancing adult earnings has helped sustain and expand the program over the past 40 years. The broad and consensus research around the EITC appears to have helped propel its federal and state expansion and led to many calling for more resources devoted to the credit. In particular, policymakers from across the aisle often agree that they should sustain, an even expand, the EITC to enhance credits for childless workers.

- ***Bipartisanship affects perceptions.*** The broad bipartisan support of the EITC's policy goals—encouraging and rewarding work while raising living standards for the poorest workers—also contributes to the program's popularity.

- ***Continuous and consistent evaluation is needed.*** The lack of formal mechanisms for evaluating tax expenditures is a shortcoming that fails to recognize the EITC as more effective than other expenditures. Evaluations should be formalized into a more coherent and comprehensive review of the $1.6 trillion tax expenditure budget.

- ***Evidence may not change politics.*** The omission of an EITC expansion in the Tax Cuts and Jobs Act of 2017 represents a failure of Congress to use existing evidence to make informed decisions about tax expenditures, as legislators prioritized other tax cuts with few or no evaluations over the EITC. ■

---

1   For an overview of the mechanics of the EITC, *see*: Gene Falk and Margot L. Crandall-Hollick. *The Earned Income Tax Credit (EITC): An Overview.* Congressional Research Service, R43805, April 18, 2018.

2   To illustrate, consider the value of the credit for taxpayers filing as "heads of households" with three or more children. In 2018, the phase-in range for this household is between $0 and $14,290 in annual income, with these taxpayers receiving an additional wage subsidy of 45 percent on additional earnings. The plateau is between $14,290 and $18,660 in annual income, with taxpayers receiving the maximum credit of $6,431 in this income range. The phase-out range is between $18,660 and $49,194 in annual income, with taxpayers in this range losing 21.06 percent of each dollar earned to a lower credit.

3   The maximum credit is $3,461 for households with one child and $5,716 for households with two children.

4   Importantly, note that these error rates largely reflect taxpayer confusion and misunderstanding, and should not wholly be characterized as "fraud." For a detailed discussion, *see*: Robert Greenstein, John Wancheck, and Chuck Marr. *Reducing Overpayments in the Earned Income Tax Credit.* Center on Budget and Policy Priorities, January 31, 2019. Available at: https://www.cbpp.org/research/federal-tax/reducing-overpayments-in-the-earned-income-tax-credit.

5   Margot L. Crandall-Hollick and Joseph S. Hughes. The Earned Income Tax Credit (EITC): An Economic Analysis, U.S. Library of Congress. Congressional Research Service, R44057, 2018.

6   Chuck Marr. "Reagan's Actions Made Him a True EITC Champion." *Off the Charts*, August 1, 2014.
     Available at: https://www.cbpp.org/blog/reagans-actions-made-him-a-true-eitc-champion.

7   Chuck Marr. "Obama Wisely Calls for Boosting the EITC for Childless Adults." *Off the Charts*, January 13, 2016.
     Available at: https://www.cbpp.org/blog/obama-wisely-calls-for-boosting-the-eitc-for-childless-adults.

8   Chuck Marr, Emily Horton, and Brendan Duke. "Brown-Khanna Proposal to Expand EITC Would Raise Incomes of 47 Million Working Households." *Off the Charts*, October 10, 2017. Available at: https://www.cbpp.org/research/federal-tax/brown-khanna-proposal-to-expand-eitc-would-raise-incomes-of-47-million-working.

9   Hilary W. Hoynes. "Building on the Success of the Earned Income Tax Credit." The Hamilton Project, The Brookings Institution, June 19, 2014. Available at: https://www.brookings.edu/research/building-on-the-success-of-the-earned-income-tax-credit/. As Hoynes explains, equivalence adjustments allow for the comparison of standards of living between multi-person households and single-person reference households. In other words, they allow for comparisons in the standard of living among households of different sizes.

10   Chris Edwards and Veronique de Rugy. "Earned Income Tax Credit: Small Benefits, Large Costs." The Cato Institute. *Tax and Budget Bulletin* (73), October 14, 2015. Available at: https://www.cato.org/publications/tax-budget-bulletin/earned-income-tax-credit-small-benefits-large-costs.

11   Janet Holtzblatt and Janet McCubbin. "Whose Child Is It Anyway? Simplifying the Definition of a Child." *National Tax Journal*, 56(3): 701-718, September 2003.

12   For an excellent summary of the evidence on the EITC, *see*: V. Holz and J.K. Scholz. "The Earned Income Tax Credit." In: R. Moffitt (ed.). *Means-Tested Transfer Programmes in the United States* (Chicago: University of Chicago Press, 2003). *See also:* Austin Nichols and Jesse Rothstein. *The Earned Income Tax Credit (EITC).* National Bureau of Economic Research, 2015. Available at: https://www.nber.org/papers/w21211.pdf.

13   Nada Eissa and Jeffrey B. Liebman. "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics*, 11(2): 605-637, May 1996.

14   Nada Eissa and Hilary W. Hoynes. "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit." *Journal of Public Economics*, 88: 1931-1958, August 2004.

15   William N. Evans and Craig L. Garthwaite. "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health." *American Economic Journal: Economic Policy*, 6(2): 258-290, May 2014.

16   Jacob E. Bastian and Katherine Michelmore. "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes." *Journal of Labor Economics*, 36(4): 1127-1163, October 2018.

17   Jacob E. Bastian and Maggie R. Jones. "Do EITC Expansions Pay for Themselves? Effects on Tax Revenue and Public Assistance Spending." Working paper, March 15, 2019.
     Available at: https://drive.google.com/file/d/1GbBeeQzfGH9fF9Y1u5rS55Sn3eStBWE7/view.

18   Hilary W. Hoynes, Jesse Rothstein, and Krista Ruffini. *Making Work Pay Better Through an Expanded Earned Income Tax Credit.* The Hamilton Project, The Brookings Institution, October 2017. Available at: https://gspp.berkeley.edu/assets/uploads/research/pdf/making_work_pay_expanded_eitc_Hoynes_Rothstein_Ruffini.pdf.

19   Hilary W. Hoynes. "Building on the Success of the Earned Income Tax Credit." The Hamilton Project, The Brookings Institution, June 19, 2014.
     Available at: https://www.brookings.edu/research/building-on-the-success-of-the-earned-income-tax-credit/.

20   Alicia H. Munnell and Abigail N. Waters. "Keeping Older People in the Labor Force." Presentation: "Improving Opportunities for Those Who Want to Work Longer" event at The Brookings Institution, Washington, D.C., January 24, 2019.
     Available at: https://www.brookings.edu/wp-content/uploads/2018/12/3-Munnell-Walters.pdf.

21   Olivia Sutphen. "The State of the Earned Income Tax Credit." National Conference of State Legislatures. *LegisBrief*, 26(3): 1-2, 2018.
     Available at: http://www.ncsl.org/LinkClick.aspx?fileticket=UxTWaDvKN7Y%3d&tabid=32022&portalid=1.

22   Samantha Waxman. "State EITC Expansions Will Help Millions of Workers and Their Families." *Off the Charts*, July 31, 2018.
     Available at: https://www.cbpp.org/blog/state-eitc-expansions-will-help-millions-of-workers-and-their-families.

23   Benjamin H. Harris, Eugene Steuerle, and Caleb Quakenbush. *Evaluating Tax Expenditures: Introducing Oversight into Spending through the Tax Code.* Urban Institute-Brookings Institution Tax Policy Center, July 10, 2018. Available at: https://www.taxpolicycenter.org/publications/evaluating-tax-expenditures.

2. DISABILITY POLICY

Saving Disability Insurance with the First Reforms in a Generation

Ted McCann is a fellow at the Bipartisan Policy Center. He was previously the assistant to the speaker for policy, managing welfare, education, workforce, pension, telecommunications, and Social Security issues for former House Speaker Paul Ryan.

Nick Hart is a fellow at the Bipartisan Policy Center and chief executive officer of the Data Coalition. Previously, he served as the director of BPC's Evidence Project. He also previously worked at the White House Office of Management and Budget covering Social Security and anti-poverty issues.
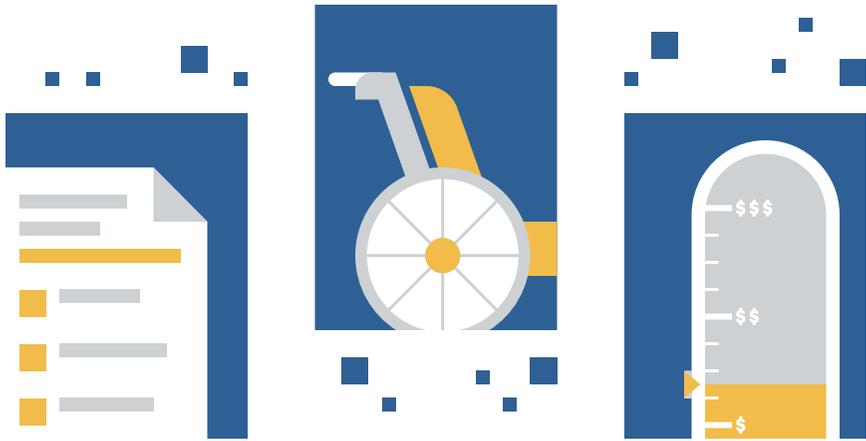
For 150 million workers in the American labor force, the Social Security Disability Insurance (SSDI) Program offers a critical backstop that provides income support if a worker cannot remain employed due to a medical impairment. Through payroll taxes, the American workforce contributes to this government insurance system that mitigates the consequences of long-term disability on employment. As of 2019, the SSDI program has 11 million beneficiaries and costs $140 billion per year, and it remains available as a form of insurance for the millions of workers in the contemporary American labor force. Given its size and scope, the program is often the center of major policy debates about the federal budget and how to best deliver anti-poverty services while encouraging employment.

ISSUE BACKGROUND

In 2016, for the first time since 1983, one of the trust funds that support the Social Security program was facing a shortfall in funding. The shortfall could have prevented the timely and complete payment of anticipated benefits. Actuaries projected that the SSDI Trust Fund would run out of money by the end of the year. If that had happened, the program would be solely funded each year based on the revenues generated, a challenge given the demographics of the program suggest it needed more funding in the near-term. By some estimates, if Congress and the president had not addressed the funding shortfall, benefits would have been reduced by about 20 percent.[1]

Several key factors contributed to the funding shortfall, chief among them that the program now covered more women in the labor force. However, there are other factors that added to increased costs, such as improper payments, high administrative costs, and a lack of innovation. With numerous critics of the SSDI program from across the political spectrum,

the run-up to 2016 was an auspicious time to look toward SSDI reform—especially given the need to address the funding crisis while balancing partisan goals, improving administrative effectiveness by reducing improper payments, and strengthening the overall program structure.

Congress took the first step by sending a signal about how it would address the funding shortfall. In January 2015, the U.S. House of Representatives included a directive in the rules that guide the legislative chamber's operations requiring the body to meet certain conditions. Many members of Congress did not understand that Social Security is composed of two distinct trust funds—one that covers the retirement program (Old-Age and Survivors Insurance) and one that covers disability insurance. The House passed a set of rules that prevented simply transferring money from the retirement trust fund to the disability trust fund, without also including policies that would improve long-term balance between revenues and expenditures.

This technical requirement added to the House rules was a marker set by Republicans; they wanted to achieve efficiencies in program expenditures in exchange for addressing the funding solvency. It was also a partisan maneuver by Republicans, opposed by many Democrats, that created a backlash from disability groups who opposed larger reforms to the program that could result in reduced payments for beneficiaries. The action created a commitment mechanism for House Republicans that would require them to make a serious legislative push toward making reforms to Social Security that would have a meaningful fiscal effect.

Following the House rule, the White House weighed in on a preferred approach to addressing the funding shortfall as part of its annual budget submission to Congress. The president's fiscal year 2016 budget proposal declared that the federal government needed a politically acceptable plan to address the funding shortfall within the current fiscal year. The administration proposed a temporary fix, transferring money from another part of Social Security to fill in the gap.[2] The proposal recognized the approach that Congress had used previously to address funding shortfalls. The White House's budget proposal also included several small reforms to modify the program that would generate budget savings in line with the expectation of the initial congressional actions on the topic.

Both Republicans and Democrats in Congress as well as the executive branch acknowledged the necessity and time sensitivity for addressing the funding shortfall, which was a signal of consensus on the core problem. But how to go about fixing the shortfall required additional dialogue.

Partisan goals for reform

Not surprisingly, Republicans and Democrats had overlapping and diverging goals in considering the financial sustainability of disability insurance. In the House, then-Social Security Subcommittee Chairman Sam Johnson laid out four goals that Republicans wanted to achieve when reforming the SSDI program:

1. Ensure benefits continue to be paid to individuals with disabilities and their family members who also rely on them;

2. Prevent a 20 percent across-the-board benefit cut;

3. Make the SSDI program work better; and

4. Promote opportunity for those trying to return to work.

These goals provided little in the way of specificity, but House Republicans proposed multiple bills that strategically advanced the position. Additionally, at the beginning of the 114th Congress, House Republicans made it explicit in the rules package that improving the financial outlook for the combined Old-Age and Survivors Insurance and SSDI trust fund was a goal of theirs as well.

On the other side of the aisle, congressional Democrats and the White House also had well-defined goals:

1. Ensure that the SSDI program remains on solid financial footing; and

2. Prevent any unnecessary programmatic changes that may harm individuals with disabilities or their families.

Improper payments

Improper payments arise when individuals are either paid too much or too little based on their eligibility and history of earnings. From the beneficiary's perspective, the SSDI program's rules and reporting requirements can be confusing, and it is not clear how large benefits should be in a given time period. From a government perspective, there are clear rules used to calculate benefits–rules that beneficiaries must legally follow.

Both Republican and Democratic staff of the House Ways and Means Committee had long expressed concerns about the high level of improper payments to beneficiaries on the SSDI program.[3] As a consequence, and due to a policy of tracking such payments, the Social Security Administration could determine the general causes that contributed to a lack of knowledge about real-time earnings from beneficiaries and administrative computational errors that affected calculated benefits.[4]

The practical and political consequences of improper payments are significant. If a beneficiary receives overpayments for an extended period, it is possible he or she could receive tens of thousands of dollars in improper benefits that were not legally permissible. For a beneficiary with limited earnings or assets, finding out that he or she owes tens of thousands of dollars back to the federal government presents a severe financial burden for that person and his or her family.

Politically, some Republicans thought that overpayments could also discourage beneficiaries from attempting to return to work. The work-incentive rules in the program are incredibly complex, and it can be difficult for an individual to understand how and when he or she can work. Against the backdrop of possible financial repercussions for overpayments, SSDI beneficiaries often decide that it is not worthwhile to return to work.[5]

Program structure

As noted above, the rules for implementing the SSDI program are complex. Congressional Republicans perceived that the structure of the SSDI program discouraged beneficiaries from returning to work. One rule, for example, establishes a period of time in which beneficiaries can begin to work—the trial work period. During this period, beneficiaries begin receiving some incentives when they cross an $850-a-month threshold; for a defined time period, the program does not reduce disability benefits to compensate for newly earned income. A separate rule establishes a threshold at which the program deems an individual back in the workforce for "substantial gainful employment" and therefore no longer eligible to receive SSDI benefits. This level begins at $1,080 of earnings per month.

There are two main reasons why Republicans thought that the trial work period and the substantial gainful employment rules discouraged work. First, individuals with disabilities face a difficult time obtaining and sustaining costly health insurance; the SSDI benefits can obviate the cost of health insurance because the program's benefits are eligible for Medicare. Second, some qualitative information suggests that beneficiaries are concerned they will be assessed overpayments if they work too much.[6] Thus, Republicans said that these issues, combined with the complexities of return-to-work incentive programs, drove down beneficiaries' work participation rates.

Democrats also long acknowledged the challenges with returning to work for program beneficiaries, but that knowledge was relatively limited about how to encourage such approaches. Both Republicans and Democrats agreed that improving recipients' work participation and reducing overpayments were laudable goals. However, there was disagreement over how best to achieve those goals.

EVIDENCE AVAILABILITY

With clear political goals and the issues well-framed for consideration, the White House and Congress could consider the body of evidence to determine a consensus-based approach to move forward in addressing the SSDI program's future. Fortunately, to support these policy discussions, the Social Security Administration had a vast amount of administrative data, or information collected through the course of implementing the program, that could be readily analyzed because they were well-organized and high-quality data. The agency also funded a series of demonstration projects to test out new ideas and potential policy options for consideration over the prior decade, though with mixed results.

Evidence about program characteristics and improper payments

All this capability was available to support the political and policy debates about potential reforms. Much of the overarching program background and framing was summarized in a report released by the White House in July 2015, with the goal of focusing public and political attention on the need to address the funding shortfall.[7] The descriptive statistics and trends provided in the report about the program's beneficiaries were intended to characterize how the program works, who the program serves, and why benefits only address a portion of beneficiaries' pre-disability earnings.[8] The report also identified that perceptions of fraud were an issue the program faces, though in practice a rare occurrence.[9]

Evidence about program incentives

In Congress, committee staff spent six months working with the Social Security Administration to develop a better understanding of the characteristics of individuals earnings and the challenges they faced, as well as the administrative challenges to reducing complexity and improving incentives. During this time, congressional and agency staff analyzed administrative data and developed new insights, including that overpayments in a related program—Supplemental Security Income (SSI)—were typically much smaller, even though SSI program has more beneficiaries. SSI beneficiaries also typically saw higher work participation rates than the SSDI program. Based on the information at hand, Republican staff concluded that the difference in work rates was likely due to a combination of beneficiaries' fear of losing health insurance, fear of overpayments, and a sharp disincentive to work embedded in rules of the SSDI program.

Looking at the same data analysis, Democratic staff in Congress generally argued that earnings reported by program participants were an accurate reflection of work capabilities. This perspective was not necessarily consistent across the Democratic field, however. A former economist at the White House noted several years prior that "labor supply disincentives are inherent in any transfer program with imperfect screening for need."[10] In other words, without perfect information about how to assess individual needs of beneficiaries in the workforce, there will likely be cases where some beneficiaries could reasonably return to work.

Past demonstration projects to test innovations

Parts of this collective knowledge and the perceptions about incentives were based on studies that Congress mandated in 1999, called Ticket to Work and the Benefit Offset National Demonstration. Both projects have been criticized for being unsuccessful at achieving the goal of helping beneficiaries return to work.[11]

Ticket to Work aimed to provide some program beneficiaries with the ability to join a network for employment supports and to have third-party providers support efforts to improve earnings and employment. Study results were disappointing as most who were eligible never sought to use the services; thus, it had little impact on SSDI outcomes of interest.

The Benefit Offset National Demonstration sought to test whether providing beneficiaries with better information about benefits in addition to slight changes in the rules would increase the likelihood of individuals returning to work. The initial results from the study were underwhelming and suggested the additional services did not increase earnings for beneficiaries.[12]

A separate study looking at mental health disorders and their effects on returning to work; concluded that beneficiaries often have insufficient access to services and treatments. Thus, the conditions themselves impede returning to work.[13] But when those with mental health disorders do return to work, beneficiaries are more likely to have earnings, though not statistically more likely to leave the program.[14] At the time, long-term follow-up information was not available from that study, leaving a vast amount of uncertainty about whether the odds of employment truly improved for the longer time frame policymakers were interested in understanding.

This vast array of information, including past experimental evaluations, descriptive statistics, and on-demand data-analysis capabilities, was available to support policymakers in 2015 to develop a consensus solution to address the broader funding dilemma for the SSDI program.

EVIDENCE USE

When Congress and the White House were considering the SSDI reforms as part of the Bipartisan Budget Act of 2015, the application of evidence focused on four key areas: (1) clarifying program rules, (2) incentivizing beneficiaries to return to work, (3) ensuring funding was available to sustain current benefits, and (4) filling knowledge gaps for future reforms. Evidence served as a foundation for Republican and Democratic staff in bridging political divides about developing a solution for proposed disability reforms.

Applying evidence to reduce overpayments

Evidence in the form of descriptive statistics and trend analysis provided by the Social Security Administration offered insights about the need for strategies to reduce overpayments. Both Republican and Democratic staff in Congress, noting the relative complexity of the SSDI program versus SSI, agreed that some simplifications in rules were well justified.

Making the program administratively simpler and more straightforward for the beneficiary to understand was one strategy to reduce overpayments based on lessons learned in other programs. Congressional policymakers translated this goal into a strategy to modify how the program counted wages to determine benefit payments. Prior to the Bipartisan Budget Act of 2015, the program counted wages at the point in time when they were *earned*, not when they were paid. Changing the default for how this calculation occurred meant beneficiaries and caseworkers could use paychecks to validate and verify information.

EVIDENCE SERVED AS A FOUNDATION FOR REPUBLICAN AND DEMOCRATIC STAFF IN BRIDGING POLITICAL DIVIDES ABOUT DEVELOPING A SOLUTION FOR PROPOSED DISABILITY REFORMS.

Another change to reduce the administrative complexity of the program was a new approach for either estimating or imputing wage data online. The SSI program already used this feature, which was not previously allowed for the SSDI program. By making the small change to the point in time at which earnings are determined (earned versus paid), the estimation of earnings was also possible. The need for this small policy reform was informed by acknowledged gaps in data quality from self-reported data, which had to be addressed in order to reduce potential inadvertent overpayments.

Encouraging beneficiaries to return to work

Despite past research and demonstration projects aimed at most effectively encouraging beneficiaries to return to work, policymakers faced an array of uncertainty about how to change the rules to do so. This was an area of strong disagreement between Republicans and Democrats, especially when it came to acknowledging how much the administrative program disincentivized work through the "cliff effect."

Without sufficient information about how to inform this decision, policymakers still had to take the best action possible. Republican congressional staff favored an approach called a "benefit offset," which credits beneficiaries for a certain period or for a certain portion of their earnings, ensuring that if they return to work, the program does not immediately deem them ineligible. Democratic staff perceived a benefit offset could unfairly harm individuals who were working more than the offset, and therefore it would have some negative consequences for current beneficiaries. Using the analysis provided by the Social Security Administration, congressional staff concluded the typical wage for a working individual on disability insurance was one that maintained a critical threshold.

Congressional staff on both sides of the aisle viewed this analysis as vindication for their respective views. For Republicans, the result of the analysis suggested individuals were keeping earnings just below this threshold because the beneficiaries did not understand the work incentives, were afraid of receiving overpayments, or did not want to risk losing health insurance provided by the program. Separately, Democratic staff concluded that the threshold was below "substantial gainful employment," and, therefore, individuals were clearly working at their full capacity.

Despite the disagreements about how to interpret the information, both Republicans and Democrats agreed about the need to ensure individuals were able to work at their full capacity. In this case, instead of modifying the rules with vast uncertainty, policymakers opted to test different approaches to learn more about what effect such a benefit offset proposal would have on individuals. The law included direction for the Social Security Administration to develop and launch a targeted pilot project and evaluation to inform future efforts to reform the program.

Reallocating funding to sustain benefits

Since the original issue at hand, which instigated the need for program reform, was a disparity between revenue and expenditures, Congress and the White House quickly agreed that fulfilling the reallocation of funds from one part of Social Security to another was desirable and unavoidable in the near-term. This goal was, in part, fulfilled because of the availability of an extensive amount of information about program participant characteristics made available by the agency, stakeholders, and even White House documents.[15] For example, one analysis highlighted the number of beneficiaries in each state, their average earnings prior to entering the program, and the amount they received in insurance payments as a way to highlight the geographic distribution of the participants in the program as well as the fact that individuals receive only a portion of their pre-program earnings through the insurance payments.[16]

With the framing about the need for benefits well established, policymakers quickly agreed on a financing solution that was contingent on the other reforms. The consensus reallocation was based on the proposal developed by the White House and was consistent with the requirements built into the controversial House rules package.[17]

Preparing for future reforms

Finally, policymakers in both the White House and Congress recognized the need to continue learning about how to improve the program. While the Social Security Administration conducted past demonstrations and evaluations, the agency's ability to consistently test new ideas was limited based on the way the program's law was written.[18] Accepting a proposal from the White House—which was a priority for the Obama administration in negotiating reforms—policymakers agreed to renew this ability to test new ideas, albeit only for a limited period.[19]

While not discussed here, the Bipartisan Budget Act of 2015 also included numerous other reforms for different parts of Social Security that were similarly informed by available evidence or program insights from relevant analyses. These additional reforms included changing medical-review requirements for cases, expanding fraud investigation capabilities, changing the rules about certain credible sources of medical records to justify claims, increasing penalties for fraud or errors, enabling data matching to electronic payroll providers, and closing certain loopholes in the programs. While the available evidence was not perfect, the important policy debate highlights both that policymakers are eager to use the best information available and that they must make decisions despite uncertainty, decisions that aim to achieve political goals and program improvements.

LESSONS

- *Analysis of administrative records was foundational.* At multiple points in the policy debates about how to proceed in decision-making, analysis of existing administrative data was incredibly informative in helping policymakers understand the starting point and even the bounds of potential reforms. Without these data collected by the government, policymakers would remain uncertain about how to develop solutions and how to agree on them.

- *The lack of impact evaluations limited action.* Because the Social Security Administration had limited success in completing impact evaluations for the SSDI program and in determining various interventions or reforms for consideration, policymakers were constrained in how much they could act. While some could view this as a beneficial feature for those who argue against any program change, the lack of information also prevented bipartisan consensus on potential reforms that could have drastically improved beneficiaries' quality of life.

- *Existing evidence did not eliminate political value systems.* As the example of the program incentives policy change highlights, when faced with uncertainty policymakers still applied political value systems when making a final decision. In this case, they decided not to take action until further information was available because of the political disagreement. But the value systems also enabled and supported the generation of new knowledge with the goal of using that evidence in future reforms.

- *Staff served as critical decision-makers in support of elected or appointed policymakers.* During the policy discussions for potential reforms, congressional staff and senior appointees at the White House and the Social Security Administration facilitated the dialogue about reforms and engaged in the effort to use evidence. Their collective advice resulted in recommendations to elected members of Congress or the president in determining whether to pass and sign the Bipartisan Budget Act of 2015.

- *The use of evidence was iterative and varied.* Analysts developed some evidence during the debates about policy reforms, rather than relying solely on evidence that existed prior to the discussions. Because of this, policymakers had a portfolio of evidence available to them that included statistics, performance metrics, implementation studies, and some impact evaluations. ◼

1   White House. *Social Security Disability Insurance: A Lifeline for Millions of American Workers and Their Families.* Office of Management and Budget and the National Economic Council, 2015.
     Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/ssdi_national_report_7-17-2015.pdf.

2   White House. *President's Fiscal Year 2016 Budget, Appendix: Social Security Administration.* Office of Management and Budget, 2016.

3   U.S. Congress, House of Representatives, Committee on Ways and Means, Subcommittee on Social Security. *The Financial Risk of Returning to Work: Hearing before the Subcommittee on Social Security.* 114th Congress, 1st session, 2015.

4   Social Security Administration. "Reducing Improper Payments."
     Available at: https://www.ssa.gov/improperpayments/RSDI_majorCauses.html#sb=2.

5   Social Security Administration. "Welcome to the Work Site." Available at: https://www.ssa.gov/work/.

6   U.S. Congress, House of Representatives, Committee on Ways and Means, Subcommittee on Social Security. *The Financial Risk of Returning to Work: Hearing before the Subcommittee on Social Security.* 114th Congress, 1st session, 2015.

7   White House. *Social Security Disability Insurance: A Lifeline for Millions of American Workers and Their Families.* Office of Management and Budget and the National Economic Council, 2015.
     Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/ssdi_national_report_7-17-2015.pdf.

8   Shaun Donovan and Jeffrey Zients. "Social Security Disability Insurance: A Lifeline for Millions of American Workers and Their Families." *The White House*, July 17, 2015. Available at: https://obamawhitehouse.archives.gov/blog/2015/07/17/social-security-disability-insurance-lifeline-millions-american-workers-and-their-fa.

9   Social Security Administration. *Follow-up Report on Anti-Fraud Efforts Requested at the January 16, 2014 Hearing on Disability Fraud.* 2014.
     Available at: http://ssa.gov/legislation/SSA_Report_to_Chairman_Johnson_2_14_2014.pdf.

10   Jeffrey B. Liebman and Jack A. Smalligan. *Proposal 4: An Evidence-Based Path to Disability Insurance Reform.* The Hamilton Project, The Brookings Institution, February 2013. Available at: http://www.hamiltonproject.org/assets/legacy/files/downloads_and_links/THP_15WaysFedBudget_Prop4.pdf.

11   Jason J. Fichtner and Jason S. Seligman. *Saving Social Security Disability Insurance: Designing and Testing Reforms Through Demonstration Projects.* Mercatus Center, George Mason University, 2018. Available at: https://www.mercatus.org/system/files/fichtner-social-security-disability-mercatus-working-paper-v1.pdf.

12   Social Security Administration. "Benefit Offset National Demonstration."
     Available at: https://www.ssa.gov/disabilityresearch/offsetnational.htm.

13   William D. Frey, Robert E. Drake, Gary R. Bond, et al. *Mental Health Treatment Study: Final Report.* Westat, July 2011. Available at: https://www.ssa.gov/disabilityresearch/documents/MHTS_Final_Report_508.pdf.

14   Julia Baller, Haiyi Xie, Gary Bond, et al. "Long-Term Follow-Up of the Mental Health Treatment Study." Presented at the 5th Annual Meeting of the Disability Research Consortium, Washington, D.C., August 2, 2017. Available at: https://www.mathematica-mpr.com/-/media/internet/conferences/2017/drc-annual-meeting-2017/summaries/baller.pdf?la=en.

15   For example, see statistics by congressional district released by the SSA in 2014 that were available for the 2015 debate: Social Security Administration. "Congressional Statistics, 2014."
     Available at: https://www.ssa.gov/policy/docs/factsheets/cong_stats/2014/index.html.

16   White House. *Social Security Disability Insurance: A Lifeline for Millions of American Workers and Their Families.* Office of Management and Budget and the National Economic Council, 2015.
     Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/ssdi_national_report_7-17-2015.pdf.

17   *See: Bipartisan Budget Act of 2015*, Sec. 833. Public Law 114-74, 2015.

18   Nick Hart, Jason Fichtner, and Jack Smalligan. "Improving Social Security Disability Insurance: Building a Culture of Innovation and Experimentation to Identify Future Reforms." *Benefits Quarterly*, First Quarter: 26-32, 2019. Available at: https://www.iscebs.org/Resources/BQ/Pages/BQ-executive-summaries-2019.aspx.

19   White House. *President's Fiscal Year 2016 Budget, Appendix: Social Security Administration.* Office of Management and Budget, 2015.
     Available at: https://www.govinfo.gov/content/pkg/BUDGET-2016-APP/pdf/BUDGET-2016-APP-1-30.pdf.

3. TRAINING POLICY AT THE ONSET OF THE GREAT RECESSION

Too Important to Let Evidence Intercede

Peter Mueser is the Frederick A. Middlebush Chair in Economics and economics professor at Department of Economics and Truman School of Public Affairs, University of Missouri, Columbia.

Kenneth Troske is the Richard W. and Janis H. First Endowed Chair of Economics at the University of Kentucky. He served as a member of the U.S. Commission on Evidence-Based Policymaking.

Policymakers' twin goals of assuring that the skills of the workforce align with the needs of the modern economy and that workers with skill deficits receive support have resulted in a long line of laws that seek to provide worker training. Modern federal support for job training dates to the 1960s, when President John F. Kennedy's administration introduced the Manpower Development and Training Act as part of its anti-poverty program.[1]

Enacted in 1998 at a time of near-full employment, the Workforce Investment Act (WIA) replaced the 1982 Job Training Partnership Act (JTPA) incorporating the welfare-reform agenda that was dominant at the time. Responding to the perceived need to modernize job training, WIA introduced a new system of centralized employment and job-training centers, established a process for combining job-search and job-training activities, used training vouchers to allow participants greater choice among providers, and included significant changes in the governance structures overseeing training at the state and local levels. Two WIA programs served adults: the Adult Program served those with poor work histories, and the Dislocated Worker Program served recently unemployed workers.

ISSUE BACKGROUND

WIA directed the U.S. Department of Labor to undertake at least one multisite evaluation of WIA by the end of fiscal year 2005. The law directed that WIA undertake an experimental evaluation.[2] Although several studies of WIA were undertaken in the law's early years, none of these studies met the experimental evaluation requirement.

In 2005, the White House Office of Management and Budget (OMB) assigned the WIA program a rating of "adequate" from its Program Assessment Rating Tool. But WIA received

low marks for its evaluation efforts, again because the program lacked an experimental evaluation, a point that was stressed in meetings between the OMB and the U.S. Labor Department.[3,4] In response, in 2007, the Labor Department released plans for two evaluations of WIA: (1) an experimental evaluation, based on a random assignment of eligible participants either to training or to a control group, that cost over $20 million with final results released over a decade later and (2) a second evaluation of WIA with a non-experimental design that used existing program data, receiving over $1 million in funding and having results available in 2008.

The nonexperimental evaluation used data for participants in the WIA Adult and Dislocated Worker programs in 12 states from 2003 to 2005, to evaluate impacts on employment and earnings for up to four years following program entry. The comparison group consisted of either those receiving job-search services in state offices or those receiving Unemployment Insurance benefits, depending on the state. The study employed statistical techniques to adjust for differences in characteristics and prior employment between participants and the comparison groups. The nonexperimental evaluation results became available in December 2008—just as interest in job training rose in response to the plight of unemployed workers, whose numbers were spiking with the onset of the recession.

The experimental evaluation collected data from individuals who applied to receive WIA services and who were deemed eligible at 28 local area offices between November 2011 and April 2013. The study then assigned applicants to one of three groups—(1) individuals who received full WIA services (including possibly job training); (2) individuals who only received the core and intensive services, which provided counseling and other individualized services; and (3) individuals who received only core services, comparable to job-search assistance freely available at state job centers. Analysts estimated treatment effects by comparing the average outcomes of individuals in the three groups. Final results from the experimental study became available in December 2018, more than 11 years after Labor Department initially announced plans for the study, 10 years after the release of results from the nonexperimental study, and four years after WIA was replaced by the Workforce Innovation and Opportunity Act.

EVIDENCE AVAILABILITY

Until the 1980s, evaluations of job-training programs relied primarily on nonexperimental methods. A widely cited 1978 review examining the effects of training programs made almost no reference to experimental studies, as the few studies that had used random-assignment methods were not viewed as directly applicable to existing large-scale programs.[5] A significant shift occurred with the publication of a 1986 study, which showed that results from an experimental, random-assignment, job-training study could not be obtained by using nonexperimental methods.[6] These conclusions would be challenged in subsequent decades as some observers perceived that only random-assignment methods could lead to causal findings from program evaluations.

Responding to this view, the federal government chose to fund an experimental, random-assignment evaluation of the law that preceded WIA in the 1980s, the JTPA. That evaluation focused on disadvantaged workers—those with unstable work histories and low earnings. The study found that job training had statistically significant but modest effects on participants' earnings.[7] The researchers did not attempt to consider job training for dislocated workers or for individuals who became unemployed after extended periods of stable employment.

Since the publication of the 1986 study and after publication of the JTPA evaluation results, researchers debated the validity of the dichotomy between experimental and nonexperimental evaluations. Critics of social experiments argue that randomized experiments have limited usefulness in policymaking.[8,9] In 1997, researchers published a set of papers using data from experiments to identify strategies that might be successful in estimating program impacts from nonexperimental data.[10,11,12]

A growing number of nonexperimental studies in Europe, covering a variety of programs providing both job-search and job-training assistance, took advantage of exceedingly detailed administrative data. In an international meta-analysis of training program evaluations in 2009, researchers found that longer-term job-training programs tended to have small or negative impacts on employment or earnings in periods of less than a year but that impacts often turned positive in the second or third years.[13] They also concluded that "research designs used in recent nonexperimental evaluations are not significantly biased relative to the benchmark of an experimental design."[14] A 2006 meta-analysis reached a similar conclusion.[15]

Before 2008, the Labor Department had overseen several studies of WIA. Two studies focused primarily on implementation.[16,17] In 2005, researchers examined labor-market outcomes in seven states for WIA participants who had completed the program in the period from July 2000 to June 2002, but data limitations and the focus on the early years of program implementation (WIA was adopted in 2000 in most states) raised questions about the validity of reported estimates.[18]

In 2008, the best estimate of the impacts of the WIA program were based on the JTPA experimental study, supported by nonexperimental findings implying modest positive effects of job-training programs across several countries. Given that the JTPA study focused exclusively on disadvantaged workers, conclusions for other unemployed workers would have been more uncertain. Although both JTPA and WIA provided similar training, the passage of two decades and a variety of administrative differences between the programs suggested that there could be substantial differences in program efficacy.

EVIDENCE USE

In 2008, researchers submitted the final report for the WIA nonexperimental evaluation to the Labor Department. The impact of the study on policy, at least in the short run, was quite limited. In large part, policymakers cherry-picked results to support activities that they wished to pursue for reasons unrelated to the study's findings. Some critics who exclusively

favored experimental methods questioned the nonexperimental design, which made it easier for those who disagreed with the study's conclusions to dismiss it. Despite the limited short-term policy effects, the study received positive attention from researchers, and it has arguably added to the weight of evidence that may be relevant in the long run.

The study's results suggested that by two to three years after program entry, participants in the WIA Adult Program experienced increased expected earnings of 15 to 26 percent, due in large part to increased employment. For participants in the WIA Dislocated Worker Program, the report concluded that the benefits of participation were either small or nonexistent and that any returns were unlikely to compensate for time spent in training.

Several months before completion of the final evaluation report, the Labor Department asked the Coalition for Evidence-Based Policy, now part of Arnold Ventures, to comment on the study's design and preliminary results. The coalition expressed doubt as to whether the nonexperimental methods would be successful in providing valid estimates of program impact. Although the study's authors disputed the basis for their conclusions, noting that their criticisms would imply the rejection of any feasible nonexperimental evaluation, the Labor Department invited the coalition to provide a review of the final report. Looking back, Labor Department staffer Jonathan Simonetta says he believes that these reviews caused policymakers to downplay the study's results.

Immediately after the study was issued at the end of 2008, the Obama administration took office amid a financial crisis that would signal the onset of the worst recession since the 1930s. Within a month of assuming the presidency, in January 2009, President Obama had signed into law the American Recovery and Reinvestment Act, which allocated $3.45 billion for additional job training, including $500 million for the WIA Adult Program and $1.25 billion for the WIA Dislocated Worker Program. The political commitment to provide retraining to unemployed workers was apparently central to the administration's policy, and it seems unlikely that the report influenced the administration's position.[19] The Economic Report of the President cited the study as supporting the efficacy of WIA, without distinguishing between the Adult and Dislocated Worker programs.[20]

Although the direct impact of the study on federal policy was likely minimal, a variety of news outlets cited the study, including *The New York Times* and *The Atlanta Journal-Constitution*.[21,22] In academic policy circles, the report appears to have received some attention. Several times in 2009, the study's authors made presentations based on the report, including to the Recovery and Reemployment Research Conference, set up by the Labor Department, and to a meeting with the European Commission in Washington. The study's authors summarized the results of the report in a chapter of a book based on this last conference.[23] In March 2011, a congressional staffer contacted one of the authors of the study, indicating that a pre-publication version of the book had received the attention of those preparing for a congressional hearing. The study's authors provided a summary of the findings for the hearing.

Following the submission of the report, the authors undertook several additional analyses. The revised paper was published in 2013.[24] The study found a place among serious job-training discussions, and especially in discussions of the Workforce Innovation and Opportunity Act, the law that replaced WIA. The original study also received a prominent place in a 2011 review of WIA research and job training.[25] Recent studies have tended to support the findings of the report. A subsequent nonexperimental study of the Trade Adjustment Assistance Program, which served unemployed workers similar to those served by the WIA Dislocated Worker Program, also found that participants gained little in terms of employment or earnings from the initiative.[26]

Researchers finally released results from the experimental WIA evaluation in 2018, and they were less informative than anticipated due to unforeseen problems with the study's design. The study found that those individuals assigned to receive intensive services experienced improvements in employment outcomes 30 months after assignment to the program. In contrast, assignment to the training treatment was not associated with any employment benefit. However, because those assigned to the training treatment were only somewhat more likely to obtain training than those in the control groups, estimates of the impact of training are uncertain. Although the effects of the Adult and Dislocated Worker programs were not statistically different, the power to distinguish differences was quite limited.[27]

LESSONS

- *Researchers' intended use of evidence isn't guaranteed.* Results from the nonexperimental study were relevant for the development of policy implementation at the time, but, in the face of political exigencies, they had little immediate role. In the longer run, however, the results undoubtedly contributed to the general understanding of the efficacy of training programs. When analyses of the Trade Adjustment Assistance Program showed little impact, this confirmed the difficulties of retraining experienced workers. In part, researchers must accept that their work will contribute to general knowledge, which will be of value in framing policy decisions in the long run. In retrospect, in comparison with the experimental study, the nonexperimental study was not only much less expensive but was at least as useful in contributing to this knowledge.

- *Research questions should be specific and direct.* To inform those who are implementing a program that their program may be of little value does not provide them with actionable information. What is most useful is information about what kinds of activities are most likely to be successful in achieving the desired outcomes.

- *Do certain groups of people benefit more from particular programs?* Are particular approaches more likely to work? To have a substantial immediate policy impact, studies must frame research questions to respond to issues faced by the users. In making most decisions, the relative expense of alternatives weighs heavily, yet cost information is seldom of sufficient quality or granularity for the analyst to make useful inferences. Program administrators must collect cost information more consistently and make it available for research purposes.

- *When federal funding is involved, states should be required to provide data.* Research is not possible without detailed and high-quality data. Researchers contacted all 50 states with a request for data, but ultimately only 12 participated. While the sample was sufficient to undertake the analysis, the issue of whether the omitted states were systemically different from those included reduces the generalizability of the conclusions. Given that the federal government funds WIA and other similar programs, policymakers could pass legislation requiring states to provide these data. The burden of providing these data is relatively modest, and the failure to do so merely reflects the lack of meaningful incentives faced by the state. ■

1   *Manpower Development and Training Act of 1962.* Public Law 87-415, 1962.

2   *Workforce Investment Act of 1998.* Public Law 105-220, 1998.

3   U.S. Government Accountability Office. *Performance and Accountability Report, Fiscal Year 2008.* GAO-09-1SP, November 14, 2008.

4   Jonathan Simonetta, who oversaw research activities in the Labor Department's Employment and Training Administration, observed that OMB staff repeatedly referenced the department's failure to satisfy WIA evaluation requirements.

5   Orley Ashenfelter. "Estimating the Effect of Training Programs on Earnings." *The Review of Economics and Statistics*, 60(1): 47, 1978.

6   Robert J. LaLonde. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review*, 76(4): 604-620, 1986.

7   Larry L. Orr, Howard S. Bloom, Stephen H. Bell, et al. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study* (Washington, D.C.: The Urban Institute Press, 1996).

8   James J. Heckman and Jeffrey A. Smith. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives*, 9(2): 85-85, 1995.

9   James J. Heckman, Robert J. LaLonde, and Jeffery A. Smith. "The Economics and Econometrics of Active Labor Market Programs." In: Orley Ashenfelter and David E. Card (eds.). *Handbook of Labor Economics, Vol. 3* (Amsterdam: Elsevier, 1999), 1865-2097.

10 James J. Heckman, Hidehiko Ichimura, and Petra E. Todd. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies*, 64(4): 605-654, 1997.

11 James J. Heckman, Hidehiko Ichimura, and Petra E. Todd. "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* 65(2), 261-294, 1998.

12 James J. Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. "Characterizing Selection Bias Using Experimental Data." *Econometrica*, 66(5): 1017-1098, September 1998.

13 David Card, Jochen Kluve, and Andrea Weber. "Active Labor Market Policy Evaluations: A Meta-Analysis." *IZA*, Discussion Paper No. 4002, February 2009.

14 Ibid., 26

15 David Greenberg, Charles Michaelopoulos, and Philip Robin "Do Experimental and Nonexperimental Evaluations Give Different Answers about the Effectiveness of Government-Funded Training Programs?" *Journal of Policy Analysis and Management*, 25(3): 523-552, 2006.

16 Burt S. Barnow and Christopher T. King. *The Workforce Investment Act in Eight States*. U.S. Department of Labor, ETA Occasional Paper, 2005.

17 Social Policy Research Associates and TATC Consulting. *The Workforce Investment Act after Five Year: Results from the National Evaluation of the Implementation of WIA*. U.S. Department of Labor, Employment and Training Administration, June 2004.
    Available at: https://www.doleta.gov/reports/searcheta/occ/papers/spr-wia_final_report.pdf.

18 Kevin Hollenbeck, Daniel Schroeder, Christopher T. King, and Wei-Jang Huang. *Net Impact Estimates For Services Provided Through the Workforce Investment Act*. U.S. Department of Labor, Employment and Training Administration, October 2005.

19 U.S. Department of Labor officials repeatedly defended federal support for job training—especially for dislocated workers. "'These programs are really working,' said the assistant secretary of labor, Jane Oates. 'These are folks who clearly want to go back to work and we're able to help them get back to work. The investment in job training is one that's not only going to pay off in the short term, it's going to help us be more competitive in the long term.'" *See*: Peter S. Goodman. "After Training, Still Scrambling for Employment." *The New York Times*, July 18, 2010. Available at: https://www.nytimes.com/2010/07/19/business/19training.html.

20 "A recent study found that Workforce Investment Act training programs for adults boosted employment and earnings, on average." *See*: Council of Economic Advisers. *Economic Report of the President*. February 2010, 229. Available at: https://www.govinfo.gov/content/pkg/ERP-2010/pdf/ERP-2010.pdf.

21 Peter S. Goodman. "After Training, Still Scrambling for Employment." *The New York Times*, July 18, 2010. Available at: https://www.nytimes.com/2010/07/19/business/19training.html.

22 Dan Chapman. "Sharper Skills Slow Cut to Joblessness." *The Atlanta Journal-Constitution*, July 26, 2010. Available at: https://www.ajc.com/business/sharper-skills-slow-cut-joblessness/sncHYM6PX6wsuh6KpeX6MJ/.

23 Douglas J. Besharov and Phoebe H Cottingham (eds.). *The Workforce Investment Act: Implementation and Experiences and Evaluation Findings* (Kalamazoo, MI: W.E. Upjhon Institute for Employment Research, 2011).

24 Carolyn J. Heinrich, Peter R. Mueser, Kenneth R Troske, et al. "Do Public Employment and Training Programs Work?" *IZA Journal of Labor Economics*, 2(6): 1-23, 2013.

25 Paul T. Decker. "Ten Years of WIA Research." In: Douglas J. Besharov and Phoebe H. Cottingham (eds.). *The Workforce Investment Act: Implementation and Experiences and Evaluation Findings* (Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2011), 315-346.

26 Peter Z. Schochet, Ronald D'Amico, Jillian Berk, et al. *Estimated Impacts for Participants in the Trade Adjustment Assistance (TAA) Program Under the 2002 Amendments*. Final Report, U.S. Department of Labor, August 2012. Available at: https://wdr.doleta.gov/research/FullText_Documents/ETAOP_2013_10_Participant_Impact_Report.pdf.

27 Kenneth Fortson, Dana Rotz, Paul Burkander, et al. *Providing Public Workforce Services to Job Seekers: 30-month Impact Findings on the WIA Adult and Dislocated Worker Programs*. Mathematica Policy Research, May 2017. Available at: https://www.mathematica-mpr.com/our-publications-and-findings/publications/providing-public-workforce-services-to-job-seekers-30-month-impact-findings-on-the-wia-adult.

4. FAMILY FIRST

Funding Practices that Keep Kids Safe

Kira Fatherree is a fiscal and program management advisor for the Office of Analytics, Technology, and Planning at the Allegheny County, Pennsylvania, Department of Human Services. She was formerly a program examiner for the White House Office of Management and Budget.
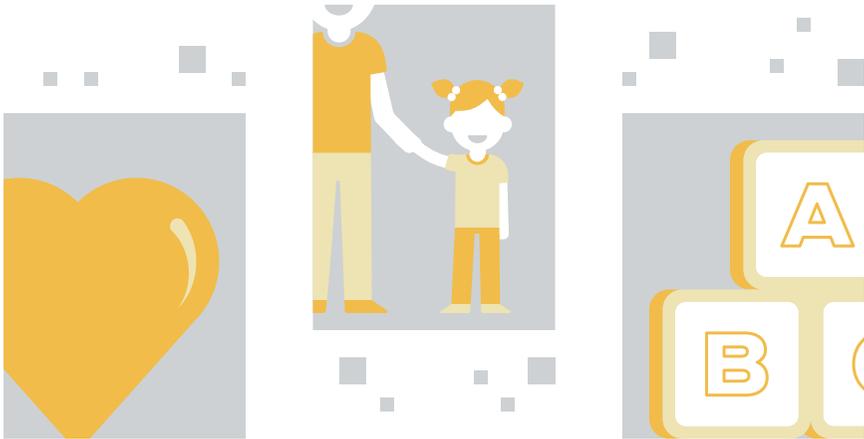
In 2017, there were over 4 million allegations of child abuse and neglect made to public child welfare agencies concerning more than 7 million children nationwide.[1] The fact that nearly 10 percent of all children in the United States were subject to such allegations poses a serious public health and public-policy concern. Of the children found to be maltreated, the highest percentages suffered from neglect (75 percent) and physical abuse (18 percent). Furthermore, emphasizing the seriousness and urgency of this issue, more than 1,700 children died from abuse and neglect nationwide in 2017.[2]

Public child welfare agencies administer an extensive set of programs intended to protect children from harm, as they are charged with ensuring their safety, permanency, and well-being. While partially funded by the federal government, public child welfare agencies are operated by individual state and/or county governments and receive and investigate reports of child abuse and neglect. These agencies decide how to respond to and investigate allegations, and if a child must be removed from the home. Ideally, if safety issues exist or factors indicate a risk of harm, the family can receive targeted services and support, while enabling the child to stay at home safely.

Child welfare agencies strive only to remove children from their families when absolutely necessary because the experience can be traumatic for children and parents, can lead to family instability, and can be a strain on the child welfare system. Decades of research has shown that children who must be removed from their family of origin, do best when they grow up with at least one loving caretaker.[3] Unfortunately, removal is necessary for some situations to protect children from harm. If removed, a child can be placed in settings

ranging from kinship care, where the caregiver is a relative or person known to the child; a nonrelative foster home, which is a family-like environment; or a residential care program (which includes group homes and institutional care). While federal law requires that children are placed in the least restrictive, most family-like setting available, child welfare agencies are constrained by availability of foster parents as well as the behavioral and physical health needs of a child.

Broadly, the child welfare system refers to a continuum of services "designed to ensure that children are safe and that families have the necessary support to care for their children successfully."[4] In 2018, recognizing opportunities for improvements to the federal government's role in this system, federal policymakers enacted the Family First Prevention Services Act, making the most expansive changes to this program in 40 years by one piece of legislation.[5] An active advocacy community played an instrumental role in the enactment as groups pleaded with and pressed Congress to better align federal funding with what has been shown to best serve children and families. While not a perfect funding realignment, the law presents a fundamental shift and leverages evidence for the benefit of vulnerable children.

ISSUE BACKGROUND

Implementing child welfare services nationwide requires substantial resources, totaling $30 billion in 2016.[6] Over half of these costs were covered by state and local governments, while the federal government contributed slightly less than half.[7] However, the amount that the federal government covered varies by state, ranging from 17 percent in Delaware to 78 percent in Louisiana.[8] These wide variations in funding speak to the differences of child welfare systems in each state and to the total amount of funding—mostly based on the level of the state investment—and they have important implications for system equity, the array of services available, and the approach and effectiveness of service delivery across states.

The largest federal resource for states in the child welfare continuum is foster care, paid for largely under Title IV-E of the Social Security Act. Foster care (also known as out-of-home care) is a temporary arrangement where children are placed with a relative or nonrelative caregiver, or in another type of residential setting, and protected against harm. The amount of federal funds provided to states for foster care is driven by claims for reimbursement. States can submit claims to the federal government for the eligible portion of the placement costs when a child from a qualified low-income household is removed from their home and placed into an out-of-home placement: a relative or nonrelative foster family home, a group home, an institution, or another setting. The federal government provided about $6 billion to state governments in 2018 for foster-care-related reimbursements. The federal government gives financial payments to state governments only after a child is removed from their home. This creates a perverse financial incentive for states that reinforces a practice that does not serve children best. The rate of federal funds spent on removals relative to efforts that prevent the need for removals is eight to one.[9]

With limited federal funding, states must bear the burden of funding family prevention and intervention services to prevent the need for foster care placement when services can make a difference. Only a limited amount of federal funds (approximately $700 million annually) is available for prevention services, which constrains efforts to strengthen families and to provide services early enough to meet families' needs and avert a crisis from occurring.

Making the federal financing of foster care more complicated and nuanced, federal law dictates that states are only eligible to receive foster care (title IV-E) reimbursement for children from families who would have met the poverty standards in the Aid to Families with Dependent Children program, precisely as it existed in 1996. Based on how the federal foster care law was written, the poverty measure that determines eligibility has remained static since 1996. For example, to qualify for the family income portion of the criteria, a family's income is limited to $674 per month for a family of three in Wyoming and $1,082 per month for a family of three in Florida, and a family must have less than $10,000 in resources available to them.[10,11] Without any increases in the threshold for over 20 years, the number of families with monthly incomes below the old standards has continuously decreased. With fewer families meeting the criteria, the federal government provided fewer resources to states based on the funding formula. While some people would advocate removing the old poverty standards, Congress has made clear that doing so would be prohibitively expensive given federal funding priorities. Therefore, policymakers have only adjusted federal foster care funding incrementally and on the margins.

Congress and stakeholders are well aware of the inadequacies of the federal role in the child welfare system. To combat the lack of funding for innovative prevention approaches and the focus on out-of-home placements, Congress strategically enabled states to do more with the limited foster care federal funds through Title IV-E Child Welfare Demonstration Projects. Building off legislation from 1994 that first created waivers to increase the flexibility of federal foster care funds, the Child and Family Services Improvement and Innovation Act of 2011 resurrected and modified child welfare demonstration waivers, adding a required innovative practice and evaluation component.[12] The waivers were "designed to improve state effectiveness in protecting children and assisting families as well as point the way to potential broader national reforms to benefit children and families in the coming years."[13] This authority relinquished the stringent federal foster care eligibility requirements and required jurisdictions to test new approaches with the same amount of funding that they would have received absent the waiver.[14]

Twenty-eight jurisdictions, including the District of Columbia, and a tribal organization, took advantage of the opportunity to operate under the waiver authority.[15] However, in the early and mid-2010s, as the legal authority for waivers neared expiration, and without a new law in place to carry the flexibilities forward, Congress faced more pressure to produce a more permanent child welfare finance reform and to do so in a bipartisan way. Waivers were a temporary fix for some of the problems of the federal financing structure and included

an expiration date of September 30, 2019, to force Congress to act. As conditions ripened for reform, members of Congress identified key issues, and staff began to draft legislation highlighting how federal reform could tackle them, including:

1. The lack of adequate funding for family-strengthening efforts that prevent the need for foster care and the traumatic experience of home removal;

2. The investment in out-of-home placement settings that fund group home care and does not serve children best; and

3. How federal funding was being used for some practices and services that were ineffective in accomplishing key outcomes of safety, permanency, and well-being.

In particular, the interventions were in some cases shown to not routinely produce positive outcomes or to prevent maltreatment and entry into foster care. Some practices were identified as actually harmful for children, like group home settings when the child did not have needs that required the level of care provided in that type of environment and stayed for an inappropriate amount of time.[16,17,18]

EVIDENCE AVAILABILITY

The child welfare field became more aware of the trauma experienced by children coming into the system. With this knowledge, the field began to build a trauma-informed system that strives not to exacerbate or add to traumatic experiences. Thus, decision-makers in the executive branch as well as members of Congress started to pay attention to child welfare issues when they increasingly heard about children being placed into group settings. There was a consensus in the research community that group settings should be used only when alternatives were not available. In 2014, leading researchers released a consensus statement to that effect.[19] Given that federal funds are one source that state and local jurisdictions use to pay for the high costs of group home settings, a change in federal funding structure was a lever the federal government was able to use to impact state behavior as it relates to placement options.

Aligned with the heightened interest and to better understand the use of group home placements, the U.S. Department of Health and Human Services (HHS) conducted a nationwide observational analysis using administrative records. There is a federal requirement for states to report on a variety of indicators through the Adoption and Foster Care Analysis and Reporting System, which in turn provides data that can generate useful insights about program operations and the children served. Because the child welfare system is decentralized and operated at the state and local levels, the federal data reporting requirements proved beneficial for HHS staff to analyze the information about trends and then make policy recommendations based on findings.

In 2015, HHS's observational analysis concluded that group home placements were being overused as a foster care placement, especially for youth under 12 years old where family-like settings are most appropriate. The study noted that group home care is appropriate and necessary when used for specialized cases of behavioral or mental health needs, and should be used to temporarily stabilize youth to prepare them for a return to a family setting when possible.[20] The HHS analysis also found that:

- Children spend an average of eight months in group care settings, with a broad range of state averages;

- Children ages 12 and younger comprised an unexpectedly high percentage (31 percent) of children who experienced a congregate care setting, with the point in time ranging from 6 percent to 69 percent; and

- Half of the children ages 13 and older in foster care entered a group home setting at some point, with a quarter entering a group home as their first placement and more than 40 percent entering due to a child behavior problem and no other clinical or mental disability.

The HHS report provided national statistics that finally suggested what many observed anecdotally and gave a straightforward conclusion that group homes were being overused.

Compounding those findings were other reports of the peril and inadequacy of foster care for youth who age out of the system (between ages 18 and 21 depending on the state), meaning they leave foster care without a permanent home. A 1999 federal law established a data system known as the National Youth Transition Database; it required states to follow specific cohorts of youth over time to collect additional outcome information to enable a collective assessment. Of the foster youth who age out of the system, at age 21, only half reported having full or part-time employment, 20 percent reported being incarcerated in the past two years, only 3 percent reported earning a college degree, and a quarter reported having given birth or fathered a child within the past two years.[21] While there are of course success stories, on the whole, the data indicate how the foster care system is not preparing youth who age out with tools for positive life outcomes and youth who age out of group home settings have poorer outcomes than their peers in family-like settings.[22,23]

The National Youth Transition Database outcome statistics give further credibility to the longstanding knowledge that children do best with their family and in family-like settings. Multiple research studies found that children in foster care "experience feelings of confusion, fear, apprehension of the unknown, loss, sadness, anxiety, and stress."[24] Moreover, while child welfare is intended to keep children safe, the process of removal "unintentionally increased the vulnerability of this already fragile population."[25] The findings on foster care experiences adversely impacting children added to the urgency of preventing foster care as an option; a landmark study conducted from 1995 to 1997 found a

relationship between the number of adverse childhood experiences a person experienced and a variety of negative outcomes in adulthood.[26] Negative outcomes included, for example, physical and mental health impairments, substance use disorders, and risky behaviors. The research regarding adverse childhood experiences and the not only short-term but lifelong negative impacts emphasized the importance and need for prevention efforts.

EVIDENCE USE

When Congress and the executive branch were determining how to best reform the federal policies related to child welfare, policymakers leaned on the array of existing research to inform the changes included in the Family First Prevention Services Act (FFPSA). These reforms included key actions related to improving and increasing prevention services, reducing the use of group homes, and identifying more high-quality foster families.

Increasing prevention services

Given the trauma and negative outcomes generally associated with abuse, neglect, and foster care placements, preventing any maltreatment (primary prevention) and preventing subsequent maltreatment from occurring (secondary prevention) are critical for an effective child welfare system. The presence of interventions that can successfully mitigate harm is fundamental to a robust child welfare system that values prevention. Prevention interventions that focus on reducing risk factors and strengthening families can vary, targeting parenting skills, child behavior issues, substance use disorders, and mental health issues. Child welfare agencies look to needs assessments of their population and evaluations of these activities to understand the proper service array to offer in their local context.

Given the prevalence of evidence that prevention of maltreatment and removal is the best way to mitigate negative impacts of foster car, FFPSA includes a new source of federal support for interventions to strengthen families. This fundamental change to federal foster care funding allows a state to claim reimbursement for the prevention of foster care placement and, further, requires services to be evidence-based, based on rigorous standards, and focused on parenting skills, mental health and substance abuse prevention and treatment programs, without regard to a child's family income level. In practice, this means that children and families receive appropriate supports in their homes, when possible, which is consistent with the available research and evaluation findings of how to support families staying intact and how to best ensure children can reach their potential. The prevention services in FFPSA have the potential to ensure that children do not experience the kinds of abuse, neglect, and household dysfunction that can lead to negative impacts on their development.

Reducing the overuse of group homes

The HHS report on group home placements demonstrated further support for congressional action on group homes, earned support from the executive branch, and served as compelling evidence of the need for reform. In this case, HHS effectively leveraged the administrative

data collected to generate findings that impacted policy. A main provision of FFPSA makes federal funding for group home settings stop after 14 days unless specific criteria are met. This, in effect, limits states from engaging in group home placements except for temporarily and when it is the necessary placement, which is where research suggests the use is most effective. Public child welfare agencies must ensure the setting is the appropriate level of care for a specific child based on an assessment and that the facility has the proper support and experts in place to provide necessary services. Stopping the federal funding for group homes unless there is a documented assurance that the placement is appropriate for a child or youth, will directly impact state behavior given the high costs of these placement settings.

Recruitment of foster families

If a child must be removed from his or her home and a family member is unavailable to care for a child, the next best situation is to place the child in a foster family home before the child can either be reunified with their family or adopted. Congress also included in the legislation a small amount of new funding, to support states in the recruitment of high-quality foster families. While modest, it is a start to further develop and strengthen home settings. Expanding the availability of high-quality foster family homes is a critical goal in best serving children in a loving environment, with or without a family member. When children must be removed from their birth families, another family-like setting is the preferred approach when possible. The legislation also established federal funding for evidence-based kinship navigator programs that help find and support kin caregivers and support this type of placement, which is shown to be best for kids if home renewal is necessary. A growing body of research suggests that children living with extended family fare better than with nonrelative foster parents. Placements with a relative result in greater placement stability and fewer behavioral and emotional problems.[27]

FFPSA is a transformational shift that garnered the movement underway in states and backed the changes with a meaningful shift in federal funding. Policymakers crafted the legislation to use the anticipated federal savings from limiting payments for group home settings toward increasing funding for evidence-based programs to prevent removals. Thus, in its design, Congress, through members and their staffs, applied the available research evidence to inform the design of FFPSA. Given that child welfare systems are complex and the federal government has a relatively remote role, there was not a direct one-to-one relationship of evidence to policy change. Instead, a large body of evidence was developed, strengthened over years, and synced with a bipartisan movement to ensure the well-being of vulnerable children that served as the wind behind the sail of FFPSA.

LESSONS

- ***Using funding to change state behavior.*** Federal funding was the main lever policymakers could use to impact state behavior and support child welfare agencies in moving practice toward what research suggests serve kids best. FFPSA incentivizes the use of evidence-based practices for services to prevent the need for foster care placements, and if placements must occur, it realigns funding to cover costs when a child is in the least restrictive and most family-like settings, especially with people who a child already knows. While change to the child welfare system can be slow, FFPSA used the most powerful federal levers of financing to ensure systems had more support and encouragement to change practices. For success in employing evidence to tackle a problem, a change must occur in the proper place of influence. Based on the structure of federal financing of child welfare, the funding stream was the exact spot.

- ***There is a need to evaluate evidence-based policy activities.*** With these central provisions of law set to go into effect starting October 2019, the potential effectiveness is yet to be seen. However, even with a foundation of evidence-based programs and services, evaluations will be able to convey the overall effectiveness of the legislation. In the short-term, based on the extensive preparation taking place and conversations occurring throughout the child welfare field and in jurisdictions, FFPSA has made its mark by increasing the focus on prevention and child well-being.

- ***Data collection was crucial in highlighting gaps and problems.*** While federal requirements can be perceived as overly intrusive and burdensome, data collected at the federal level was critical in this situation and provided a clear picture of what was happening in all 50 states, even with such different state systems and their respective nuances. FFPSA successfully based the legislation on evidence, built from data, that made the clear case for reforms to better serve children and families. ■

---

1   U.S. Department of Health and Human Services. *Child Maltreatment 2017.* Administration for Children and Families. Children's Bureau, 2017.
     Available at: https://www.acf.hhs.gov/sites/default/files/cb/cm2017.pdf.

2   Ibid.

3   Mary Dozier, J. Kaufman, R. Kobak, et al. "Consensus Statement on Group Care for Children and Adolescents: A Statement of Policy of the American Orthopsychiatric Association." *American Journal of Orthopsychiatry*, 84(3): 219-225, 2014.

4   For the FY 2018 funding numbers used to make the ratio calculation, see the CRS report: Emilie Stoltzfus. Child Welfare Funding in FY2018. U.S. Library of Congress, Congressional Research Service, R45270, July 30, 2018. Available at: https://fas.org/sgp/crs/misc/R45270.pdf.

5   U.S. Department of Health and Human Services. *What Is Child Welfare: A Guide for Educators.* Administration for Child and Families. Children's Bureau, June 2018.
Available at: https://www.childwelfare.gov/pubPDFs/cw_educators.pdf.

6   *Bipartisan Budget Act of 2018*, Title VII. Public Law 115-123, 2018.

7   Kristina Rosinsky and Sarah Catherine Williams. *Child Welfare Financing SFY 2016: A Survey of Federal, State, and Local Expenditures* (Bethesda, MD: Child Trends, December 2018). Available at: https://www.childtrends.org/wp-content/uploads/2018/12/CWFSReportSFY2016_ChildTrends_December2018.pdf.

8   Ibid. The main federal funding sources in child welfare come from titles IV-B and IV-E of the Social Security Act.

9   Ibid.

10   Wyoming Department of Health. *Title IV-E, Foster Care & Adoption Assistance.* Director's Unit for Policy, Research, and Evaluation, June 2017. Available at: https://wyoleg.gov/interimCommittee/2017/02-0717APPENDIXB.pdf.

11   Florida Department of Children and Families. *Eligibility Simplification IV-E Foster Care Desk Reference.* Center for Child Welfare, 2006.
Available at: http://centerforchildwelfare.org/kb/GenIVE/IV-E%20Eligibility%20Simplification%20Guide.pdf.

12   *Welfare Indicators Act of 1994.* Public Law 103-432, 1994.

13   U.S. House of Representatives. *Child and Family Services Improvement and Innovation Act.* 112th Congress, 1st sess. H. Rep. 112-210. Ways and Means Committee, 2011.

14   The program required the waivers to be cost neutral. States operating under a waiver received a capped allocation based on a projection of the level of costs over the demonstration project period absent the waiver.

15   Some jurisdictions approved to operate a wavier terminated the agreement early.

16   Richard P. Barth. *Institutions vs. Foster Homes: The Empirical Base for the Second Century of Debate.* University of North Carolina, School of Social Work, Jordan Institute for Families, 2002.

17   Bethany R. Lee, et al. "Outcomes of group care for youth: A review of comparative studies." *Research on Social Work Practice*, 21(2): 177-189, 2011.

18   Annie E. Casey Foundation. *Every kid needs a family: Giving children in the child welfare system the best chance for success.* 2015. Available at: http://www.aecf.org/m/resourcedoc/aecf-EveryKidNeedsAFamily-2015.pdf.

19   Dozier et al. "Consensus Statement on Group Care for Children and Adolescents.

20   U.S. Department of Health and Human Services. *A National Look at the Use of Congregate Care in Child Welfare.* Administration for Children and Families. Children's Bureau, May 2015.
Available at: https://www.acf.hhs.gov/sites/default/files/cb/cbcongregatecare_brief.pdf.

21   U.S. Department of Health and Human Services. *Highlights from the NYTD Survey: Outcomes Reported by Young People at Ages 17, 19, and 21 (Cohort 1) Data Brief #5.* Administration for Children and Families. Children's Bureau, November 2016. Available at: https://www.acf.hhs.gov/sites/default/files/cb/nytd_data_brief_5.pdf.

22   Richard P. Barth. *Institutions vs. Foster Homes: The Empirical Base for the Second Century of Debate.* University of North Carolina, School of Social Work, Jordan Institute for Families, 2002.

23   Jennifer E. Macomber, et al. *Coming of Age: Employment Outcomes for Youth Who Age Out of Foster Care Through Their Middles Twenties.* U.S. Department of Health and Human Services. Office of the Assistant Secretary for Planning and Evaluation. 2008. Available at: https://www.urban.org/sites/default/files/publication/31216/1001174-Coming-of-Age-Employment-Outcomes-for-Youth-Who-Age-Out-of-Foster-Care-Through-Their-Middle-Twenties.PDF.

24   Delilah Bruskahs. "Children in Foster Care: A Vulnerable Population at Risk." *Journal of Child and Adolescent Psychiatric Nursing*, 21(2): 70-77, 2008.

25   Ibid.

26   Vincent Felitti, et al. "Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults." *American Journal of Preventative Medicine*, 14(2): 245-58, 1998.

27   David Rubin, et al. "Needs of Kinship Care Families and Pediatric Practice." *American Academy of Pediatrics*, 139(4): 2-9, 2017.

5. LLAMA, LLAMA, CHILD SUPPORT UNDER OBAMA

Vicki Turetsky is an independent research and policy consultant for child support and social services. She previously served as the commissioner for the Office of Child Support Enforcement in U.S. Department of Health and Human Services during the Obama administration.
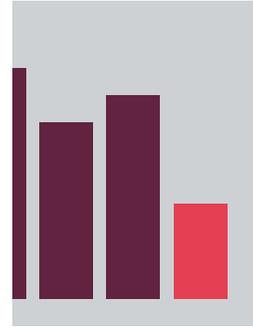
The purpose of the child support enforcement program is to increase financial support to a child living apart from a parent by collecting child support payments from the parent and sending the support to the child's household.[1] The program identifies parents, sets child support orders, and enforces child support payments. Nationwide, the child support program serves one in five children, making child support one of the largest income support programs for children.[2]

Congress passed and President Gerald Ford signed into law the child support enforcement program in 1975 under Title IV-D of the Social Security Act. The federal Office of Child Support Enforcement (OCSE) in the U.S. Department of Health and Human Services oversees the program. Every state and several tribes administer a child support program funded with federal and state matching funds. In addition, states receive federal funding based on a series of performance measures.[3] The child support program is highly automated, maintaining an extensive set of interstate databases, including the National Directory of New Hires.[4]

In 2011, President Barack Obama issued an executive order directing federal agencies to conduct a review of existing regulations.[5] OCSE conducted a comprehensive regulatory review and consulted widely with state agencies, courts, parent groups, and other stakeholders to identify rules that were outdated, inefficient, or ineffective. Some rules for the child support program had not been updated in decades. One goal of the rulemaking process was to incorporate evidence-based policies and practices into the program. In 2016, the Obama administration issued a regulation to modernize program operations.[6] The development of the rule relied extensively on research to ensure that government instituted the most effective known policies.

ISSUE BACKGROUND

The child support program has a long history of innovation. However, it took more than a culture of innovation, performance measurement, and commitment to children to see cases as families, to understand how the program's actions could impact parent-child relationships, and to address barriers to payment. While administrators understood that some parents were low-income, the program did not meaningfully address poverty as a barrier to collecting support dollars. Nor was the program generally viewed as an anti-poverty strategy.

Limited focus on poverty

Child support payments can be a critical source of income for low-income children. More than one-quarter of custodial parents have incomes below the federal poverty level.[7,8] Among poor custodial families who receive payments, child support averages more than 40 percent of family income.[9] However, the reality is that often, the lowest-income custodial families do not receive regular child support payments to support the financial needs of children.

Most noncustodial parents want to provide for their children, and most do pay child support. Three-fourths of child support payments are collected through automatic payroll withholding, similar to income taxes. This means that child support collection depends on the noncustodial parent's ability to find and keep stable employment. Thus, the best predictor of compliance with child support orders is a noncustodial parent's monthly gross earnings. Most parents who fall behind on child support payments have unstable employment and low earnings.[10,11]

Two decades of legislative changes had positioned the child support program to find parents and money. However, by 2009, the evidence was mounting that the program was less effective and sometimes counterproductive in its efforts to collect child support when noncustodial parents had low earnings. Even for low-income parents who want to support children, the design of the system poses challenges. Mothers and fathers from the same low-income communities often have similar barriers to employment. Noncustodial parents with limited education and marketable skills, an intermittent work history, lack of reliable transportation, and sometimes a criminal record often struggle to find and maintain work at a sufficient wage to support themselves and pay their child support obligations.[12]

The importance of child support income for low-income families made it imperative to identify and implement evidence-based policies to increase regular support payments by noncustodial parents without undermining family relationships and community stability. There was evidence that some child support practices were discouraging employment in the regular economy and pushing parents underground. A growing body of research identified payment barriers and disincentives and tested alternative approaches to increasing support payments and reducing unmanageable debt.

Slowing program innovation

In 1998, Congress enacted a child support performance-based incentive funding system. Under the Child Support Performance Incentive Act, the federal government pays a portion of federal funding based on state performance on a series of standard measures. The performance funding was set up as a competition among states, rewarding states that performed better and penalizing states that performed worse on the measures.

The Child Support Performance Incentive Act was the culmination of a five-year effort, led by OCSE that operationalized goal-setting, strategic planning, and measurement as a result of the Government Performance and Results Act of 1993 (GPRA), a law designed to improve government performance management and strategic-planning systems.[13] One outcome of GPRA was the development of a collaborative, consensus-building approach to strategic planning, policy implementation, and operational changes. Using this collaborative approach, OCSE and states developed performance indicators aligned with core program priorities. State administrators used the performance data to set priorities, allocate program resources, and demonstrate return on investment.

Initially, the new performance system helped to create a performance culture and spur state innovation, as states implemented the new technology and enforcement authorities enacted in the 1996 welfare law. Between 1998 and 2002, collection amounts shot up by 40 percent and some collection rates doubled during a robust economy.[14,15] However, as the performance system matured, innovation slowed as states adjusted their operations to maximize incentive payments and avoid penalties. The program was slow to adapt to underlying shifts in the economy affecting low-wage parents.

By 2009, the program was beginning to struggle with the fall-out from the Great Recession, including federal and state funding cuts, state staffing declines, parental unemployment affecting collections, aging technology, and reduced attention from Congress and advocates. Before the recession, several states had begun to implement promising family support initiatives, often funded by OCSE through its competitive grant programs. However, most states did not consider initiatives such as specialized case management for incarcerated parents and employment services for unemployed parents to be part of the core child support program. When the federal grant funding went away, many of these initiatives were not sustained.

A program culture geared toward data, but not research

To carry out its program functions, the child support program is a significant consumer of data generated and matched from federal and state agencies, employers, financial institutions, and other sources to locate parents and assets and collect child support payments. The program also uses data to measure, monitor, and improve performance.

However, the goal of improving performance measures did not necessarily align with the goal of incorporating evidence-based practices. Performance measurement does not focus

on long-term impacts like child well-being. Child support administrators understood data management. They were not as familiar with research principles and methods. Administrators asked, "How do our numbers look?" but usually not, "What does the research say?"

By 2009, there were a growing number of research studies in the field. However, compared with other human-services programs like health care, child welfare, and Head Start, the evidence base was small. Studies described caseload characteristics, provided ethnographic insights into the lives of people in the cases, and identified benefits and problems associated with existing practices. A smaller number of studies tested promising program innovations. The research was beginning to answer the question of "What might work?"—but it was nowhere near answering the question, "What works best and in what context?"

One likely reason for the comparatively small number of child support studies was that legislators and administrators historically viewed the program as legal and operational, rather than policymaking and programmatic. However, the main reason for the small number of studies was due to limited research funds. The basic source of federal research funding is section 1115 of the Social Security Act and states, tribes, territories, and the District of Columbia are eligible for the funds. Although a relatively small allotment, these funds have the virtue of a matching fund component that allows states to match research funds with regular program funds—increasing the total number of dollars available for state research.

Section 1115 grants must be used to fund state pilots or demonstration projects and have an evaluation element. The OCSE typically used the funds for an array of state program projects without requiring a rigorous research component. The grants seeded many projects that tried out innovative program approaches, yet millions of dollars were spent without producing useful evidence on whether these approaches worked.

EVIDENCE AVAILABILITY

Two decades ago, most research evidence about child support was from small-scale nonexperimental studies and program data analysis. One early experiment was Parents' Fair Share, a large-scale multistate demonstration funded by OCSE in the early 1990s. The Parents' Fair Share demonstration tested the effectiveness of employment services, enhanced child support case management, parenting classes, and peer support in increasing the employment, earnings, and child support payments of low-income noncustodial parents of children receiving public assistance. Other early experiments included state demonstrations to pass through child support payments to families receiving cash assistance, instead of keeping the money to repay cash assistance.[16,17]

Most child support research has been conducted by nonprofit research firms (MDRC, Urban Institute, Center for Policy Research), state universities (University of Wisconsin, University of Maryland, University of Texas, Indiana State University, Iowa State University), and state

agency research units (Washington state and Orange County, California). Based on the existing research, administrators now know far more about the causes of nonpayment and the consequences of traditional child support policies and practices:

- *Parents' inability to pay support*: The main reason parents do not meet their financial obligations is a lack of resources to pay required child support. In practice, most parents who pay sporadically or who do not pay at all have limited earnings, posing a major challenge to achieving program goals to support children. In contrast, parents who fully pay their support obligations are more likely to be employed and have significantly higher earnings. As earnings increase, compliance increases.[18,19,20]

- *Unrealistic payment requirements*: Many states assume parents earn at least full-time minimum wages, even when there is evidence to the contrary. This assumption can be flawed. In one study, parents issued support orders based on imputed, or attributed, income actually earned 72 percent less than the amount listed on the child support worksheets. As a result, payment requirements are set at unachievable levels. The result is accumulating debt. Existing research suggests compliance and payments fall off by parents at all income levels when support orders are set higher than about 20 percent of gross earnings. Levels above 20 percent of income could in turn decrease payments.[21,22,23]

- *Uncollectible debt*: A nine-state study of the causes and collectability of child support debt conducted by the Urban Institute found that parents with reported incomes of $10,000 or less owed 70 percent of the unpaid debt balance. Incarceration is one reason for the buildup of large uncollectible debts because few parents can work or make child support payments while in prison. Other studies found that unmanageable child support debt discourages employment and ongoing support payments.[24,25,26,27,28,29,30]

- *Employment services*: The Parents' Fair Share multistate demonstration resulted in small increases in employment and child support payment rates and increased earnings for the hardest-to-employ noncustodial parents. Two state studies with strong nonexperimental designs, Strengthening Families Through Stronger Fathers (New York) and Noncustodial Parent Choices (Texas), found that noncustodial parents offered a set of employment and other services were employed at higher rates and therefore paid more child support. Also, the New York study found that the parents had higher earnings.[31,32,33]

EVIDENCE USE

While a good deal of evidence existed on certain aspects of the program, key features of the program design lacked robust evidence to inform regulatory or programmatic decisions. Even so, the Obama administration proceeded with a regulatory update that prioritized where knowledge existed about how to make meaningful improvements.

OCSE administrators had several interrelated goals for modernizing the child support program. To successfully promote more effective policies and practices, the program culture had to change. For the program culture to change, child support administrators and line workers needed to access and understand the effects of existing practices on low-income families and adopt better alternatives. To identify better alternatives, the program needed to build the evidence base. To build the evidence base, administrators needed to see the value of applying research in the child support program. In addition, researchers needed to have access to funding and recognize that policymakers were interested in their work.

In other words, research evidence became the frame for highlighting the circumstances of low-income families, the basis for implementing improved policies and practices, and the aim of OCSE research capacity-building. OCSE administrators interjected the language of research into policy and practice discussions. The national strategic plan, updated through a state consensus-building process, included a section on building the evidence base.[34] OCSE used evidence to tell the stories of low-income parents in the child support caseload and to frame acceptable policy responses. The question asked at every turn was, "What would actually work to increase parental support for children?"

Goal one: Strengthening program effectiveness for low-income families

OCSE reviewed regulations and identified several that had not been revised since the program began in 1975. They consulted closely with state administrators and courts to determine which regulations to update. OCSE proposed new rules in several areas, including child support guidelines, civil contempt hearings, and debt buildup during incarceration. One organizing principle of the regulatory changes was that child support obligations should be based on "earnings, income and other evidence of ability to pay." In addition, the proposed rule allowed states to use federal funding for a limited set of employment services to increase the ability of low-income noncustodial parents to pay child support.

The preamble to the rule documented the specific research findings and studies that contributed to the rules.

Goal two: Building a family-centered mission

The child support program mission had been shifting for years from a debt recovery focus to a longer-term family support mission. This was partly in response to emerging welfare reform research about the role of child support income in families leaving cash assistance, research on debt accumulation, noncustodial parent employment research, the pass-through studies, and research examining the relationship between parenting time and child support payment. Federal legislation and initiatives introduced during the Bill Clinton and George W. Bush administrations built on these research findings.

During the Obama administration, OCSE highlighted evidence to frame a family-centered approach to child support that combined enforcement strategies with case management

and services. This approach focused on understanding caseload demographics and the effects of poverty, identifying the reasons for nonpayment through data analytics and early intervention, implementing specialized case management for cases with missed payments, and incorporating procedural justice principles. The model centered around five evidence-based practices: (1) right-sized orders based on ability to pay; (2) debt management; (3) support collections passed through to families; (4) employment services; and (5) parenting time.

The approach was based on family-centered and "two-generation" service-integration models in other fields, including child welfare, education, and health care. The national strategic plan and the 2016 federal rule included the family-centered model, which received broad support from states, counties, courts, and advocates.

Goal three: Building an evidence-based culture

While OCSE continued to emphasize performance-data trends, it coupled performance data with discussions about the research evidence. While some administrators initially expressed concern that the implementation of evidence-based practices would compete with performance goals, particularly cost-effectiveness, OCSE showed how evidence-based practices could improve performance. In presentations and program documents, OCSE administrators emphasized the research evidence that suggested better ways to do business. OCSE invited researchers to give presentations to staff, and highlighted new research at conferences, in social media, and in direct communications to child support administrators and their policy staff across the country. State administrators actively participated in an ongoing exchange of information and insights.

Goal four: Building a body of research evidence

Recognizing gaps in existing knowledge, OCSE used three strategies to help build a strong evidence base.

First, OCSE incorporated a rigorous evaluation component into federal projects funded with section 1115 grant funds. OCSE commissioned an experimental evaluation of demonstrations in the areas of noncustodial parent employment and alternatives to contempt hearings and incarceration. Eight state project sites received grants, with one state in charge of managing the research. Also, OCSE managed a number of smaller pilot implementation studies or braided funding with larger demonstrations conducted by other federal agencies in the areas of homeless veterans, parenting time, asset accounts, and prisoner reentry. OCSE regularly convened grantees to establish "learning communities."

Recently published results of the noncustodial parent employment demonstration showed the program benefits outweighed the costs in 10 years. Specific findings included a substantial improvement in the satisfaction with the child support program, which was a major achievement since distrust ran high among program participants at enrollment. It also increased the likelihood of working and the amount of earnings. Parents' sense of responsibility for their children and their contact with their children also increased.[35,36]

Second OCSE sought increased state capacity to conduct child support research. OCSE provided a series of grants to state child support agencies that agreed to partner with state universities to conduct state-specific research, with the goal of increasing state-funded child support research through such partnerships. OCSE and its partners designed a series of behavioral-economics research grants to show states how to conduct rapid-cycle research, saving states money and time before implementing small policy changes.

Finally, to strengthen OCSE's capacity to manage research and to use grant funds, OCSE was reorganized to create a division structure that was responsible for research and staffed with appropriate experts.

LESSONS

- *Culture change may be needed before policy change.* OCSE had two tasks: first, to persuade state agencies and courts that research evidence of what works—rather than values, legal precedent, or operational efficiency alone—should guide program policies and procedures; and second, to marshal the evidence in support of specific policy changes. That OCSE was successful speaks to the ongoing dialogue within the child support program, the commitment to performance, and the willingness of state administrators to engage, consider the evidence, and come to a consensus. In the comments on the proposed regulation, stakeholders—with few exceptions—strongly supported the proposed rule. OCSE administrators worked hard to modify the rule to address operational concerns while still honoring the research. Although the rule took several years to promulgate, that time allowed consensus for change to build, contributing to a longer-term culture shift.

- *Resource gaps can limit evidence availability and use.* Over time, the child support program has developed a basic evidence base. However, it is not very large and not very specific. For example, OCSE administrators can point to rigorous studies that support lower orders or the efficacy of employment services, but have little research on whether one intervention model is better than another. In large part, this is because the child support field has limited resources for research. Even though there is more interest than before in child support research, it may be unlikely to garner sufficient research funding to build a stronger evidence base.

- *Congress, not the evidence, has the final say.* In the proposed rule, OCSE included language to fund a limited set of employment services based on the research evidence for the subpopulation that needed and wanted services. The existing statute led the Obama administration to conclude that it had authority to promulgate a rule to pay for limited employment services for noncustodial parents. Commenters overwhelmingly supported this provision, and OCSE did not expect that the cost would be high based on its analysis. However, rulemaking authority is only as broad as Congress thinks it should be. Key Republicans in Congress challenged some provisions of the proposed rule, including the employment provision, based on administrative overreach. Compounding this issue are the rules of engagement between Congress and the executive branch in place during the rulemaking process that limit free and open dialogue. The administration ultimately removed the employment provision before finalizing the regulation, and the remaining rule received congressional support. In the end, the policy decision was based on political deference, not evidence. ■

1   This chapter is dedicated to the memory of two OCSE employees who devoted their careers to improving the child support program before they each passed away last year. Barbara Addison was the lead author of the Flexibility, Efficiency and Modernization in Child Support Enforcement Programs adopted in 2016. Adrienna Johnson helped prepare program data for publication every year, providing public access to the data. It seems particularly fitting to recognize their work in this chapter.

2   Melody Morales. "The Child Support Program Provides More Support to Families in 2016." *Office of Child Support Enforcement Data Blog*, December 1, 2017. Available at: https://www.acf.hhs.gov/css/ocsedatablog/2017/12/the-child-support-program-provides-more-support-to-families-in-2016.

3   Duties of Secretary. U.S. Code 42 §§ 652(g) and Incentive Payments to States. U.S. Code 42 §§ 658a.

4   Federal Parent Locator Service. U.S. Code 42 §§ 653.

5   The White House. "Executive Order 13563 of January 18, 2011, 'Improving Regulation and Regulatory Review.'" *Code of Federal Regulations*, Title 3, 2011. Available at: http://www.whitehouse.gov/the-press-office/2011/01/18/improving-regulation-and-regulatory-review-executive-order.

6   U.S. Department of Health and Human Services. "Flexibility, Efficiency and Modernization in Child Support Programs." *Federal Register*, 81(244): 93492-93569, December 20, 2016. Available at: https://www.govinfo.gov/content/pkg/FR-2016-12-20/pdf/2016-29598.pdf.

7   Elaine Sorensen, Arthur Pashi, and Melody Morales. *Characteristics of Families Served by the Child Support (IV-D) Program: 2016 Census Survey Results*. U.S. Department of Health and Human Services, Office of Child Support Enforcement, November 2018. Available at: https://www.acf.hhs.gov/css/ocsedatablog/2018/12/2016-characteristics-of-families-served-by-the-iv-d-program.

8   Timothy Grall. *Custodial Mothers and Fathers and Their Child Support: 2015, Current Population Reports*. U.S. Census Bureau, 2018.

9   Elaine Sorensen. *The Child Support Program is a Good Investment: The Story Behind the Numbers.* U.S. Department of Health and Human Services, Office of Child Support Enforcement, Admiration for Children and Families, December 2016. Available at: https://www.acf.hhs.gov/sites/default/files/programs/css/sbtn_csp_is_a_good_investment.pdf.

10 Letita Logan Passarella. *Maryland's Child Support Caseload: Examining Obligors Who Owed Support in July 2017.* University of Maryland School of Social Work, Ruth Young Center for Families and Children, July 2018. Available at: https://familywelfare.umaryland.edu/reports1/cscaseload_obligors.pdf.

11 Office of Child Support Enforcement. *Preliminary Report Fiscal Year 2017.* 2017. Available at: https://www.acf.hhs.gov/sites/default/files/programs/css/fy_2017_preliminary_data_report.pdf?nocache=1529610354#page33.

12 *See, for example*: Elaine Sorensen and Chava Zibman. "Getting to Know Poor Fathers Who Do Not Pay Child Support." *Social Service Review* 75:3 (2001): 420-434.

13 *Government Performance and Results Act of 1993.* Public Law 103-62, 1993.

14 U.S. Department of Health and Human Services. *Fiscal Year 1998 Annual Report to Congress*. Office of Child Support Enforcement, December 1998. Available at: https://www.acf.hhs.gov/css/resource/fy1998-annual-report.

15 U.S. Department of Health and Human Services. *Fiscal Year 2002 and Fiscal Year 2003 Annual Report to Congress*. Office of Child Support Enforcement, December 2003. Available at: https://www.acf.hhs.gov/css/resource/fy2002-and-fy2003-annual-report.

16 Cynthia Miller and Virginia Knox. *The Challenge of Helping Low-Income Fathers Support Their Children: Final Lessons from Parents' Fair Share.* Manpower Demonstration Research Corporation, November 2001.

17 Daniel R. Meyer and Maria Cancian. *W2 Child Support Demonstration Evaluation Phase I: Final Report.* University of Wisconsin, Institute for Research on Poverty, 2001.

18 *See, for example*: Mark Takayesu. *How Do Child Support Order Amounts Affect Payment and Compliance?* Orange County Department of Child Support Services, 2013. Available at: http://ywcss.com/sites/default/files/pdfresource/how_do_child_support_orders_affect_payments_and_compliance.pdf.

19 *See also:* Carl Formoso and Jo Peters. *Determining the Composition and Collectability of Child Support Arrearages, Col. 1: The Longitudinal Analysis.* Washington State Department of Social and Health Services, 2003. Available at: https://www.dshs.wa.gov/sites/default/files/ESA/dcs/documents/cvol1prn.pdf.

20 *See also*: U.S. Department of Health and Human Services. *The Establishment of Child Support Orders for Low Income Non-Custodial Parents*. Office of Inspector General, OEI-05-99-00390. Available at: https://oig.hhs.gov/oei/reports/oei-05-99-00390.pdf.

21   Steven Eldred and Mark Takayesu. *Understanding Payment Barriers to Improve Child Support Compliance.* Orange County Department of Child Support Services, Research and Reporting Unit, September 2013. Available at: http://ywcss.com/sites/default/files/pdf-resource/understanding_payment_barriers_to_improve_child_support_compliance.pdf.

22   *See also*: Lauren A. Hall, Letita Logan Passarella, and Catherine E. Born. *Who Pays Child Support? Noncustodial Parents' Payment Compliance.* University of Maryland, School of Social Work, Family Welfare Research and Training, May 2014. Available at: http://www.youngwilliams.com/sites/default/files/u258/um_ssw_who_pays_child_support_-_ncp_compliance.pdf.

23   *See updated study*: Natalie Demyan and Letitia Logan Passarella. *Actual Earnings and Payment Outcomes Among Obligors with Imputed Income.* University of Maryland School of Social Work, Ruth Young Center for Families and Children, 2018. Available at: https://familywelfare.umaryland.edu/reports1/guidelines_imputedincome.pdf.

24   Carolyn Heinrich, Brett C. Burkhardt, and Hilary M. Shager. "Reducing Child Support Debt and its Consequences: Can Forgiveness Benefit All?" *Journal of Policy Analysis*, 30(4): 24-43, 2011.

25   Maria Cancian, Carolyn Heinrich, and Yiyoon Chung. "Does Debt Discourage Employment and Payment of Child Support? Evidence from a Natural Experiment." Institute for Research on Poverty, Discussion Paper no. 1366-09, 2009.

26   Elaine Sorensen, Liliana Sousa, and Simon Schaner. Assessing Child Support Arrears in Nine Large States and the Nation. Urban Institute, July 11, 2007.

27   Pamela Caudill Ovwigho, Correne Saunders, and Catherine E. Born. The Intersection of Incarceration & Child Support: A Snapshot of Maryland's Caseload. University of Maryland School of Social Work, Family Welfare Research and Training Group, July 2005. Available at: https://familywelfare.umaryland.edu/reports1/incarceration.pdf.

28   Harry Holzer, Paul Offner, and Elaine Sorensen. Declining Employment Among Young Black Less-Educated Men: The Role Of Incarceration and Child Support. Urban Institute, 2004. Available at: https://www.urban.org/sites/default/files/publication/58036/411035-Declining-Employment-among-Young-Black-Less-Educated-Men.PDF.

29   Carl Formoso and Jo Peters. Determining the Composition and Collectability of Child Support Arrearages Volume 1: Longitudinal Analysis. Washington State Department of Social and Health Services, May 2003. Available at: https://www.dshs.wa.gov/sites/default/files/ESA/dcs/documents/cvol1prn.pdf.

30   Nancy Theonnes. Child Support Profile: Massachusetts Incarcerated and Paroled Parents. Center for Policy Research, May 2002.

31   Kye Lippold, Austin Nichols, and Elaine Sorensen. Strengthening Families Through Stronger Fathers: Final Impact Report for the Pilot Employment Programs. Urban Institute, October 2011. Available at: https://www.urban.org/research/publication/strengthening-families-through-stronger-fathers-final-impact-report-pilot-employment-programs.

32   Daniel Schroeder and Nicholas Doughty. Texas Non-Custodial Parent Choices: Program Impact Analysis. The University of Texas at Austin, Lyndon B. Johnson School of Public Affairs, August 2009. Available at: https://raymarshallcenter.org/files/2005/07/NCP_Choices_Final_Sep_03_2009.pdf.

33   Cynthia Knox and Virginia Miller. The Challenge of Helping Low-Income Fathers Support Their Children: Final Lessons from Parents' Fair Share. Manpower Demonstration Research Corporation, November 2001. Available at: https://www.mdrc.org/sites/default/files/full_529.pdf.

34   Office of Child Support Enforcement. National Child Support Strategic Plan for 2015-2019. U.S. Department of Health and Human Services, Administration for Children and Families, 2016. Available at: https://www.acf.hhs.gov/sites/default/files/programs/css/national_child_support_strategic_plan_2015_2019.pdf.

35   Quinn Moore, Katherine Anne Magnuson, and April Yanyuan Wu. Benefit Cost Analysis Findings from the Child Support Noncustodial Parent Employment Demonstration (CSPED), University of Wisconsin, Institute for Research on Poverty, 2019.

36   Maria Cancian, Daniel R. Meyer, and Robert Wood. Final Impact Findings from the Child Support Noncustodial Parent Employment Demonstration (CSPED). University of Wisconsin, Institute for Research on Poverty, 2019.

6. ALPHA, BRAVO, CHARLIE

Reforming the Department of Defense Child Care Program

Linda Smith is currently the director of the Early Childhood Initiative at the Bipartisan Policy Center. She served as the deputy director of the Office of Family Policy for child care and youth, and ultimately became the director.

Beverly Schmalzried is the former director of Child and Youth Development Programs for the U.S. Air Force.
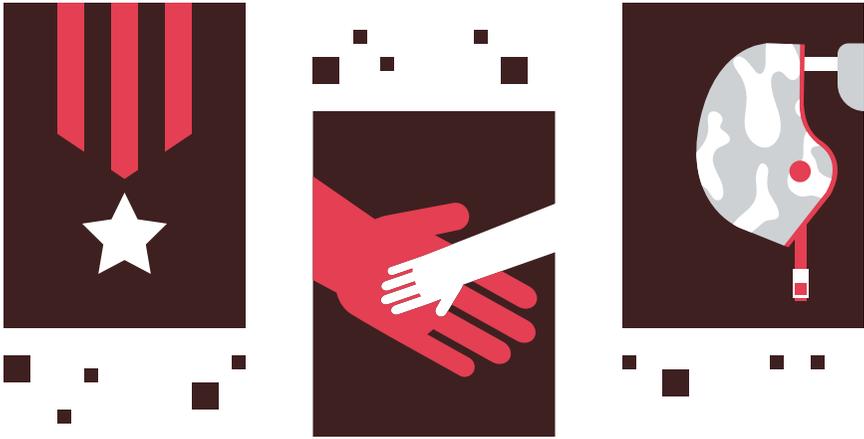
Military installations are a microcosm of American society. They reflect most American communities in their composition, services, and infrastructure. In addition to the military mission, installations have hospitals, fire and police services, grocery and department stores, schools, recreation centers, and housing. The major difference between military installations and civilian communities is the round-the-clock nature of the military mission. Military installations face unique challenges, such as a much younger demographic, a highly mobile workforce, and a lack of extended family to support family functions. These factors all increased the demand for child care in the late 1970s that far outpaced the supply.

Although the military, to varying degrees, began to address the need for child care beginning in the late 1970s, Congress held hearings in December 1987 and passed the Military Child Care Act (MCCA) of 1989.[1,2] The MCCA put a spotlight on military child care and the inconsistencies among how services were provided by the Army, Navy, Air Force, and Marine Corps.[3]

After Congress passed the MCCA, the Department of Defense (DoD) began addressing the three major problems of child care that now confront the nation in the 21st century: cost, quality, and availability of care.[4] This case study describes how DoD, in coordination with each of the military services, used evidence to develop and implement policies required by the MCCA, to improve the quality of military child care, and to initiate corresponding strategies to ensure accountability for implementation.

ISSUE BACKGROUND

Following the Vietnam War, in 1973 the nation established an all-volunteer military. This transformed the military from the "draft" to an all-volunteer force. To compete for the "best and the brightest" to operate increasingly high-tech weaponry, communications systems,

and equipment, the military had to expand its recruitment and retention strategies. The transition to an all-volunteer force made military families increasingly important to the performance of military readiness.[5] It was paramount to start retaining married individuals with families instead of prioritizing unmarried individuals without children.

Prior to the shift to an all-volunteer force, a typical military family consisted of senior officer wives and children, whom society expected to play a supporting role in their husbands' or fathers' careers. Even as the force began to change, service members were typically young unmarried men who served only briefly before rejoining the civilian world to begin their careers and start a family. Other demographic changes included an increase in the number of dual military couples and single parents serving in the military.

At the same time, women entering the workforce (and the military) sharply increased, and changes in military personnel policies allowed women to remain on active duty and have children. Most women on active duty are of childbearing age, and researchers estimate that at least 15 percent will become pregnant while on active duty. The proportion of women in the active component of the military has grown from about 2.5 percent in 1973 to 16 percent in 2018.[6] The enlistment of women resulted in a new requirement for full-time child care, especially for infants and toddlers, since the military required women to return to duty after a six-week maternity leave.

The quality of child care is dependent in large part on the quality of the interactions between children and their caregivers or teachers. Due to the needs of very young children, providing quality child care is labor intensive, and the cost of personnel is the single biggest cost driver in early childhood programs, accounting for approximately 80 percent of costs nationwide.[7] A critical indicator of the quality of child care is the stability and competency of the workforce. Before the MCCA, according to anecdotal reports, employee turnover was high. In a post-MCCA survey, 70 percent of the respondents said staff retention had been a major problem before the implementation of the MCCA.[8]

Accountability, quality, and cost challenges

During the 1980s, the country confronted a series of child sexual-abuse scandals in child care settings. Military child care was no exception. However, the high profile of the allegations in military programs (the Presidio of San Francisco and West Point Military Academy) captured the attention of the national media and ultimately prompted congressional hearings. The spotlight on child safety was intense, and DoD had to address safety and resolve it promptly.

Concern about the quality of military child care facilities began to surface in the 1980s. The House Committee on Appropriations' Surveys and Investigations staff reported that some of the child care centers visited were in old buildings originally constructed for other purposes, such as barracks, dining halls, and bowling alleys. The staff concluded that the poor building conditions contributed to "program inadequacies."[9]

Also, during the 1980s, military service leadership recognized the need for more child care and considered transforming what had been the "nursery" model of hourly care (primarily provided by volunteers who supported social functions) to a full-day program for single parent and dual-working families. The quality of the programs became an issue because military parents needed child care regularly for longer days.

In addition to quality and availability concerns, parents were vocal about the costs, especially for lower-ranking enlisted personnel. Prior to the MCCA, individual installation commanders set child care fees high enough to cover the cost of operating the child care center, but military families, especially on the lower-income scale, struggled to be able to afford these costs.

EVIDENCE AVAILABILITY

As with the current challenges facing child care across the country, improving the military's child care was not a simple task, and neither were the solutions. Each proposed policy change had an impact on thousands of children, their parents, the child care workforce, and mission readiness—none of which could be taken lightly. The use of data and research was essential to support multimillion-dollar budget increases. The availability of evidence depended on the core area being addressed.

Program and facility quality

Word about the lack of basic health and safety conditions in many child care facilities reached members of Congress. The Armed Services Committees (both House and Senate) asked the Government Accountability Office (GAO) to examine the condition of child care facilities, construction requirements, program operation, and methods for controlling costs. The GAO's 1982 report, *Military Child Care Programs: Progress Made, More Needed*, found that "many child care centers currently in use are neither safe nor suitable... [A]dditional facilities are needed in the Marine Corps to accommodate demand. The majority of centers in the Army and Navy and 20 percent in the Air Force need upgrading."[10]

The same 1982 GAO report also concluded that "DoD-wide minimum standards are lacking for important program elements including total group size, caregiver/child ratios, educational activities, staff training, and food service."[11] When DoD designated child care centers as community facilities in 1978, it gave the services authority to develop their child care regulations.[12] Subsequently, the services began developing service-wide child care regulations based on state standards and the research and expertise of the major social welfare and professional early childhood organizations.

Workforce

Anticipating that Congress would take steps to improve the child care workforce, DoD established a task force to study the issue and make recommendations for appropriate wage and training programs.[13] The task force examined data from the non-appropriated fund

and general schedule personnel databases from the military services and reviewed possible training and wages paid for equivalent work on installations.

Costs

The 1982 GAO report also found the fees charged in military centers were generally lower than in civilian centers, often by as much as 25 to 50 percent. Commanders were charging low fees under the mistaken impression that much lower-ranking personnel were using the centers; in reality, few lower-ranking individuals had children.[14]

GAO had surveyed government agencies in the Washington metropolitan area and found that civilian centers often did set child care fees by family income, and private centers sometimes reduced rates for families whose income was not sufficient to cover the full cost of child care. GAO recommended that the military services increase the fees for most users in order to improve the quality of the care offered but continue to offer lower fees to ranks E-1 through E-3.[15]

EVIDENCE USE

An advocacy group, the Military Family Wives Association, now the National Military Family Association, played a key role in bringing military families' concerns about cost and quality issues to DoD's attention.[16] A 1988 GAO report pointed to the variations in fees within the military services and even among bases within the same branch.

In response to issues of cost, quality, and availability of child care, Congress conducted hearings in 1988 and passed the MCCA a year later. The MCCA included 35 specific requirements. The DoD Office of the Inspector General conducted a second and lesser-known study of child care. Their report, issued in September 1990, contained 50 additional recommendations. The MCCA covered some, but there were new requirements for the military to address.

The use of evidence in the decision-making process was paramount. The decisions DoD leaders were making had major implications for the investment and use of taxpayers' dollars. Meeting the child care needs of more than a million active duty personnel and their families would ultimately require hundreds of millions of dollars. The allocated funds would affect the lives of over 3 million active duty and civilian personnel and their families. The military child care program was highly visible within Congress, and members were sure to closely scrutinize every decision. The MCCA itself required six major reports back to Congress, and DoD routinely provided briefs on the program to both Senate and House staff. The program was so visible that staff regularly apprised the secretary of defense and military leadership of its status.

The MCCA mandated changes to the child care program at a time when the defense budget was being reduced due to the fall of Communism.[17] Because the changes required significant funds and no appropriation accompanied the MCCA, funds had to be taken from already

shrinking accounts, leaving the program at odds with some military commanders and comptrollers.[18] To say there were tensions, as a result, would be a gross understatement.

While the MCCA dictated with some specificity the policies to follow to improve military child care, DoD had to use evidence from a variety of sources to achieve them. Evidence came from many sources, including the military services, the Defense Manpower and Data Center, GAO reports, civilian workforce data, both appropriated and non-appropriated fund budgets, research reports from Logistic Management Institute and RAND Corporation, military personnel reports, and research in early childhood education, children's health and nutrition, and effective training practices.

Facility quality policies

As a result of the 1982 GAO report, *Military Child Care Programs: Progress Made, More Needed*, Congress approved DoD's request for appropriated funds to construct new child care facilities for the first time in the 1982 budget.[19]

Military officials, prompted by GAO recommendations, agreed that a uniform DoD-wide design guide for child care centers could reduce both the cost and the time required for construction of new facilities.[20] In developing the uniform design guide, program officials, as well as designated service engineers and architectural representatives, relied heavily on evidence-based fire standards for child care centers developed by the National Fire Protection Association. They also based the new facility requirements on the median of state licensing standards related to the number of square feet required per child as well as the number of toilet and other facility features required to provide child care.

Program quality policies

Beginning in the late 1970s, the military services began developing service-wide child care regulations based on state standards and the research and expertise of the major social welfare and professional early childhood organizations. However, the 1982 GAO report concluded that "DoD-wide minimum standards are lacking for important program elements including total group size, caregiver/child ratios, educational activities, staff training and food service."[21]

In response, in 1988, DoD commissioned a review of state child care standards. A report from the Logistic Management Institute provided data on state standards for critical elements such as ratios, group size, and basic health and safety requirements.[22] A DoD instruction was issued in 1993, and based on the report, DoD decided to use the median of state standards. This is important because many still believe the DoD standards are higher than state standards. The distinction is that while the standards may be the median, the enforcement of them is very closely monitored. The services were required to modify their existing regulations to come into compliance with the DoD instruction.[23]

Further gains in quality were necessary to meet the MCCA's requirement for at least 50 military child care centers to become nationally accredited. To become accredited, the military services had to demonstrate they had achieved the quality of the National Association for the Education of Young Children Accreditation Standards. The association used current research to establish an accreditation standards to help parents identify high-quality early childhood programs.[24] Similar research was conducted by the National Association for Family Child Care to establish the accreditation standards for family child care homes and used by DoD to guide standards for family child care homes.[25]

Workforce policies

The task force on personnel completed its work around the time Congress passed the MCCA. The MCCA had two directives. First, it standardized a training program. Second, it required completing the training as a condition of employment. The MCCA required "a program to test competitive rates of pay to improve the competency and stability of the workforce." Based on the recommendations of the task force, DoD conducted a pilot program soon after passage.

To improve the competency of the workforce, DoD staff looked to the results of studies completed for the Department of Health, Education, and Welfare report *Children at the Center: Final Report of the National Day Care Study.*[26] The report demonstrated that child care-related education and training showed a moderately strong and consistent relationship to the measure of the quality of care but little relationship to cost. The study recommended that staff providing direct care to children receive training in child-related education and care. Because of the mobility of the military child care workforce, traditional training approaches, such as an institution-based certificates or diploma programs, could not be used.

In 1977, the Army received funding from the Department of Health and Human Services to develop staff training materials and administrative guides. By September 1980, the Army had developed 16 training manuals and guides. These materials covered child development from infancy through school-age, planned appropriate educational activities, and provided guidance on managing all aspects of military child care centers.[27] A few years later, the Navy funded a non-appropriated fund contract to develop training modules based on more current research. These modules were later expanded to family child care and school-age care with Army funding. In 1993 DoD adopted use of the modules to train all military child care workers.

Accountability policies

In the late 1970s, prior to the MCCA passing, the military services had begun to conduct inspections of its child care centers so that the centers could participate in the U.S. Department of Agriculture (USDA) Child and Adult Care Food Program. To receive USDA funds to help pay for the cost of the meals and snacks served in child care centers, the military programs had to be state licensed. In lieu of this licensing, the services developed their own inspection and certification requirements in order to receive USDA funds.[28]

Later, in response to the MCCA, DoD established a system of accountability that included multilevel inspections.[29] Beginning in 1990, each military child-development program received a minimum of quarterly unannounced inspections, with one from a higher command specialist. In turn, each military service conducted at least one unannounced inspection in each major command, and DoD conducted one unannounced inspection of each military service. In cooperation with the military services, DoD developed and issued an inspection checklist based on the best-known indicators of quality in child care and, in large part, on the National Fire Protection Association's Life Safety Codes and *Caring for Our Children: The National Health and Safety Performance Standards: Guidelines for Out-of-Home Care.* The latter standards, published for the first time in 1992, were the product of a five-year national project funded by the U.S. Department of Health and Human Services and included a comprehensive set of health and safety standards. The resulting DoD standards became the minimum for all inspections. All military child-development specialists were trained in advance on the use of the inspection checklist.

Finally, some decisions to inspect specific installations were based in part on calls from parents and the public to the DoD Child Abuse and Safety Hotline, which the MCCA required. Substantiated calls required an unannounced follow-up inspection.[30]

Establish cost and achieve affordability

All of the requirements in the MCCA were "unfunded mandates." In other words, there was no increase in the DoD budget to fund the improvements, rather funding had to be identified and reprogrammed from existing operating and maintenance accounts.[31] Although in 1978, DoD had designated child care centers as facilities for which the federal government had responsibility, the programs were primarily viewed as services to be paid for by the user fees.[32] When the MCCA passed, the total appropriated budget for child care was $89.9 million and included only 1,222 positions funded with taxpayer funds. Parent user fees paid most of the cost, and there was little expectation that funds appropriated for the military mission would go toward operating child care programs.

All of this was complicated by the fall of Communism in 1989 and the accompanying drawdown of the military both in numbers of active duty and in budgets.

The MCCA required parent fees based on total family income with those earning less paying less. The total revenue raised from parent fees had to match taxpayer dollars on roughly a dollar-for-dollar ratio that in turn was enough to pay for the general schedule child care staff. Although the pay scales for the actual military members were a matter of public record, the incomes of nonmilitary spouses were not available. DoD and the military services developed a variable income fee schedule designed to meet these requirements so that parents could pay based on their total family income.

Establishing the DoD child care fee structure required estimating the non-appropriated fund and parent fees cost of providing care under very dynamic conditions. During the initial years of the MCCA, the services sought to add appropriated fund positions for program

management and caregiving staff. Funding the program management and at least half of the caregiving staff was critical to reducing the cost of program operations that parent fees paid for. During the first years of the new fee structure, delays in obtaining funding for and filling appropriated fund positions while also setting parent fees based on income led to non-appropriated fund losses at some installations and disgruntled installation commanders and higher-ups at headquarters.[33]

Increase availability

The MCCA required DoD to study the expected demand for child care for military and civilian personnel in order to provide a plan to meet demand and to estimate the cost. Using the 1989 GAO report *Military Child Care, Extensive, Diverse, and Growing* as the benchmark of how much care was available, DoD and the military services developed a formula based on available data from the Defense Manpower Data Center on military families to project the need. The formula used data on the total number of military personnel, the number of married military personnel, single parent military, military with civilian working spouses, the number of children by age category, and the number living on-base versus off-base. DoD staff tested the formula at select installations and made adjustments based on the findings. DoD then applied the formula to all military bases, and used the results to project the potential need for child care. Also, the military bases tracked waiting lists, which, while an imperfect measure, provided additional information. The baseline set for available care was reported to Congress and included a five-year plan to meet the need.[34] This document also served as a budget justification moving forward.

By 1995, more than 95 percent of all military child care programs were nationally accredited by the National Association for the Education of Young Children, compared with 10 percent nationwide at the time.[35] The accreditation then and now remains the gold standard in child care quality. The amount of care available to families had increased from 53,000 to 154,000 spaces, and staff turnover had been reduced to 30 percent (the average military family transferred about every three years.)[36] Before the MCCA, employee turnover in military child care programs ranged from 65 to 300 percent.[37] In 1997, as a result of the work done, then-President Bill Clinton proclaimed the Military Child Care Program a national model.[38]

One policy decision stands out as the most critical and fundamental to the creation of a comprehensive, high-quality child care system: apply all requirements of the MCCA to all components of the system, including child care centers, family child care homes, school-age care, and part-day programs. The MCCA itself required many of the changes apply to centers only. After carefully considering the ultimate consequences of improving just one component, DoD decided that any child receiving care in a setting sponsored by the DoD must maintain certain levels of safety and quality. Had DoD not made that decision, the likelihood was high that all parents would abandon family child care or school-age programs in favor of the centers, creating yet another crisis.

LESSONS

- *Evidence-based policymaking requires multiple perspectives and flexibility.* Fixing military child care required a commitment from Congress, DoD leadership, the military services and installation commanders. Data and research informed decisions to the extent possible. When data did not exist, DoD used proxies or conducted pilot programs to test approaches and determine the best course of action. When pilots yielded information, the department was willing to adjust policies to ensure the highest-quality care was available.

- *Evidence can improve understanding of multifaceted policies.* While tackling multiple issues at one time was challenging for the DoD and the military services, the goal of high-quality, affordable, and available child care for military families could not have been achieved by fixing only one piece of the system or one piece at a time. For example, paying staff more without requiring higher levels of training and competence would have increased parent fees without significant improvements to quality. Likewise, issuing standards without enforcing them would have achieved minimal improvements. Implementing a graduated-fee policy without significant underwriting of the cost of care from nonparent sources would have forced higher-income parents to subsidize lower-income families at such a high rate that they would leave the system—a common dilemma in civilian child care. Even with all these elements in place, effective management is necessary and having access to useable evidence was an aspect of this. Currently, few of these elements are in place in the civilian child care sector, and it is unlikely the civilian sector can duplicate the military's success without major structural and funding changes at a national level.

- *It is hard to make policy decisions when there are knowledge gaps.* There were notable gaps in evidence when the DoD began implementing the MCCA. Chief among the gaps were the lack of data on the cost of care, the actual need for care, and how to measure quality. These gaps were ultimately closed as data became available during the process. Depending on the military service, the quality of the installation programs varied. There were no minimum DoD-wide health and safety standards, and no information available on the actual quality of experiences children were having. There was also no information on the demand for care and who was actually using the different types of care (center-based, after-school care, or family childcare). There were no nationally recognized, research-based instruments to determine quality. Perhaps the biggest gap concerned the actual cost to provide care in the various settings for the different age groups of children, making it difficult to develop budget projections. ■

1   U.S. Government Accounting Office (GAO). *Military Child Care Programs: Progress Made, More Needed.* GAO/FPCD-82-30, June 1, 1982. Available at: https://www.gao.gov/assets/140/137668.pdf.

2   U.S. House of Representatives. "Department of Defense. Onsite Child Care Initiative Hearing." Subcommittee of the Committee on Government Operations, 100th Congress, December 15, 1987.
Available at: https://catalogue.nla.gov.au/Record/2785367.

3   Gail L. Zellman and Anne S. Johansen. *Examining the Implementation and Outcomes of the Military Child Care Act of 1989.* RAND Corporation, 1998.
Available at: https://www.rand.org/content/dam/rand/pubs/monograph_reports/2009/MR665.pdf.

4   Ibid.

5   M. Clever and D.R. Segal. "The Demographics of Military Children and Families." *Future Child,* 23(2): 13-39, Fall 2013.

6   U.S. Department of Defense. *Table of Active Duty Females by Rank/Grade and Service.* Active Duty Master Personnel File, Military Academies, January 2018. Available at: https://www.dmdc.osd.mil/appj/dwp/rest/download?fileName=rg1801_female.pdf&groupName=milRankGrade.

7   Child Care, Inc. *Operating Budgets for Child Care Center.* 2001. Available at: http://www.earlychildhoodnyc.org/resourceguide/resources/OperatingBudgets-Child%20Care%20Inc%20publication.pdf.

8   Zellman and Johansen. *Examining the Implementation and Outcomes of the Military Child Care Act of 1989.*

9   GAO. *Military Child Care Programs,* 1982.

10   Ibid., i.

11   Ibid., i.

12   Ibid.

13   Nancy Duff Campbell, Judith C. Appelbaum, Karin Martinson, and Emily Martin. *Be All That We Can Be: Lessons from the Military for Improving Our Nation's Child Care System.* National Women's Law Center, April 2000.
Available at: https://nwlc-ciw49tixgw5lbab.stackpathdns.com/wp-content/uploads/2015/08/military.pdf.

14   GAO. *Military Child Care Programs,* 1982.

15   Ibid.

16   National Military Family Association. "About Us: Our Story."
Available at: https://www.militaryfamily.org/about-us/our-story/.

17   Zellman and Johansen. *Examining the Implementation and Outcomes of the Military Child Care Act of 1989.*

18   Ibid, 192.

19   *Department of Defense Appropriation Act.* Public Law 97-114, 1981.

20   GAO. *Military Child Care Programs,* 1982.

21   Ibid.

22   Robert L. Crosslin and Trevor L. Neve. *Acceptable and Affordable Child Care Services for Military Family.* Logistics Management Institute, 1998.

23   U.S. Department of Defense. *Instruction: Child Development Programs.* Instr. 6060.2, January 19, 1993.

24   Academy for Early Childhood Program Accreditation. "History of NAEYC Accreditation."
Available at: http://oldweb.naeyc.org/academy/HistoryNAEYCAccreditation.asp.

25   National Association for Family Child Care. *Quality Standards for NAFCC Accreditation.* 2017.
Available at: https://www.nafcc.org/file/bfae1239-d67e-41d9-820d-96c059842fac.

26   Administration for Children, Youth and Families. *Children at the Center: Final Report of the National Day Care Summary.* 1979. Available at: https://eric.ed.gov/?id=ED168706.

27   GAO. *Military Child Care Programs,* 1982.

28   Ibid.

29   Zellman and Johansen. *Examining the Implementation and Outcomes of the Military Child Care Act of 1989.*

30   Ibid, 83.

31   Ibid, xix.

32   GAO. *Military Child Care Programs*, 1982.

33   Gail L. Zellman and Anne S. Johansen. "Chapter 8." In: *Examining the Implementation and Outcomes of the Military Child Care Act of 1989*. RAND Corporation, 1998, 109-128.
     Available at: https://www.rand.org/content/dam/rand/pubs/monograph_reports/2009/MR665.pdf.

34   U.S. Government Accountability Office. *The Potential Demand for Child Care Within the Department of Defense and a Plan to Expand Availability*. 1992.

35   Campbell et al., *Be All That We Can Be*.

36   Zellman and Johansen. *Examining the Implementation and Outcomes of the Military Child Care Act of 1989.*

37   Department of Defense, *A Briefing to Senator Ted Kennedy: The Department of Defense Military Child Care*. Office of Family Policy, 1997.

38   Campbell et al., *Be All That We Can Be.*

7. BABY, IT'S COLD OUTSIDE

Modernizing Energy Assistance Grants

Nick Hart is a fellow at the Bipartisan Policy Center and chief executive officer of the Data Coalition. He previously served as the director of BPC's Evidence Project. He also previously worked at the White House Office of Management and Budget covering energy assistance issues.

Since the 1980s, the U.S. government has provided energy subsidies to support low-income families in paying their heating and cooling bills. Historically when families were unable to meet basic energy payments, utility companies may have discontinued services. As a consequence, families faced freezing temperatures in the winter or excessive heat in the summer—all of which can contribute to a range of negative public health outcomes.

Policymakers determined that establishing a program to counteract the potential for negative effects from extreme weather conditions was a public health priority. This was the genesis of the Low-Income Home Energy Assistance Program (LIHEAP), established in 1981. Utility companies favored the approach because it meant they were not on the hook to disconnect services from needy families in the depths of winter. Congressional representatives from cold weather states appreciated the financial grants to their states in the winter; representatives from warm states valued the assistance provided to their states in the summer.

The history of the program and the feel-good factor of supporting needy families led to stable funding for years, with support from a bipartisan and geographically diverse coalition in Congress. However, as budgetary pressures increased after the 2000s, the program also struggled to justify its allocation of funds to state grantees and whether the program used those funds most effectively to target the neediest populations. Numerous executive branch budget proposals, from both Republicans and Democrats, proposed cutting the funding allocated to the LIHEAP block grant.

Determining how much money to allocate to the program was related to understanding an array of factors. The total funding level directly affects the resource supply to provide services. But demand is harder to determine. While the government has information generally about the number of households eligible to receive the subsidies, the actual

scope of the population in need is less clear. Knowledge about need is based on housing characteristics like size and age, family financial stability, heating and cooling systems, home energy efficiency, health conditions of individuals in a household, and, importantly, the weather conditions in a given region and year.

For a program intended to address pressing needs in reducing energy burdens for low-income households, administrators knew relatively little about how to best fulfill the unknown demand for program services and what the most effective strategies for doing so were. One theory was that the best approach was to increase funding for LIHEAP to provide larger or more subsidies. Another theory was that improving general household income supports would be more effective. Still another theory emphasized the need for potential program reforms to ensure that resources addressed the long-term needs of the most burdened households.

ISSUE BACKGROUND

Traditionally, the vast majority of federal funding spent on LIHEAP grants supported direct heating and cooling assistance, or emergency assistance used to replace furnaces and air conditioners during times of crisis. A smaller amount of funding was also available to support energy efficiency upgrades with the goal of improving features of the housing stock over time in order to reduce the costs of heating and cooling.

LIHEAP permits state grantees incredible flexibility to determine how to allocate resources to best meet individual state needs and priorities. In practice, some states ask low-income households to apply for assistance at the beginning of winter, then take the total funding available and distribute it across all eligible applicants. Other states prioritize resources for the elderly and children. Some states encourage high levels of spending on energy efficiency; others virtually none. This mix of state strategies for implementing the program is characteristic of federal block grants, but it also reflects the range of climate, geographic, population, housing, and fuel characteristics across the country.

These characteristics also create tension points for program implementation. Inevitably, each winter there will be a cold spell that will translate into a surge in fuel prices in some part of the country. In the past, LIHEAP had a "contingency fund" that provided flexible funding to address these types of needs. But while weather can be predictable, less predictable is where the need will be greatest. Then in 2012, Congress ceased appropriating flexible contingency funding, meaning states and the federal government had to plan for unexpected changes in temperatures and fuel prices.

In 2014, the price of propane in the United States skyrocketed. For households in the upper Midwest, the high price of propane meant many could not fill their tanks when empty. Paying to refill a propane tank is a lump-sum cost that may cover fuel for multiple months or even an entire winter. The fact that propane prices would spike in the winter was predictable. That year, fall in the Midwest had been wet, so farmers depleted the supply of propane by drying crops prior to sending them to market. The shortage in total propane

supply drove up the price of propane across-the-board, including for residential deliveries for heating homes.

Beyond the context of energy prices and the regional politics of the program, funding LIHEAP is considered sacrosanct in Congress. But the executive branch—regardless of political party in control of the White House—tends to suggest flat or decreased funding for the block grant. Nearly every year of any administration, Republican or Democrat, the funding level for LIHEAP is determined by a combination of political goals about program optics and a need to meet broad spending targets for the entire budget. As a result, the decision about the LIHEAP budget request is anything but evidence-based; it is determined by staff in the White House, political appointees, and budget scorekeepers. That is not to say information and evidence did not play a role historically in program operations, but the funding level was traditionally more politically influenced. In the budget, LIHEAP is a political football.

In 2015, as the Obama administration was preparing its final budget proposal to Congress, the White House became interested in identifying "good government" initiatives with bold new ideas that also achieved broad administration goals. Coincidentally, within the White House complex, an active discussion about how to improve LIHEAP was already underway. Policymakers tacked LIHEAP onto budget framing that involved staff in the White House's Office of Management and Budget, the Domestic Policy Council, the U.S. Department of Health and Human Services (HHS), and external, non-governmental stakeholders. The stakeholders were collectively trying to determine a proposed funding level alongside a strategy for improving the long-term viability and effectiveness.

EVIDENCE AVAILABILITY

A vast range of evidence was available about LIHEAP operations and program participants' need, though the program lacked impact evaluations to characterize the extent to which the program affected intended outcomes.

Performance measures

For years, a series of performance measures that tracked the energy burden across sensitive subpopulations informed HHS implementation of LIHEAP. Individual state grantees largely report the data to support the performance measures on an annual basis. A 2005 evaluation suggested that high-burden households tend to receive more LIHEAP funding than households with lower energy burdens, though there were substantial gaps in benefit targeting.[1] HHS nonetheless routinely presented performance measures to Congress and to the American public in the budget and funding justifications.

Energy information

The Energy Information Administration at the U.S. Department of Energy collected data in a survey it operated on behalf of HHS. The Residential Energy Consumption

Survey specifically added questions about LIHEAP receipt to support the development of performance measures as well as other program insights.[2] While useful, the Residential Energy Consumption Survey is an intermittent survey rather than an annual one, leaving HHS with updates about LIHEAP participants' characteristics and energy-consumption patterns once every four or more years.

The Energy Information Administration also provides HHS and the American public with extensive information about energy prices, including predictions for fuel demand and costs over the winter. The short-term energy outlook includes projections for different fuel types in the winter and consideration of regional price variations.[3] This information is first available in the fall and the administration updates it monthly through the winter, with summary descriptive statistics published as open data.

Weather and climate information

The U.S. Department of Commerce's National Oceanic and Atmospheric Administration makes information available about how much variation from the norm different parts of the country's temperatures are. For example, using ambient temperatures, the agency estimates how much energy is needed to warm a building to a comfortable temperature. The information is a proxy for energy demand broadly across the country and applies for both cool and warm climates. Both historic and projected statistics are available.[4]

Population eligibility characteristics

The U.S. Census Bureau's American Community Survey provides statistics about household incomes and eligibility to serve as a proxy for need across the country. The Census Bureau provides statistics about poverty levels and household members at varying geographies through public-use datasets, which can provide insights about eligible and vulnerable populations, with consideration of age or disabilities in the household. The Census Bureau also collects and analyzes data about utility costs, which HHS uses to allocate some grant funding.

Housing characteristics

The U.S. Department of Housing and Urban Development collects additional data through the American Housing Survey.[5] In conjunction with the Census Bureau's Current Population Survey, the Department of Housing and Urban Development can generate statistics about the characteristics of homeowners in poverty on a biennial basis. For example, the housing survey can determine the share of households below the federal poverty line that rent versus own and that also have an elderly member of the household. While less specific to the LIHEAP-recipient population, these data are timelier than those collected by the Energy Information Administration. The survey also allows for stratification of housing and individual characteristics useful to understand trends in the housing stock over time.

Taking all of this information together, a robust depiction of LIHEAP-eligible and LIHEAP-recipient households emerges—a depiction analysts can pair with broader

climate, weather, and fuel price statistics to provide insights for policymakers about projected trends and performance.

EVIDENCE USE

In preparing for the 2016 White House budget proposal, Obama administration officials requested recommendations about how to develop a final decision on the LIHEAP funding level as well as any policy adjustments that might be reasonable for the program. Career and political staff considered the entire array of evidence in weighing whether to suggest policy changes to the program and, if so, what those changes would be. However, the determination about the funding level was largely based on politics and broader budgetary targets for the budget.

Every year, the Office of Management and Budget (OMB) runs a routine process for formulating the president's budget proposal. In September, agencies submit their budget requests to the White House for consideration, then the staff at OMB review, analyze, and refine the proposals with input from political appointees and agencies. It is a process of weighing priorities, incorporating the president's agenda, and promoting effectiveness and efficiency in government operations.

When considering potential LIHEAP changes, staff across OMB came together to discuss the intersection of multiple programs that relate to funding energy costs, addressing low-income household needs, and related energy efficiency operations. For example, the Supplemental Nutrition Assistance Program (SNAP, previously known as food stamps) can support some utility payments, meaning low-income households have multiple potential resources for addressing high energy costs or burdens. The staff had to weigh potential changes in an array of other programs as well as LIHEAP in formulating recommendations about how to proceed. In addition, the Obama administration had broad goals about achieving certain energy efficiency outcomes, which are relevant for LIHEAP because a portion of grant funds can go toward efficiency activities.

For LIHEAP, analysts at OMB developed a comprehensive analysis of information from the range of sources about demand, supply, projected price changes, and other programmatic features in a memorandum provided to the OMB director and the senior policy staff. The memo presented tradeoffs and acknowledged the gaps in knowledge that remained due to a lack of impact evaluations about program outcomes.

Traditionally in the OMB process, the career staff present these findings along with a recommendation to the OMB director. For LIHEAP, the recommendation was to develop substantial reforms about the long-term sustainability of the program, essentially shifting resources to the efficiency part of the program to develop improvements to the housing stock over time. Another prong of the recommendation was to create a policy that better linked funding to changes in weather patterns and prices in order to reflect the real-world changing conditions. A final prong recommended allocating funding for research, evaluation, and

innovation in the program. OMB's political officials agreed, based on the descriptive analysis available, that long-term changes made sense for the program.

In the ensuing weeks and months before the White House published the budget, OMB and HHS officials refined the proposal. The final proposal, presented in the president's fiscal year 2016 budget included funding for innovation and research, provided some funding for the traditional program, increased resources for energy efficiency activities, and developed a funding mechanism that would provide resources to state grantees when there were substantial anomalies in temperatures, fuel prices, or the eligible population for the program.[6] While the public justification of the policy reforms did not present the detailed references or analysis, OMB and HHS officials agreed on the analytical basis for the decisions reached about policies in the LIHEAP budget proposal for the year.[7]

While evidence was largely unused for deciding the actual funding level, the policy reforms proposed for LIHEAP were based on a variety of data sources and a collation of descriptive statistics or trends about key aspects of the program. Even then, substantial gaps and uncertainties remained about the eligible population, grantee implementation, and other program characteristics. But decision-makers were able to apply the evidence available to inform the most substantial proposed reforms to LIHEAP since the 1980s.

In the end, Congress did not adopt the administration's proposal and, in subsequent years, provided additional funding to the program. But the development of the proposal spurred broader discourse among stakeholders about better developing stronger evidence for potential reforms to the program.

LESSONS

- *Politics is inevitable for certain funding decisions.* When making the funding decisions about LIHEAP, politics was an inevitability that affected the level of funding requested and provided. However, the program lacked—and still does—impact evaluations about program outcomes that could affect future perceptions about whether the program achieves intended goals.

- *Decisions happen without perfect evidence.* The information available for LIHEAP decisions, including for proposing major reforms, includes vast uncertainties about the program operations, grantee performance and implementation, and characteristics of the population that might be relevant for aspects of the decision. But at the same time, much information was known. Policymakers were comfortable proposing reforms even in the face of uncertainty, and they recognized that they would still need to make budget decisions without perfect information.

- ***Motivated leadership could rely on credible staff.*** Career staff prepared sophisticated analysis underlying the decisions about the LIHEAP reforms at OMB. The staff established credibility with political appointees and senior career officials in making recommendations, while relying on a variety of trusted government data sources. The recommendations were within reason and plausible for policymakers to consider, so motivated leadership at the career and political level could advance the dialogue about reforming the policies with trust that at least analytically the decisions were based on the best available information.

- ***Data have to exist to be used***. Staff used many government data sources to develop a rationale for how to reform policy, and all of those datasets were publicly available. Open data can provide analysts information to make powerful and compelling arguments about what actions to take. Information had to be available first before analysts could use it. This suggests that attention to data infrastructure and availability is a key aspect of ensuring policy analysts can effectively operate in dynamic environments for decision-making.

- ***A learning agenda for LIHEAP would be productive.*** Given the gaps in knowledge about LIHEAP when reforms were on the table in 2014, a strategic plan to learn more about the program moving forward would be useful for policymakers. There has been some discussion about stakeholders developing a plan to study the program's effectiveness at achieving long-term outcomes. This in turn could support the production of research that would be useful for future reforms and that OMB staff and other policymakers would likely use in determining how to shape energy assistance activities in coming years. The proposed reforms from 2015 demonstrate that in the context of LIHEAP, policymakers were eager to identify and use available information to help inform the framing and extent of reforms—filling in knowledge gaps would only improve this capability. ■

1   Applied Public Policy Research Institute for Study and Evaluation. *LIHEAP Energy Burden Evaluation Study: Final Report*, July 2005. Available at: http://www.appriseinc.org/reports/LIHEAP%20BURDEN.pdf.

2   U.S. Department of Health and Human Services. *2015 RECS LIHEAP Household Administrative Data Matching.* Administration for Children and Families. Office of Community Services. Division of Energy Assistance. Transmittal No. LIHEAP-AT-2017-03. January 19, 2017.
Available at: https://www.acf.hhs.gov/ocs/resource/2015-recs-liheap-household-administrative-data-matching.

3   U.S. Energy Information Administration. *Short-Term Energy Outlook*. April 2019.
Available at: https://www.eia.gov/outlooks/steo/.

4   National Oceanic and Atmospheric Administration. "Degree Days Statistics: States and Cites." Climate Prediction Center. Available at: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/cdus/degree_days/.

5   U.S. Census Bureau. "American Housing Survey." Available at: https://www.census.gov/programs-surveys/ahs.html.

6   Nick Hart and Jeannie L. Chaffin. *2016 President's Budget: LIHEAP*. March 23, 2015.

7   U.S. Department of Health and Human Services. *Fiscal Year 2016 Justification of Estimates for Appropriations Committees*. Administration of Children and Families, 2015, 27-49.

8. WHAT WILL IT TAKE TO END FAMILY HOMELESSNESS?

The Family Options Study

Anne Fletcher is a social-science analyst for the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development. She was also a senior policy and research analyst for the U.S. Commission on Evidence-Based Policymaking.

The federal government has been funding efforts to address family homelessness since the 1980s. The passage of the McKinney-Vento Homelessness Assistance Act of 1987 was the first significant legislative action designed to address homelessness nationwide.[1] It authorized the Homeless Assistance Grants program within the U.S. Department of Housing and Urban Development (HUD). In fiscal year 2018, HUD received just over $2.5 billion in Homeless Assistance Grant funding.

While homelessness emerged as a problem of sufficient magnitude to induce a federal response in 1987, local communities have been grappling with homelessness for centuries. Recognizing the value of locally driven and coordinated efforts to address homelessness, HUD awards the Homeless Assistance Grants funding to entities known as "Continuums of Care." Continuums of Care are composed of a network of providers within a given geographic region who offer housing or services to people experiencing homelessness. Using HUD funds, along with other state and local funding, communities may operate a variety of programs designed to address homelessness. These programs range from emergency shelters that provide temporary overnight accommodations to permanent supportive housing, which offers a long-term rental subsidy packaged with a flexible array of supportive services.

Each year, the members of the Continuum of Care in communities across the country come together and assemble a single application for funding to HUD, thus providing an opportunity for the jurisdiction to prioritize the kinds of programs believed to be the most critical to address homelessness in their particular community. HUD requires Continuums of Care to establish a local information technology system, a Homeless Management Information System, which collects client-level data and information on the provision of housing and services to homeless individuals, families, and persons at risk of homelessness. There is an

expectation that communities will use their local data to better understand the composition of the population experiencing homelessness within their Continuum of Care, and to structure their homelessness assistance system to meet the identified needs of the local population.

The ultimate goal of a Continuum of Care is to create a homeless assistance system designed to match the needs of those experiencing homelessness with an appropriate program that will address their needs and end their homelessness. By mid-2000s, however, the body of research on family homelessness was largely descriptive, both of the population and of the various interventions designed to address family homelessness. Little data related to the impact of these different interventions existed, and thus communities lacked solid evidence regarding "what works for whom" to enable families to exit homelessness quickly and permanently. In addition, not much was known about the costs of implementing the various interventions.

ISSUE BACKGROUND

Family homelessness is dynamic, with families moving in and out of homeless assistance programs every day. During 2017, nearly 151,000 families with children (representing almost half a million individuals) accessed the homeless assistance system through a stay in an emergency shelter or transitional housing. Different theories on the cause(s) of family homelessness have led to the rise of different types of interventions to address the problem. One theory holds that homelessness is purely an economic problem that can be addressed by providing a family with assistance in affording a housing unit. Another theory poses that while housing assistance is indeed crucial, family homelessness is a result of other challenges faced by families (such as child welfare engagement, mental health or substance abuse challenges, and unemployment) that must be addressed in order to end their homelessness.

In addition to these two broad camps, theories vary on the length of time for which assistance must be provided to a family in order to end their homelessness, with some arguing that the need for assistance is permanent, and others arguing that the need for assistance is only temporary. Debate on the appropriateness of various housing and services interventions to assist homeless families is overlaid by the scarcity, as well as the range of costs, associated with different interventions. Programs to address homelessness typically vary along three primary domains: (1) length of stay/duration of assistance; (2) provision of supportive services; and (3) housing arrangement and financing thereof.

The lack of rigorous research about homelessness interventions hampers providers from targeting their resources efficiently and limits policymakers in providing guidance to communities regarding the optimal investments for their homeless assistance funding. Congress acknowledged this limitation in HUD's 2006 appropriations by directing HUD to "undertake research to ascertain the impact of various service and housing interventions in ending homelessness for families."[2] HUD launched the Family Options Study to learn about which housing and services interventions work best for families with children experiencing homelessness and to document the costs of implementing the various interventions.

EVIDENCE AVAILABILITY

The Family Options Study, launched by HUD in 2008, is a multisite evaluation designed to measure the relative impacts of various housing and services interventions for homeless families. The study has an experimental design to generate the most rigorous evidence suitable for informing policy.

In implementing the study, HUD sought to understand which interventions yield the best outcomes for families, whether certain types of families were better served by particular interventions, whether some interventions work better in the short-term or the long-term, and the costs of the different interventions. Because each intervention studied is designed to address homelessness through different pathways, outcomes of interest extended beyond housing stability to also include family preservation, child well-being, adult well-being, and self-sufficiency.

HUD implemented the Family Options Study in 12 communities across the country. In total, over an 18-month enrollment period, 2,282 families who had spent at least seven days in an emergency shelter, and had at least one child younger than 16, were randomly assigned to receive priority access to one of four possible interventions:

1. A non-time-limited housing subsidy, usually in the form of a Housing Choice Voucher;

2. Community-based rapid re-housing, which is temporary rental assistance, that was available for up to 18 months with limited housing-related services;

3. Project-based transitional housing, which provides temporary housing for up to 24 months in an agency-controlled building, coupled with intensive supportive services; and

4. Usual care, which was a mix of housing and services that homeless families may access from the shelter on their own without direct referral to one of the other interventions.

Researchers followed families for three years after random assignment. Researchers conducted extensive primary data collection with families when the study started and again approximately 20 and 37 months after random assignment. Also, interim contacts with families every six months provided additional data related to housing status and family composition at more frequent intervals. HUD used administrative data sources to measure additional family outcomes, including local Homeless Management Information System data, HUD administrative data on the receipt of subsidized housing, earned income data, and child welfare data.

The homeless assistance system functions as a network of programs to which families receive referrals, theoretically based on an individual family's particular set of needs. For this system to work efficiently, there must be evidence available to guide the referral process (families must be well-matched to the different available programs), and there must a sufficient supply of the right kind of programs for the kinds of families that seek assistance. Also, referrals are not directives; families choose to act upon referrals, and thus the

interventions must appeal to families. If families do not want what is being offered to them, they can choose not to accept the referral.

The Family Options Study was structured to mirror real-world conditions, with two exceptions. First, the implementation of the study forced communities to allocate referrals via random assignment, rather than through their regular procedures, which typically involve some type of needs assessment. Second, the study made one intervention available, the non-time-limited housing subsidy (referred to as "subsidy-only"), that is not typically available to families through the homeless assistance system.

Other interventions tested in the study—emergency shelter, community-based rapid re-housing, and project-based transitional housing—are typically available within any community and accessed through the homeless assistance system. For the subsidy-only intervention, however, families must access it through the mainstream housing assistance system by applying for the program through a local public housing agency. Housing assistance is not an entitlement, and currently, only one out of every four eligible applicants receives housing assistance. As a result, public housing agencies typically have long waitlists, which can take years to clear. While some public housing agencies may have preferences for prioritizing persons experiencing homelessness, most waitlists include long lists of eligible families with varying of levels of immediate need for assistance and no ability for a homeless family to receive preferential treatment. Thus, the provision of a set number of slots for the subsidy-only intervention in study communities represented an expansion in the range of interventions typically available for homeless families.

Findings

Reports published in 2015 and 2016 provided evidence about the effects, relative to usual care, of giving families in emergency shelters priority access to different types of housing and services interventions.[3,4] Both the short-term (20-month) and long-term (37-month) analysis documented vastly superior outcomes for families who were randomly assigned to receive a non-time-limited subsidy when compared with the outcomes of families who were randomly assigned to the other interventions. Families offered a non-time-limited subsidy experienced less than half as many episodes of subsequent homelessness and vast improvements across a broad set of measures related to residential stability in comparison with families assigned to other interventions. The benefits of the non-time-limited subsidy extended beyond the housing domain as well, generating powerful benefits on a range of outcomes that are critical for healthy child development, including:

- A reduction in child separations (at 20 months);

- Reductions in psychological distress of the family head, economic stress, exposure to intimate partner violence, school mobility for children, and food insecurity (at both time points); and

- A reduction in behavior problems and sleep problems of children (at 37 months).

The other interventions studied had few positive effects on families.

A second key finding concerned the cost of the various interventions studied. Analysis of the cost data revealed that the benefits conferred to families who were randomly assigned to receive a non-time-limited subsidy were achieved, on average, for roughly $45,500 over the full study period. This was only 9 percent higher than the costs that accrued to the families assigned to usual care. The families assigned to usual care cost, on average, $41,000 in housing and services over the full study period and demonstrated no improved outcomes.

Community-based rapid re-housing emerged as the lowest-cost intervention studied, with an average cost of $4,000 less than the other interventions over the full 37-month period. Although not associated with the same benefits as a long-term subsidy, priority access to community-based rapid re-housing achieved roughly the same outcomes as usual care but at a lower cost.

The striking benefits observed among families assigned to the subsidy-only intervention provide support for the view that, for most families, homelessness is a housing affordability problem that can be remedied with long-term housing subsidies without associated specialized services.

It is important to note that the modestly higher cost that accrued to the families assigned to receive a non-time-limited subsidy includes only the housing and services costs associated with program participation and/or emergency shelter. These costs do not consider any cost offsets that may have been the result of providing families with stable housing. Access to a non-time-limited subsidy led to reduced mobility, reduced adult psychological distress, fewer experiences of intimate partner violence, reduced school mobility among children, greater food security, and lower economic stress—all conditions that carry high economic and social costs for both families and communities.

EVIDENCE USE

Numerous policy-relevant observations emerged from the findings of the Family Options Study. The study for the first time presented policymakers with a broad range of information about effectiveness and cost of different interventions in addressing family homelessness.

Subsidies made available to families experiencing homelessness represent an intervention that families want, that they can use, and that delivers powerful benefits in important domains that make the lives of low-income families and children better. The study reveals how families value the offer of a subsidy: 84 percent of the eligible families used it compared with 60 percent for community-based rapid re-housing and 54 percent for project-based transitional housing. The study also demonstrated families can maintain the subsidy over time, even without the offer of tailored services, with more than two-thirds of subsidy families still using the subsidy at the time of the 37-month follow-up point.

For the crisis response system, the study provides clear evidence that the high cost of emergency shelter and transitional housing programs do not yield improved outcomes for families when compared with the outcomes of families offered rapid re-housing. In the absence of a sufficient supply of long-term subsidies for families experiencing homelessness, rapid re-housing becomes the most useful response for communities seeking to assist homeless families, as the lower cost of rapid re-housing enables communities to serve more families with their limited resources.

In describing the study, then-HUD Secretary Julian Castro stated, "This can't be something that just goes up on the shelf—it needs to be something that connects with what we do."[5] The findings from the Family Options Study led HUD to seek a historic investment of $11 billion to address family homelessness through a bold fiscal year 2017 budget request.[6,7] The budget request included a 10-year, $11 billion proposal to fund additional housing vouchers and rapid re-housing assistance. HUD projected that by targeting new voucher assistance in combination with normal voucher turnover, as well as increasing the availability of rapid re-housing within communities, that communities could serve 40,000 new homeless families each year, potentially ending family homelessness over 10 years. The White House specifically articulated that the funding request was a result of the research: "This significant investment is based on recent rigorous research that found that families who utilized vouchers— compared to alternative forms of homeless assistance—had fewer incidents of homelessness, child separations, intimate partner violence and school moves, less food insecurity, and generally less economic stress."[8] HUD further cited the research as an underpinning for the budget proposal in its documentation provided to Congress to justify the request.

Unfortunately, the 2017 HUD appropriation did not include any funding in support of this proposal. And the following administration did not request additional funds as part of budget proposals from 2018 to 2020.

Despite a lack of additional dedicated funding, the study findings provide HUD with guidance for how to best use available resources within the crisis response system. If resources are too limited to offer long-term subsidies to all families experiencing homelessness, community-based rapid re-housing emerges as the most effective policy response, because it can achieve outcomes similar to those of usual care and transitional housing at a lower cost. HUD has moved to evolve the rapid re-housing program since the time of the Family Options Study, extending the maximum time that families can access the intervention from 18 to 24 months. HUD continues to encourage communities to reduce their investment in costly transitional housing programs and to shift their resources into the lower-cost and equally effective rapid re-housing program. Additionally, HUD continues to partner with communities to strengthen their coordinated entry systems to better connect people with the interventions and mainstream services they need when they arrive at a shelter.

LESSONS

- ***Sometimes evidence will not be enough to change funding priorities.*** A primary conclusion of the Family Options Study is that, in most cases, family homelessness is caused by the inability to afford housing. Thus, scaling up investments to increase the supply of affordable housing is an essential component of any strategy that purports to end homelessness among families. The homeless assistance system does not provide immediate access to long-term subsidies for most families in shelters. The 2017 budget request aimed to target significant funding to families experiencing homelessness by supporting a significant expansion in the availability of rapid re-housing and Housing Choice Vouchers dedicated to families experiencing homelessness. In the absence of additional funding to support the expansion of affordable housing opportunities or a dedicated stream of new funding to support the provision of housing subsidies dedicated to families experiencing homelessness, communities must build stronger connections between the homeless assistance system and the mainstream housing assistance system to ensure that families experiencing homelessness have equal access to housing subsidies.

- ***If policymakers did not use evidence this time, it does not mean they never will***. The study demonstrates the ability of well-conducted, large-scale, multisite evaluations to generate national-scale conclusions. The study findings were suitable to develop a proposal for a large investment in the White House budget. Even though the final investment did not fit within the government-wide funding priorities at the time, the Family Options Study demonstrates the fundamental role that stable and affordable housing plays in improving child and family well-being, and it has provided clear, actionable evidence for future policy and research agendas in the ongoing work to end family homelessness. ◼

1   Stewart B. *McKinney Homeless Assistance Act*. Public Law 100-77, 1987.

2   U.S. Senate. Committee on Appropriations. *Transportation, Treasury, The Judiciary, Housing and Urban Development, and Related Agencies Appropriations Bill*. 109th Cong., 1st sess. 2008. S. Rep. 109-109.
     Available at: https://www.congress.gov/congressional-report/109th-congress/senate-report/109.

3   Daniel Gubits, et al. *Family Options Study: Short-Term Impacts of Housing and Services Interventions for Homeless Families*. U.S. Department of Housing and Urban Development, July 2015.

4   Daniel Gubits, et al. *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. U.S. Department of Housing and Urban Development, October 2016.
     Available at: https://www.huduser.gov/portal/publications/Family-Options-Study.html.

5   U.S. Department of Housing and Urban Development. "Landmark Family Options Study Has Clear Policy Implications." *Policy Development and Research Edge*.
     Available at: https://www.huduser.gov/portal/pdredge/pdr-edge-featd-article-110716.html.

6   U.S. Department of Housing and Urban Development. *Fiscal Year 2017 President's Budget Request for the U.S. Department of Housing and Urban Development*. 2016.
     Available at: https://archives.hud.gov/news/2016/pr16-160-23-HomeAssist_for_Families.pdf.

7   Nikita Stewart. "Obama Will Seek $11 Billion for Homeless Families." *The New York Times*, February 08, 2016.
     Available at: https://www.nytimes.com/2016/02/09/nyregion/obama-to-propose-11-billion-to-combat-family-homelessness.html.

8   Daniel Gubits, et al. *Family Options Study: 3-Year Impacts*, 72.

9. CUTTING HEALTH CARE COSTS

Innovations from Pioneering Accountability in Care

Mai Pham is the vice president of provider alignment solutions at Anthem Inc. She is also the former chief innovation officer for the Center for Medicare and Medicaid Innovation.

Melissa Cohen is the staff vice president of Payment Innovation at Anthem Inc. She is also a former group director at the Center for Medicare and Medicaid Innovation.
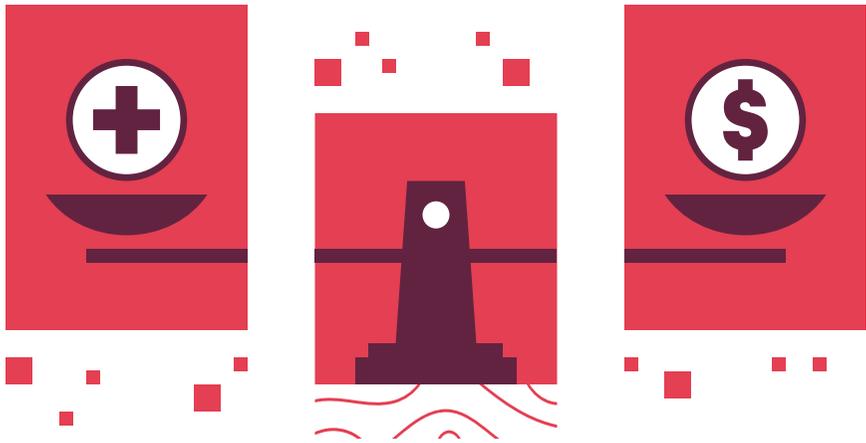
The United States has high health care costs, at least relative to other developed countries. In 2010, national spending on health care totaled $2.6 trillion, nearly 18 percent of the country's economic activity.[1] However, this spending comes without comparable gains in critical health and quality outcomes, such as life expectancy.[2]

The Patient Protection and Affordable Care Act (ACA, or Obamacare) was enacted in 2010 to address two key causes: the number of uninsured Americans and the costs of health care. The ACA included provisions to control the rise in health care spending while also improving the quality of patient care. One approach encouraged health care providers to organize activities in a way that focuses on the quality of the care delivered by taking a holistic look at the value of care. [3] This included the development of accountable care organizations (ACOs), which are groups of physicians and/or hospitals that take on the responsibility of the cost and quality of care for a patient population and move the system away from the historic fee-for-service model. ACOs are rewarded for managing chronic conditions and avoiding hospital admissions and other expensive and/or unnecessary services. The ACA established the Medicare Shared Savings Program (MSSP) as the permanent, national ACO program.

The ACA also restructured how the U.S. Department of Health and Human Services (HHS) went about testing new ways of paying for and delivering health care in order to determine which strategies were appropriate for larger-scale policy reforms. Together, Medicare and Medicaid provide government-subsidized health coverage to a large portion of the population, 47.2 million for Medicaid and 54.6 million for Medicaid as of 2010.[4] The launch of the Center for Medicare and Medicaid Innovation (CMMI) enabled HHS to test innovative payment techniques and service delivery models that could reduce costs and improve quality in the Medicare and Medicaid programs. With CMMI, the ACA gave HHS

a mechanism to partner with providers as it conducted these experiments.[5] One early test at CMMI, of the Pioneer Accountable Care Organization model, offers insights about how experienced medical providers supported innovative payment approaches and how the findings influenced changes to the MSSP.

ISSUE BACKGROUND

According to the Institute of Medicine about 30 percent of spending for health care goes to unnecessary services, administrative overhead, and other wasteful spending; in 2009, this amounted to $750 billion nationally.[6] Historically, doctors had limited incentives to reduce the costs of treatment because the core payment approach offered them fees for each service provided, which created incentives to focus on growing the volume of services rather than on providing quality outcomes or mitigating the total costs of care. The theory behind the ACO programs is that if given the opportunity to earn bonuses by reducing total spending and improving quality, providers will have more balanced incentives and will focus more on what is appropriate for their patients and less on billing for more and more services.

Medicare Shared Savings Program

In ACO programs, Medicare determines the population that an ACO will be accountable for and sets a spending target for that population. If at the end of the year, the ACO's population costs Medicare less than the spending target and meets specific quality standards, the ACO will receive a share of those savings. Conversely, if the ACO's population costs Medicare more than the spending target, that ACO may owe Medicare payment as a share of those losses. Most ACOs prefer to enter "shared savings only" contracts, which gives them limited risk exposure while they change their organizational culture and care processes to manage their population's health. More confident organizations may opt for the "shared savings and losses" contracts if they deem the financial opportunity large enough and the risk of losses manageable. Over time, HHS hopes that more ACOs will move to greater levels of risk, on the theory that greater accountability will drive better performance. In the first year of the MSSP, more than 100 organizations across the country participated, but fewer than 10 took on "shared savings and losses" contracts.[7]

Pioneer accountable care organizations

Parallel to the MSSP, CMMI designed the Pioneer ACO Model to test more ambitious and risky design elements, with a goal of determining which of those elements the MSSP should incorporate as permanent features in future years. Out of more than 70 applicants, the agency selected 32 Pioneer ACOs, including both hospital systems and large medical groups in urban and rural regions. Like the MSSP, the Pioneer ACO Model held providers to a financial target and quality scorecard. However, Pioneer was designed for providers already experienced in population health management and ready to take on financial risk for potential losses starting on day one. In exchange for taking on significant levels of financial risk, CMMI offered Pioneer ACOs a number of program features designed to enhance their chances of success.

CMMI staff assigned Pioneer organizations a set list of patients at the start of the year to begin focusing their care management efforts, rather than having a patient list that could change throughout the year as new patients came into Medicare and others sought care from different providers. Having this enhanced level of certainty helped Pioneer ACOs better target their investments in staff time to engage with complex patients and those with known chronic conditions.

CMMI also waived certain payment rules, like requiring a three-day stay in a hospital before Medicare would pay for a skilled nursing facility admission. This rule had long been in place to prevent waste in the system on the theory that providers might otherwise overuse skilled nursing facilities. Because ACOs were now holistically accountable for the cost and quality of care, CMMI hypothesized that the organizations would use the waiver responsibly, safely decreasing Medicare costs by avoiding unnecessary in-patient hospital stays for patients who were stable enough to go straight to a nursing facility instead. Testing these features required a certain level of trust that ACOs would not game or abuse the new legal flexibilities. But HHS also detailed expectations of the processes that organizations needed to set up in order to ensure that the waiver applied only to appropriate cases and that no harm came to patients. Plus, HHS would monitor their efforts.

EVIDENCE AVAILABILITY

HHS used multiple types of evidence to assess the Pioneer ACO Model. First, CMMI fielded a team of researchers who conducted a formal evaluation to answer two key questions: (1) Did the Pioneer ACO Model result in lower Medicare spending than would have otherwise occurred without harming the quality of care? and (2) Were there certain types of ACOs that were more likely than others to be successful?

Second, the Office of the Actuary within the Medicare agency played a critical role in reviewing the formal evaluation findings and conducting additional analyses to assess whether the program met criteria to qualify for expansion into permanent programs, as laid out in the ACA. In contrast to the formal evaluation, which looked retrospectively at what happened in the Pioneer ACO Model, the Office of the Actuary had responsibility for prospectively predicting the impact of program features. This required consideration of the likelihood of different behavioral responses by health care providers to specific circumstances.

Third, CMMI collected a great deal of qualitative and quantitative information on how the program performed operationally. It sponsored a learning system that brought ACO participants together to share their experiences and to exchange ideas on solving common problems, such as how to interpret data files or how to explain their transformative care work to patients. CMMI collaborated with Pioneer ACOs to create a curriculum of topics. One module, for example, brought together the ACOs that implemented the three-day-stay waiver to discuss the best way to partner with skilled nursing facilities, streamline the process of readying patients to go home, and develop useful measures for success. Through

these interactions and by creating new data tools, contract templates, legal instruments, and processes for collaboration, CMMI generated a wealth of operational lessons for the agency to consider in generalizing from the Pioneer ACO Model.

EVIDENCE USE

The independent evaluation demonstrated that Pioneer organizations saved Medicare a total of $384 million in 2011 and 2012.[8,9] A significant part of these savings came from Medicare beneficiaries using the hospital less and shifting care to doctors' offices. At the same time, the evaluation found that Pioneer organizations achieved a higher quality of care and sustained that performance throughout the evaluation period, confirming that savings did not come at the cost of reducing quality. The Office of the Actuary reviewed the evaluation findings and modeled additional analyses by using assumptions of the number of likely qualified organizations in the country, whether those organizations would begin as high-cost or low-cost providers, how strongly Pioneer program features would lure providers away from other types of contracts, and other key behavioral parameters.

Based on these analyses, the staff in the Office of the Actuary predicted that savings would accrue to the Medicare Trust Fund if Medicare's permanent ACO program, the MSSP, included specific Pioneer design elements, such as providing a patient list to the organizations at the beginning of the year and giving them the option of realizing a higher share of savings if they agree to accept increased financial risk. The Office of the Actuary presented its findings in April 2015, certifying the expectation of cost savings.[10] Based on the certification by the Office of the Actuary's staff, HHS decided to include these Pioneer ACO Model design elements into a new component of the MSSP through regulatory action.[11]

Based on CMMI's operational experience, HHS policy officials determined that the three-day-stay waiver did not result in adverse patient outcomes or questionable business practices by providers and was a cost-effective tool for case management in rehabilitation settings. Thus, it incorporated the three-day-stay waiver into the MSSP for organizations taking financial risk. HHS finalized these new MSSP features in June 2015, approximately four and a half years after the launch of the Pioneer ACO Model.

As important, HHS staff recognized that not all of the features tested through the Pioneer ACO Model were ready for large-scale adoption, but they could still be translated into program elements to test in a new ACO experiment. For example, CMMI learned that some organizations wanted to change how they receive Medicare payments: rather than a stream of payment for individual services with an additional payment of shared savings at the end of the year, some ACOs wanted their payments throughout the year to be lumped together into larger monthly payments, called "capitation," and calculated on a per-patient basis. Receiving more predictable and larger amounts at a steady pace would allow them to better plan their investments responsibly. ACOs also found that many (often the majority) of the patients assigned to them were not members of the population they were accountable for, which greatly

HHS STAFF RECOGNIZED THAT NOT ALL OF THE FEATURES TESTED THROUGH THE PIONEER ACO MODEL WERE READY FOR LARGE-SCALE ADOPTION, BUT THEY COULD STILL BE TRANSLATED INTO PROGRAM ELEMENTS TO TEST IN A NEW ACO EXPERIMENT.

diluted the value of the care interventions and presented missed opportunities to help patients in need. Therefore, they asked for the ability to approach patients CMMI may not otherwise have assigned to them to "join" their ACO.

CMMI applied these and other lessons into the design of an even more advanced program, the Next Generation Model, which began testing in 2016. Many organizations that had participated in the Pioneer ACO Model transitioned into the Next Generation Model, a testament to the value they found in working collaboratively with HHS to rapidly cycle through experimentation and learning.

LESSONS

- *Evidence use requires weighing many inputs*. In the case of determining how to proceed with the Pioneer model, the most important lesson was how to synthesize and give appropriate weights to all the inputs. The program team received feedback directly from the ACOs, the learning system modules, shared savings results, evaluation results, and even from members of Congress who received constituent questions. Nothing takes the place of a rigorous evaluation, but in government programs, multiple stakeholders and informational inputs need to be factored into policy decisions.

- *Clear goals help inform the evaluation questions and resulting policy decisions*. While the Pioneer ACO Model did demonstrate savings to the Medicare Trust Fund, when designing subsequent projects, a key question involved the importance of early savings compared with longer-term care delivery transformation and provider culture change. If HHS gave a significant portion of the savings realized back to the ACOs, could this encourage more investment in care transformation and more longer-term eventual savings? When executing a program, there will likely be tremendous pressure to show results—and quickly. Setting expectations and defining metrics for success at the outset can help mitigate some of these external pressures. ∎

1   The Henry J. Kaiser Family Foundation. *Health Care Costs: A Primer—Report.* May 2012.
    Available at: https://www.kff.org/report-section/health-care-costs-a-primer-2012-report/.

2   Elizabeth Docteur and Robert A. Bernenson. *How Does the Quality of U.S. Health Care Compare Internationally?*
    Timely Analysis of Immediate Health Policy Issues. Urban Institute, August 2009. Available at: https://www.
    urban.org/sites/default/files/publication/30596/411947-How-Does-the-Quality-of-U-S-Health-Care-Compare-
    Internationally-.PDF.

3   Shared Savings Program. U.S. Code 42 § 1395jjj.

4   U.S. Department of Health and Human Services. *2010 CMS Statistics.* Centers for Medicare and Medicaid Services,
    2010. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/
    CMS-Statistics-Reference-Booklet/2010.html.

5   Center for Medicare and Medicaid Innovation. U.S. Code 42 (2010), §§ 1315a.

6   Kaiser Health News. "IOM Report: Estimated $750B Wasted Annually in Health Care System." *KHN Morning
    Breakout*, September 07, 2012. Available at: https://khn.org/morning-breakout/iom-report/.

7   U.S. Department of Health and Human Services. "Shared Savings Program: Program Data." Centers for Medicare
    and Medicaid Services, last modified February 13, 2019. Available at: https://www.cms.gov/Medicare/Medicare-
    Fee-for-Service-Payment/sharedsavingsprogram/program-data.html.

8   L&M Policy Research. *Evaluation of CMMI Accountable Care Organization Initiatives: Pioneer ACO Evaluation
    Findings from Performance Years One and Two.* March 10, 2015.
    Available at: https://innovation.cms.gov/Files/reports/PioneerACOEvalRpt2.pdf.

9   David J. Nyweide, Woolton Lee, and Timothy T. Cuerdon. "Association of Pioneer Accountable Care
    Organization vs. Traditional Medicare Fee for Service with Spending, Utilization, and Patient Experience."
    JAMA, 313(21): 2152-2161, 2015.

10 Paul Spitalnic. *Certification of Pioneer Model Savings.* U.S. Department of Health and Human Services, Centers
    for Medicare and Medicaid Services, April 10, 2015. Available at: https://www.cms.gov/Research-Statistics-Data-
    and-Systems/Research/ActuarialStudies/Downloads/Pioneer-Certification-2015-04-10.pdf.

11 U.S. Department of Health and Human Services. "Medicare Program; Medicare Shared Savings Program:
    Accountable Care Organizations. Final Rule." *Federal Register*, 80(110): 32692-32845, June 9, 2015. Available at:
    https://www.federalregister.gov/documents/2015/06/09/2015-14005/medicare-program-medicare-shared-
    savings-program-accountable-care-organizations.

10. LET'S TALK ABOUT LUST

EPA's Underground Storage Tanks Program

Carolyn Hoskinson is the director of the Office of Underground Storage Tanks in the Environmental Protection Agency.

Mark Barolo is the deputy director of the Office of Underground Storage Tanks in the Environmental Protection Agency.

Karen Ellenberger is the outreach team leader of the Office of Underground Storage Tanks in the Environmental Protection Agency.
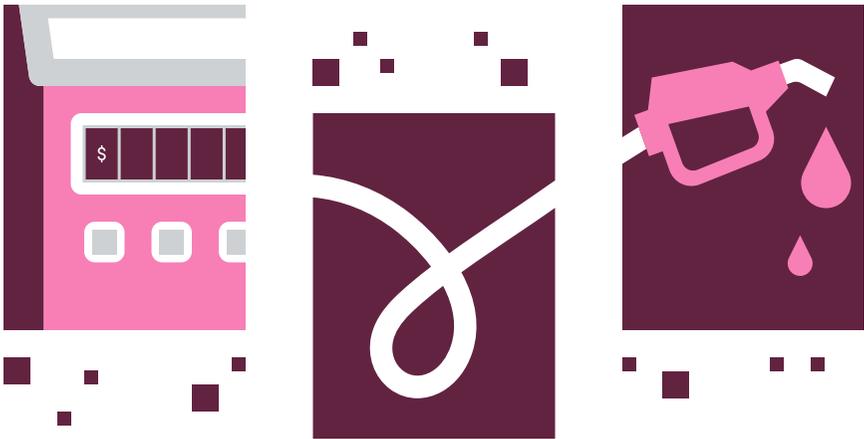
Now that we have your attention, full disclosure dictates that we tell you that LUST is an acronym for "leaking underground storage tanks," at least in the parlance of the Environmental Protection Agency (EPA). Not exactly what you expected? EPA's LUST work is important in protecting human health and the country's environment, and the program achieved meaningful reforms through an evidence and evaluation project, also known as a backlog study.

Going back more than three decades, in 1984 Congress and the president added Subtitle I to the Solid Waste Disposal Act. Subtitle I created a federal environmental program to regulate underground storage tanks (USTs) containing petroleum and certain hazardous substances, with a goal of limiting corrosion and structural defections, as well as minimizing future tank leaks. Subsequent legislative amendments in 1986 and 2005 created the Leaking Underground Storage Tank (LUST) Trust Fund to address petroleum releases and to expand eligible uses of the LUST Trust Fund to include leak-prevention activities.

In response to congressional direction, EPA created the Office of Underground Storage Tanks in 1985; at that time, there were approximately 2.1 million regulated UST systems in the United States. In 1988, EPA issued the country's first national UST regulation, and in 2015, it revised the 1988 regulation by adding additional prevention requirements and by ensuring that all USTs in the United States meet the same minimum standards.[1,2]

EPA's UST program's mission, which includes addressing leaking USTs, is to protect the country's environment and human health from UST petroleum releases. States and territories (hereafter referred to simply as "states"), EPA, and tribes accomplish this by working collaboratively with industry to prevent releases, to detect releases early, and to clean up releases. Billions of gallons of petroleum products are stored in USTs throughout the United

States, and releases from USTs are a major source of soil and groundwater contamination. The greatest potential threat from leaking USTs is contamination of groundwater, which is the source of drinking water for nearly half of all Americans.[3,4] A leaking UST can also present other health and environmental risks, including the potential for fire and explosion.

USTs are located in every community in the United States at retail facilities, such as service stations and convenience stores. At a typical service station, USTs hold tens of thousands of gallons of fuel. USTs are also located at non-retail facilities, such as municipal facilities for school buses, police, and fire stations; marinas; taxi-fleet facilities; postal and delivery service facilities; and federal facilities like military bases. As of September 2018, there were approximately 550,000 active USTs at approximately 200,000 facilities—all regulated by EPA's federal UST regulation. Because of the large size and great diversity of the UST-regulated community, states are in the best position to oversee USTs, and they are the primary implementers of the UST program. EPA has responsibility for and authority over USTs in Indian country, and we partner with tribes to implement the UST program there.

ISSUE BACKGROUND

Since the beginning of the national UST program, EPA collected and made data public from states regarding UST performance measures.[5] These data include information such as the number of active and closed tanks, confirmed releases, initiated and completed cleanups, facility compliance with UST requirements, and inspections. By subtracting the completed cleanups from the number of identified LUST releases, including newly confirmed UST releases reported, EPA determines the number of UST releases remaining to be cleaned up, known as the backlog.

One of EPA's performance measures—confirmed releases—provides data originated from UST owners or operators who identify releases from federally regulated petroleum UST systems and report the releases to states; then states confirm the releases. Fewer confirmed releases indicate better release prevention.

As of September 2018, EPA and states confirmed more than 543,000 releases from federally regulated USTs nationwide since the beginning of the national UST program. In the 1990s, EPA and states reported annual releases of between 25,000 to 66,000 releases (the highest ever in one year). Between 1989 and 1998—the first 10 years after issuing the 1988 federal UST regulation—EPA and states reported 367,000 releases, or 68 percent of all releases confirmed. Since then, a steady drop in the number of releases reported occurred. For the last 10 years, the number of releases detected each year stabilized, ranging from a high of 7,100 to a low of 5,500 per year; EPA and states reported a total of 55,000 releases between 2009 and 2018. For the last three years, EPA and states reported approximately 6,000 releases each year. The number of declining releases reported each year shows that EPA and state's prevention work is doing what it is supposed to do: helping to keep petroleum from contaminating the environment.

After confirming releases, EPA either direct responsible parties to initiate cleanup or EPA or the states directly begins the cleanup. The number of releases cleaned up each year varies. However, after the first 10 years of the UST program, cleanups have only decreased. In the 1990s, EPA and states averaged 23,300 cleanups per year; in the 2000s, averaged 15,500 each year; and for the last nine years, averaged 11,100 per year, with an average of 8,600 each year for 2016 through 2018. As of September 2018, the number of releases cleaned up nationwide since the beginning of the national UST program was approximately 478,000. That means 88 percent of all releases have been cleaned up, and 65,000 releases—or 12 percent—remain in the backlog of releases that still need to be cleaned up. Even though those remaining releases may be technically challenging, may lack funding for cleanup, or be abandoned, reducing the backlog remains a national UST program priority, and we are continuing steady progress on this effort.

During the 2000s, EPA increased its focus on performance measures and the White House Office of Management and Budget (OMB) pressed the UST program to show strong results in the number of releases cleaned up each year. This was partly due to the Government Performance and Results Act (GPRA) of 1993, which established strategic-planning and performance-measurement requirements in federal agencies. Because EPA directly implements an extremely small fraction of the releases cleaned up in the UST program, the bulk of the day-to-day oversight for cleaning up releases rests with state programs. EPA's 10 regional offices provide oversight of states, and the EPA's Office of Underground Storage Tanks provides national direction and oversight.

Each year, as part of the federal government's annual budget cycle, EPA negotiated with OMB to determine the national GPRA target for releases to be cleaned up. Once the national target was established, EPA worked with states to determine state targets. States were extremely concerned about setting targets that were achievable, because their performance success was strongly judged on whether or not they met their targets. Also, there was no incentive to agree to stretch goals if there was a possibility of missing them. We became engaged each year in divisive and unproductive debates as we attempted to negotiate these targets. Resentment, frustration, and disagreement bubbled up annually between staff in EPA and the states. Collectively, we spent a lot of time and energy negotiating and debating targets; that was time not spent working together toward completing cleanups or developing enhanced strategies to improve cleanups.

Without fail, each year states argued that all the easy cleanups were already done and only the difficult ones were left, meaning they could complete fewer and fewer each year. EPA had no data to either confirm or deny this claim, nor was there information to determine what were reasonable targets for each state to achieve. Yet, EPA was expected to lead states to meet or exceed national targets. Moreover, there was no incentive for OMB to lower the national target for this measure, given that EPA could provide no evidence to support why the target should be lowered.

EVIDENCE AVAILABILITY

EPA decided to identify and analyze what obstacles were preventing cleanup of the then approximately 100,000 releases. EPA also wanted to determine what opportunities existed to address those obstacles and develop strategies to bring more releases to closure. EPA referred to its analysis as the "2011 backlog study," and it provided a detailed, data-driven look at releases still in need of cleanup.[6]

Because states are the primary implementers of the UST program, EPA does not keep a database of releases that it could use in this analysis. States maintain and are the primary repositories of all site-specific data for the UST program; EPA maintains data about USTs in Indian Country, but that encompasses only about 0.5 percent of the federally regulated UST systems. As a result, we realized we needed to work with states to collect site-specific data on remaining releases to better determine the issues and opportunities. However, states structured their databases in ways that met their own needs, meaning their databases were quite distinct, with different data fields, terms, and levels of data quality.

Given the extraordinary time and effort required to collect and standardize this information from 56 states, territories, and Washington, D.C., and then create a dataset consistent enough to analyze, EPA instead decided to use a smaller subset of 14 states for the analysis. These states were responsible for 66 percent of the national backlog. EPA collected data from each of the 14 states, then worked with them to understand the data well enough to make consistent data points and analyze those points from a national perspective.

EPA published the results of the analysis in 2011. The study included a chapter on national findings and a state-specific chapter on findings for each state. EPA named the study *The National LUST Cleanup Backlog: A Study of Opportunities* to emphasize that the purpose of the analysis was to identify opportunities to move more releases to cleanup-completed status.[7] In the study, EPA carefully pointed out areas with data limitations or gaps. The analysis provided a critical understanding of the makeup of the backlog. It turns out the backlog is not homogeneous. Releases stayed in the backlog due to various factors, including a lack of money, technical cleanup challenges, abandoned tanks, and priority rankings. We identified unique strategies to address subsets of releases with different challenges. Understanding the data allowed us to identify which strategies would best address the variety of underlying issues.

EVIDENCE USE

Several states were concerned about the implications of being part of an analysis that defined them as having a "backlog" and the potential negative connotations of that moniker. Several states also resisted EPA pointing out that they might have inefficiencies or challenges in their programs since it might appear that their programs were unsatisfactory. Despite those concerns, several states examined the findings and realized there were aspects of their programs that merited further examination. Interestingly, some nonparticipating states found the results compelling enough that they initiated state-performed backlog studies of their own.

EPA regions used information from the analysis to achieve a greater level of specificity in discussions with states about possible state-specific strategies. Several states took the analysis, built on it, and developed strategies to clean up more releases. The bottom line is that the situation is much more nuanced than "all the easy releases are cleaned up and only the difficult ones remain."

Through the analysis, EPA identified a number of opportunities to potentially move releases through cleanup to completion. Those opportunities include: expediting site assessments, optimizing remedial design, applying risk-based decision-making, providing performance-based financial incentives to cleanup contractors, addressing orphan or abandoned releases, reviewing case files for releases close to closure, and reviewing case files for old releases with no activity.

One of the most significant results of the analysis was that, even for those intimately involved in the UST program, our perceptions of the program changed completely. We no longer present uninformed arguments, such as "You need to do more." And we no longer hear in reply, "We cannot do more." The analysis empowers us to now have more fruitful discussions about specific pockets of releases that meet certain descriptions and whether crafting specific solutions will move those releases toward cleanup and completion.

Overall, the UST program is now much more strategic in the work it does. After evaluating the evidence, EPA took specific actions that were an outgrowth of that information, and states took additional steps as well. For example, because we discovered that one of the factors forestalling cleanups was the technical complexity of some releases, EPA then developed and provided tools to help cleanup staff better address technically complex releases. Examples of tools that EPA developed and presented include webinars on high-resolution site characterization and guidance on how to clean up releases affected by petroleum-vapor intrusion. EPA is currently developing guidance for onsite remediation and is addressing common but complicated contaminants from LUST releases.

States also implemented specific strategies to address some of the identified obstacles to cleaning up releases.

- California implemented several policy initiatives to clean up more releases. One approach is called the "low-threat closure policy." The policy requires California to evaluate the risks of releases and, in cases where little to no risk is present, determine if the release can be closed even if contamination remains. For the first couple of years after California implemented the policy, it closed more than 1,000 releases per year. From 2012, when the policy was implemented, to 2018, California reduced its backlog of releases by more than half, from 7,703 to 3,128.

- Florida state law had required that money from state funds can only pay to clean up releases based on priority. But the highest priorities are often the most challenging and costly, limiting the money available to address additional releases. In 2011, Florida knew it had the highest backlog of releases in the country, and it changed its legislation to address more releases, including a voluntary program where lower-

risk releases can be assessed for low-cost closure. Since then, Florida increased the number of releases cleaned up per year, and in the past two years, it has cleaned up more than 850 releases per year. Florida reduced its backlog of releases from 13,507 in 2011 to 9,332 in 2018.

Clearly, EPA is seeing steady progress in the program's core metrics, and this is helping to produce positive environmental and health outcomes. Since the beginning of the study analyzing the backlog of UST releases in 2009 through the end of September 2018, EPA and state partners reduced the national backlog from approximately 100,000 to 65,000 releases remaining to be cleaned up.

LESSONS

- *Ask specific questions to get specific answers.* The better EPA decision makers understand the specific evidence about the UST program, the more effective and efficient we can be about working toward and achieving the program's mission of protecting the environment and human health from petroleum UST releases. In order to develop program strategies that succeed in making a difference, it is necessary to conduct an in-depth review of underlying data to understand the real issues behind the problem. We were so pleased with our progress and results that in 2014 we conducted a similar study, this time of UST releases in Indian Country. The goal was to determine how to more effectively address releases where EPA is directly responsible. In 2018, EPA updated that study and is now drawing from the findings to determine region-specific strategies to clean up more releases in Indian Country.

- *A domino effect of evidence use can occur.* Having found value from studies of the LUST backlog, we looked for other opportunities to use evidence-based analysis. Although not described in this case study, EPA performed two additional UST analyses that used evidence to better understand the UST program and its impacts. In partnership with EPA's National Center for Environmental Economics, the UST program studied the impacts of preventing and cleaning up UST releases on housing prices.[8] The study showed that, on average, there is a 3 to 6 percent depreciation when a high-profile UST release is discovered and a 4 to 9 percent appreciation after the UST release is cleaned up. Partnering with EPA's Office of Communications, Partnerships, and Analysis, the UST program studied the impacts of more frequent inspections on compliance rates at UST facilities in Louisiana.[9] Results from that study suggested that increasing inspection frequency from roughly every six years to every three years, as required under the Energy Policy Act, improved UST facility compliance by 11 percent. ■

1   U.S. Environmental Protection Agency. "Underground Storage Tanks (USTs): 1988 Underground Storage Tanks; Technical Requirements; Final Rule and Underground Storage Tanks Containing Petroleum—Financial Responsibility Requirements and State Program Approval Objective; Final Rule." 1988. Available at: https://www.epa.gov/ust/1988-underground-storage-tanks-technical-requirements-final-rule-and-underground-storage-tanks.

2   U.S. Environmental Protection Agency. "Underground Storage Tanks (USTs): Revising Underground Storage Tank Regulation—Revisions to Existing Requirements and New Requirements for Secondary Containment and Operator Training; Final Rule." 2015. Available at: https://www.epa.gov/ust/revising-underground-storage-tank-regulation-revisions-existing-requirements-and-new.

3   U.S. Environmental Protection Agency. "Chapter 6: Groundwater Quality." In: *2000 National Water Quality Inventory.* Office of Water, August 2002, 49.

4   U.S. Environmental Protection Agency. "Underground Storage Tanks (USTs): UST Performance Measures." Available at: https://www.epa.gov/ust/ust-performance-measures.

5   Ibid.

6   U.S. Environmental Protection Agency. *The National LUST Cleanup Backlog: A Study of Opportunities.* September 2011. Available at: https://www.epa.gov/ust/national-lust-cleanup-backlog-study-opportunities.

7   Ibid.

8   Dennis Guignet, Robin R. Jenkins, Matthew Ranson, and Patrick J. Walsh. *Do Housing Values Respond to Underground Storage Tank Releases? Evidence from High Profile Cases across the United States.* U.S. Environmental Protection Agency, National Center for Environmental Economics, Working Paper 16-03, March 2016. Available at: https://www.epa.gov/environmental-economics/working-paper-do-housing-values-respond-underground-storage-tank-releases.

9   Karen A. Sullivan and Achyut Kafle. *Do More Frequent Inspections Improve Compliance? Evidence from Underground Storage Tank Facilities in Louisiana.* U.S. Environmental Protection Agency, Office of Communications, Partnerships and Analysis, OCPA Working Paper 2017-05, May 2017. Available at: https://www.epa.gov/aboutepa/do-more-frequent-inspections-improve-compliance-evidence-underground-storage-tank.

11. ASSESSING CHILDREN'S HEALTH IN PUBLIC AND ASSISTED HOUSING

Lisa B. Mirel is the chief of the Data Linkage Methodology and Analysis Branch in the Division of Analysis and Epidemiology for the National Center for Health Statistics within the Centers for Disease Control and Prevention.

Irma Arispe is the director of the Division of Analysis and Epidemiology for the National Center for Health Statistics within the Centers for Disease Control and Prevention.

Veronica Helms is a social scientist for the Office of Lead Hazard Control and Healthy Homes in the U.S. Department of Housing and Urban Development.

Christine Cox is a consultant for the Division of Analysis and Epidemiology for the National Center for Health Statistics within the Centers for Disease Control and Prevention.
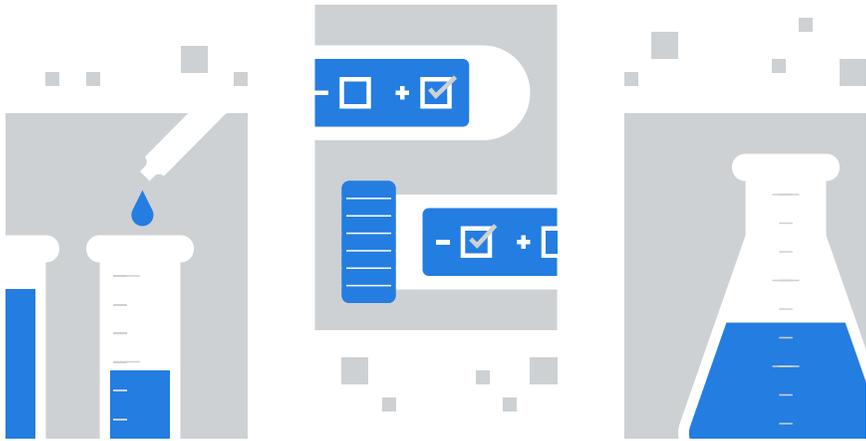
Lead exposure among children is associated with detrimental effects on development, such as damage to the brain and nervous system, slowed growth and development, learning and behavioral problems, and hearing and speech problems.[1,2,3] Beginning in the late 1970s, blood lead levels in the United States declined greatly as a result of policies aimed at the removal of lead from gasoline, residential paint, and, to a lesser extent, solder used in cans.[4,5,6] However, children are still at risk for exposure through ingestion of lead-based paint, which is particularly prevalent in older housing.[7]

Reducing the harmful effects of elevated blood lead levels in children is a key health objective for the U.S. Department of Health and Human Services (HHS) and the Centers for Disease Control and Prevention (CDC). The 2019 HHS budget affirms that "CDC will remain committed to the goal of eliminating elevated blood lead levels in children in the U.S. as a major public health problem by 2020."[8] Additionally, the HHS secretary is a co-chair on the President's Task Force on Environmental Health Risks and Safety Risks to Children, a committee that was established in 1997 by executive order to serve as the focal point for federal collaboration to promote and protect children's environmental health, including the prevention of lead poisoning.[9]

The U.S. Department of Housing and Urban Development (HUD) also prioritizes the reduction of elevated blood lead levels through the removal of lead-based paint hazards in at-risk housing units. In particular, HUD's strategic plan includes an objective to "Protect families from lead-based paint and other health hazards by making an additional 23,500 at-risk housing units lead-safe by the end of [fiscal year] 2019."[10] HUD's Office of Lead Hazard Control and Healthy Homes, which works to eliminate lead-based paint hazards, heads up this work.[11]

The National Center for Health Statistics (NCHS) and HUD collaborated to link survey and administrative data, as a means to assess the relationship between key health indicators and housing characteristics. The collaboration led to the development of new evidence used by HUD leaders to inform healthy homes policies and programs focusing on reducing elevated blood lead levels among children living in the United States. Healthy homes policies and programs focus on addressing how housing conditions can mitigate or exacerbate health.

ISSUE BACKGROUND

Substantial progress has been made over the past 40 years to reduce the number of children with elevated blood lead levels. These reductions reflect the impact of coordinated policies and programs, implemented across national, state, and local agencies, aimed at eliminating lead in vehicle emissions, paint, and consumer products marketed to children as well as reducing lead concentrations in housing, air, water, and ground.[12,13]

Lead exposure, however, continues to pose a health risk to children, especially among children residing in older housing stock, as these homes often contain deteriorating lead-based paint. Once the paint begins to deteriorate, chipping and flaking creates contaminated lead dust that becomes airborne and can be ingested. Residential lead dust exposure is highly correlated with elevated blood lead levels, which can lead to adverse health effects in children, including intellectual and behavioral deficits.[14,15] Lead-based paint hazards are present in an estimated 23 million U.S. homes, including 1.1 million homes of low-income families with one or more children under age 6.[16] In 2012, the CDC concurred with the Advisory Committee on Childhood Lead Poisoning Prevention that the primary prevention of lead exposure could best be accomplished by ensuring that all housing be lead-safe.[17]

Policymakers at HHS and HUD recognized the importance of the relationship between housing and health in the prevention of lead exposure among children, and that recognition is reflected in the strategic plans and budget proposals for both departments. Additionally, in December 2018, HUD, HHS, and the Environmental Protection Agency jointly released *The Federal Action Plan to Reduce Childhood Lead Exposures and Associated Health Impacts*, a comprehensive blueprint for reducing lead exposure and associated harms through collaboration among federal agencies with a range of stakeholders.[18]

To address the shared strategic interests of reducing lead exposure in those receiving federal housing assistance, and to evaluate the effectiveness of policies designed to reduce lead exposure, HUD approached NCHS with a request to link participants from NCHS population health surveys to HUD data resources on federal housing program participation.

NCHS was prepared to do this through its Data Linkage Program. The program is designed to link NCHS national health survey data with vital and administrative program data for eligible participants.[19,20] The resulting linked files create new and unique data resources that enable evidence-based policy evaluation and a deeper understanding of the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality. NCHS survey

data have been successfully linked to Medicare and Medicaid program data, Social Security disability and supplemental security insurance data, and cause and date-of-death information collected from the National Death Index. The linked data files are available to researchers through the NCHS Research Data Center.[21] Both NCHS and HUD worked collaboratively to securely combine person-level information from the National Health Interview Survey and the National Health and Nutrition Examination Survey with information about federal housing assistance from HUD.[22]

To undertake the desired linkage project, NCHS and HUD developed a memorandum of understanding to link NCHS survey data with HUD administrative records. The agreement included a commitment from both agencies to provide staff expertise to (1) assure high-quality linked data; (2) share recognition of the different mission of each organization and implications for the types of research questions, data analysis, and findings; (3) honor the different statutory requirements regarding data ownership, participant privacy, and data sharing present in each agency's data collections; and (4) define the conditions for subsequent use and access to the new linked data files.

NCHS and HUD strictly adhered to their applicable agency laws, policies, and procedures to protect the confidentiality of program and survey participants. NCHS is required to protect the handling and use of identifiable survey participant data according to the Privacy Act of 1974 and the Confidential Information Protection and Statistical Efficiency Act of 2002.[23] HUD was able to participate in the linkages because the HUD secretary has the authority to "undertake such programs of research, studies, testing, and demonstration relating to the mission and programs of the Department as he/she determines to be necessary and appropriate."[24] Once the linkage occurs at the individual level, the resulting linked files are also covered under the same NCHS confidentiality requirements. To analyze these new linked data resources, researchers must use the NCHS Research Data Center, a secure mechanism that provides researchers access to restricted-use data for jointly (NCHS and HUD) approved research projects. NCHS and HUD continue to work cooperatively to ensure that the linked data are only used in a manner consistent with the consent provided by program and survey participants.

EVIDENCE AVAILABILITY

HUD provides housing assistance to approximately 10 million low-income persons annually.[25] HUD's largest housing assistance program categories include housing-choice vouchers, multifamily programs, and public housing. Multifamily programs provide affordable housing through contracts with private owners of apartment buildings. Local public housing agencies manage housing-choice vouchers and public housing; they also oversee data collection and manage housing assistance program implementation. Public housing authorities own public housing, while the housing-choice voucher program gives tenants a voucher that covers part of their rent in a private-market unit.[26]

To administer its housing assistance programs, HUD collects information like household structure, household address, and detailed income information for all household members.

People living in HUD-assisted households are captured in HUD administrative data because they receive a rental subsidy or pay a below-market rent. Generally, a rental subsidy reduces gross housing costs for the tenant to about 30 percent of household income, although program rules may allow for substantial variations in that ratio.[27] While HUD administrative data contains information necessary to run its federally assisted housing programs effectively, they do not include information on participant health characteristics.

NCHS collects information on the health and well-being of the U.S. population through national surveys, some of which are fielded annually. The annual National Health Interview Survey is a nationally representative, cross-sectional household survey of 35,000 households in the United States.[28] The survey collects information on health status, access to and use of health services, health insurance coverage, behavioral health, and other health risk factors. The data helps researchers monitor trends in illness and disability and to track progress toward achieving national health objectives. The public health research community also uses the data to conduct epidemiologic and health policy analyses that characterize a wide variety of health conditions, to examine barriers to accessing and using appropriate health care, and to evaluate the effectiveness of federal health programs.[29]

The National Health and Nutrition Examination Survey is a nationally representative survey of the U.S. population, comprising about 5,000 persons from 15 different counties each year.[30] The survey assesses the health and nutritional status of adults and children in the United States through in-person interviews and physical examinations. The National Health and Nutrition Examination Survey interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel. HHS uses findings from this survey to determine the prevalence and risk factors of diseases and to set national standards for health measures, including height, weight, blood pressure, and other clinical measures, such as blood lead levels. National Health and Nutrition Examination Survey data are used in epidemiological studies and health-sciences research, which in turn help develop public health policies, direct and design health programs and services, and expand health knowledge for the nation.[31]

EVIDENCE USE

Researchers from both HHS and HUD successfully used the linked data to develop evidence through scientific research and analysis. The linkage allows for reliable national estimates of the prevalence of health conditions and health care utilization among adult and child participants receiving HUD assistance. HUD researchers published two reports that highlighted health characteristics of adults and children in HUD-assisted households. These two HUD reports represent a first step in securely sharing important health characteristics of participants receiving HUD assistance to provide insights about program operations.[32,33]

Additionally, the data linkage garnered support and interest from HUD leadership. Researchers provided presentations that highlighted key findings to the HUD secretary, HUD's assistant

secretary for policy development and research, and HUD's assistant secretary for public and Indian housing. Additionally, the HUD 2018-2022 Strategic Plan cited findings from the linked data as evidence to support the removal of lead-based paint hazards. The report states: "Analysis by HUD and the Centers for Disease Control and Prevention (CDC) of HUD tenant data linked with health survey data shows that children ages 1–5 who lived in HUD-assisted housing in 2005–2012 had lower lead blood levels than expected given their demographic, socioeconomic, and family characteristics, suggesting that HUD implementation of its lead hazard control regulations is effective in reducing exposure among children."[34,35]

Following the dissemination of this noted observational research finding based on linked National Health and Nutrition Examination Survey-HUD data, HUD proposed a new rule to further protect young children living in federally assisted housing by lowering the threshold for determining elevated blood lead levels in children living in assisted housing to match the standard used by the CDC. HUD's proposed "reference level" for lead in a young child's blood would be 75 percent lower (from 20 µg/dL to 5 µg/dL) and remain aligned with CDC recommendations in the future. This change to HUD's 17-year-old Lead Safe Housing Rule will allow for an earlier response when a child under 6 is exposed to lead-based paint hazards in their HUD-assisted homes.[36]

The availability of the NCHS-HUD linked data has allowed HUD leadership to pose other specific research questions to NCHS on topics where it needs more information. For example, senior leadership highlighted the high prevalence of learning disabilities among HUD-assisted children as a data point that merits further research. The linked NCHS-HUD data are also available to the public health research community through secure-access mechanisms and have resulted in several publications, including health characteristics (smoking and physical activity), health care access, and health insurance coverage among the housing-assisted population.[37,38,39]

The collaboration that led to linking data files has also fostered new cross-agency relationships between HUD and other agency partners, including:

- A collaboration between HUD and CDC partners to study childhood asthma among HUD-assisted children.

- A collaboration between HUD and the U.S. Department of Agriculture's Economic Research Service to examine rates of food insecurity among adults receiving rental assistance.

- A collaboration between HUD and the Administration for Community Living to examine disability and unmet health care needs among HUD-assisted older adults.

FOLLOWING THE DISSEMINATION OF THIS NOTED OBSERVATIONAL RESEARCH FINDING BASED ON LINKED NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY-HUD DATA, HUD PROPOSED A NEW RULE TO FURTHER PROTECT YOUNG CHILDREN LIVING IN FEDERALLY ASSISTED HOUSING BY LOWERING THE THRESHOLD FOR DETERMINING ELEVATED BLOOD LEAD LEVELS IN CHILDREN LIVING IN ASSISTED HOUSING TO MATCH THE STANDARD USED BY THE CDC.

LESSONS

- *Cooperation and coordination go a long way.* Bringing together data from population health surveys and administrative data from federal housing programs created new challenges and opportunities for both agencies. This project benefited from cross-agency leadership support and clearly defined data needs. Both agencies demonstrated a high level of cooperation and coordination throughout the process, including during the development of data-sharing agreements, data privacy, data exchange, and data analysis. Experts from both agencies were able to share expertise within their own data systems, which in turn helped to ensure accurate study findings, including an open assessment of each data system's limitations. Although each agency has its own statutory requirements regarding data confidentiality and privacy protections, cooperation and intent to achieve success resulted in a unique and high-quality integrated data resource.

- *Privacy and confidentiality can be assured when sharing data*. By using the NCHS Research Data Center as the secure-access mechanism for researchers' use of the linked data, both agencies can ensure that processes and procedures for accessing data continue to safeguard the privacy and confidentiality of program and survey participants. These safeguards include defining the process for evaluating and approving research proposals and statistical review of all researcher-generated analyses to assess disclosure risks.[40]

- *Combining data can produce valuable insights.* This was the first-time population health indicators from survey participants were linked to administrative data on federal housing assistance. Based on empirical evidence generated from this new and unique data source, HUD was able to promote policy changes to further safeguard the health of children living in federally assisted housing. Given this recent success, both agencies are keen to continue collaboration and expect to have an updated linked file released in the spring of 2019. ■

1   Bruce P. Lanphear, et al. "Low-Level Environmental Lead Exposure and Children's Intellectual Function: An International Pooled Analysis." *Environmental Health Perspectives*, 113(7): 894-899, 2005.

2   U.S. Environmental Protection Agency. *Integrated Science Assessment (ISA) for Lead (Final Report)*. 2013. Available at: https://www.epa.gov/isa/integrated-science-assessment-isa-lead.

3   U.S. Centers for Disease Control and Prevention. "Learn More about CDC's Childhood Lead Poisoning Data." Last modified May 4, 2016. Available at: https://www.cdc.gov/nceh/lead/data/learnmore.htm.

4   U.S. Environmental Protection Agency. Integrated Science Assessment (ISA) for Lead (Final Report). 2013. Available at: https://www.epa.gov/isa/integrated-science-assessment-isa-lead.

5   Federal Interagency Forum on Child and Family Statistics. *America's Children: Key National Indicators of Well-Being 2017*. U.S. Government Printing Office, 2017.

6   J.L. Pirkle, et al. "The Decline in Blood Lead Levels in the United States. The National Health and Nutrition Examination Surveys (NHANES)." *JAMA*, 272(4): 284-291, 1994.

7   F.G. Dewalt, et al. "Prevalence of Lead Hazards and Soil Arsenic in U.S. Housing." *Journal of Environmental Health*, 78(5): 22-29, 2015.

8   U.S. Department of Health and Human Services. *Fiscal Year 2019 Budget in Brief*. 2018.

9   National Institute of Environmental Health Sciences. "President's Task Force on Environmental Health Risks and Safety Risks to Children: About." Last modified August 30, 2017. Available at: https://ptfceh.niehs.nih.gov/about/index.htm.

10 U.S. Department of Housing and Urban Development. *HUD Strategic Plan 2018-2022*. 2018.

11 Ibid.

12 J.L. Pirkle, et al. "The Decline in Blood Lead Levels in the United States. The National Health and Nutrition Examination Surveys (NHANES)." *JAMA*, 272(4): 284-291, 1994.

13 W. Wheeler and M.J. Brown. "Blood Lead Levels in Children Aged 1-5 Years—United States, 1999-2000." *Morbidity and Mortality Weekly Report*, 62(13): 245-248, 2013.

14 B.P. Lanphear, et al. "The Contribution of Lead-Contaminated House Dust and Residential Soil to Children's Blood Lead Levels. A Pooled Analysis of 12 Epidemiologic Studies." *Environmental Research*, 79(1): 51-68, 1998.

15 Agency for Toxic Substances and Disease Registry. *Toxicological Profile for Lead*. U.S. Department of Health and Human Services, 2007.

16 U.S. Department of Housing and Urban Development. *HUD Strategic Plan 2018-2022*. 2018.

17 U.S. Centers for Disease Control and Prevention. *CDC Response to Advisory Committee on Childhood Lead Poisoning Prevention Recommendations in "Low Level Lead Exposure Harms Children: A Renewed Call of Primary Prevention."* 2012.

18 President's Task Force on Environmental Health Risks and Safety Risks to Children. *The Federal Action Plan to Reduce Childhood Lead Exposures and Associated Health Impacts*. December 2018.

19 NCHS's Research Ethics Review Board approved the linkage, which only applied to eligible NCHS survey participants. Only NCHS survey participants who have provided consent as well as the necessary personally identifiable information are linkage-eligible.

20 National Center for Health Statistics. *NCHS Data Linkage Activities*. Available at: https://www.cdc.gov/nchs/data-linkage/index.htm.

21 Information about accessing datasets through the RDC, *see*: U.S. Centers for Disease Control and Prevention. "Research Data Center." Available at: https://www.cdc.gov/rdc/index.htm.

22 Patricia C. Lloyd, et al. "Linkage of 1999-2012 National Health Interview Survey and National Health and Nutrition Examination Survey Data to U.S. Department of Housing and Urban Development Administrative Records." *Vital Health Statistics*, 1(60): 1-40, October 2017.

23 *Privacy Act of 1974*. U.S. Code 5 § 552a, *Paperwork Reducation Act*. U.S. Code 44 § 3501, Section V, and *Confidential Information Protection and Statistical Efficiency Act of 2002*. U.S. Code 44 § 101.

24 Research and Demonstrations; Authorization of Appropriations; Continuing Availability of Funds. U.S. Code 12 § 1701z–1.

25 U.S. Department of Housing and Urban Development and Office of Policy Development and Research. *Picture of Subsidized Housing: 2018*. 2018. Available at: https://www.huduser.gov/portal/datasets/assthsg.html.

26 Patricia C. Lloyd, et al. "Linkage of 1999-2012 National Health Interview Survey and National Health and Nutrition Examination Survey Data to U.S. Department of Housing and Urban Development Administrative Records." *Vital Health Statistics*, 1(60): 1-40, October 2017.

27 P.C. Lloyd and V.E. Helms. *NCHS-HUD Linked Data: Analytic Considerations and Guidelines*. National Center for Health Statistics, March 2016.

28 "Detailed descriptions of the National Health Interview Survey sample design have been described elsewhere. See: V.L. Parsons, et al. "Design and Estimation for the National Health Interview Survey, 2006-2015." *Vital Health Statistics*, 2(65): 1-53, 2014.

29 U.S. Centers for Disease Prevention. "National Center for Health Statistics: About the National Health Survey." Last modified January 16, 2019. Available at: https://www.cdc.gov/nchs/nhis/about_nhis.htm.

30 Detailed descriptions of the National Health and Nutrition Examination Survey sample design have been described elsewhere. See: C.L. Johnson, et al. "National Health and Nutrition Examination Survey: Sample Design, 2011-2014." *Vital Health Statistics*, 2(162): 1-33, 2014.

31 U.S. Centers for Disease Prevention. "National Center for Health Statistics: About the National Health and Nutrition Examination Survey." Available at: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm.

32 Veronica E. Helms, Jon Sperling, and Barry L. Steffen. *A Health Picture of HUD-Assisted Adults, 2006-2012*, by U.S. Department of Housing and Urban Development, Office of Policy Development and Research, March 2018.

33 U.S. Department of Housing and Urban Development and Office of Policy Development and Research. A Health Picture of HUD-Assisted Children, 2006-2012. March 2018.

34 U.S. Department of Housing and Urban Development. HUD Strategic Plan 2018-2022. 2018.

35 K.A. Ahrens, et al. "Housing Assistance and Blood Lead Levels: Children in the United States, 2005-2012." *American Journal of Public Health*, 106(11): 2049-2056, 2016.

36 U.S. Department of Housing and Urban Development. The Lead-Safe Housing Rule, 24 C.F.R. § 35. 2017. Available at: https://www.hud.gov/program_offices/healthy_homes/enforcement/lshr.

37 V.E. Helms, B.A. King, and P.J. Ashley. "Cigarette Smoking and Adverse Health Outcomes Among Adults Receiving Federal Housing Assistance." *Preventative Medicine*, 99: 171-77, 2017.

38 M.S. Wong, E.T. Roberts, C.M. Arnold, and C.E. Pollack. "HUD Housing Assistance and Levels of Physical Activity Among Low-Income Adults." *Preventing Chronic Diseases*, 15(94): 1-12, 2018.

39 Alan E. Simon, et al. "HUD Housing Assistance Associated With Lower Insurance Rates and Unmet Medical Need." *Health Affairs*, 36(6): 1016-1023, 2017.

40 Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitation Methodology.* Statistical Policy Working Paper No. 22. 2005.

12. LEVERAGING RESEARCH TO SUPPORT THE NATION'S FOOD SAFETY NET

Patricia Bailey is a senior program analyst in the WIC Program Integrity and Monitoring Branch in the Food and Nutrition Service of the U.S. Department of Agriculture.

Amy Herring is the chief of the WIC Program Integrity and Monitoring Branch in the Food and Nutrition Service of the U.S. Department of Agriculture.

Elizabeth Barrett is a senior program analyst in the WIC Program Integrity and Monitoring Branch in the Food and Nutrition Service of the U.S. Department of Agriculture

Sarah Widor is the director of the Supplemental Food Programs Division in the Food and Nutrition Service of the U.S. Department of Agriculture.

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) provides healthy foods, infant formula, health care referrals, and nutrition education to low-income, nutritionally at-risk pregnant and post-partum women, infants, and children under 5.[1] The U.S. Department of Agriculture's (USDA) Food and Nutrition Service (FNS) administers WIC, which was launched as a pilot program in the 1970s. FNS provides grants to state agencies, usually health departments, to fund service delivery.

State agencies use cost containment strategies to keep food costs relatively low, and these strategies, coupled with annual program funding, have enabled WIC to serve all eligible individuals seeking program benefits over the past several years. In fiscal year 2017, WIC served approximately 7.3 million people, or participants, each month. In the same year, WIC provided $3.6 billion worth of nutritious food to participants. WIC provides foods that meet the specific nutritional needs of participants, including milk, eggs, whole grains, fresh fruits and vegetables, and more. In addition, WIC also provided $2 billion to state agencies to operate the program and to provide nutrition education, including breastfeeding promotion and support.

WIC is a cost-effective program that improves the health and nutrition of low-income women, infants, and children. WIC makes moms healthier, results in longer and safer pregnancies, and improves dietary outcomes for infants and children.[2]

Beyond serving WIC participants, the program supports the broader community through its grocery store, or vendor, authorization policies. The presence of WIC-authorized vendors improves the availability and variety of healthy foods, which means that all shoppers have greater access to fruits, vegetables, and whole grains, regardless of whether they participate in the program or not.[3]

This case study will discuss how FNS identified a nationwide WIC challenge, employed strategically designed research, and developed actionable guidance to improve state agency compliance with federal regulations.

ISSUE BACKGROUND

WIC prescribes participants specific foods designed to meet their nutritional needs during critical periods of growth and development. For example, a pregnant mom, her 10-month-old baby, and her 4-year-old son each receive different WIC foods since they have different needs.

To purchase WIC foods, most participants use either paper checks or electronic benefit transfer (EBT) cards that specify the amounts and types of foods prescribed. Participants shop for foods at WIC-authorized vendors and pay using their WIC checks or EBT cards. Vendors then request reimbursement for those foods from the WIC state agency.

With the exception of the cash-value benefit, which allows participants to purchase fresh fruits and vegetables, WIC provides specific amounts and types of food. As a result, participants do not have to worry about the cost of WIC foods as they shop. In order to ensure that this does not result in high costs to the program, WIC regulations require state agencies to have cost containment policies and procedures in place to mitigate two types of costs:

- **Food costs**: State agencies select foods for WIC authorization from among those that meet federal nutritional standards. During this process, they consider food costs and, for example, may choose to authorize generic brands or enter into rebate contracts with manufacturers. WIC regulations require state agencies to have a rebate contract in place for infant formula. In fiscal year 2017, infant formula rebates saved the program $1.7 billion.

- **Vendor costs**: State agencies are responsible for all aspects of vendor management, including selecting, authorizing, and overseeing vendors. They must select and authorize vendors with competitive prices and ensure that vendors charge WIC appropriately for foods.

The Child Nutrition and WIC Reauthorization Act of 2004, and the subsequent WIC Vendor Cost Containment Final Rule, published in 2009, made significant updates to vendor cost containment requirements.[4,5] Some of these include:

- Grouping vendors based on common characteristics that affect food prices. These groupings are known as "vendor peer group systems." Each state agency must assess its system at least every three years to ensure it remains effective.

- Ensuring that vendors charge competitive prices for food.

- Establishing price limitations (that is, the maximum price a vendor may be reimbursed for food items sold to WIC participants).

Together, these three requirements promote good stewardship of taxpayer dollars by providing a framework for state agencies to use as they contain vendor costs, thus protecting the program's ability to serve all eligible women, infants, and children who apply.

In order to ensure that vendor cost containment systems are effective, state agencies must develop and implement statistically sound vendor peer group systems. This involves conducting complex statistical analyses to determine which factors play the largest roles in vendor pricing in the state. After conducting these analyses, state agencies may choose to group vendors by size, location, ownership structure, or other factors.

By grouping vendors together based on common characteristics that affect food prices, state agencies can make informed decisions about how much each store should charge for WIC foods. Vendor peer group systems allow the state agency to compare one vendor's prices with the prices of similar vendors, or "peers." For example, a state agency could use its peer group system to determine whether a small, family-owned corner store's price for WIC foods is similar to other small stores and make WIC vendor authorization decisions accordingly. A state agency could also determine the maximum price it is willing to pay that store for a carton of eggs based on the prices at similar stores.

By allowing for peer-to-peer comparisons, vendor peer group systems create a straightforward method for determining whether a vendor is competitively priced and for establishing price limitations. Authorizing competitively priced vendors and paying them appropriately are critical to containing program costs. Ineffective vendor peer group systems undermine vendor cost containment efforts, weakening the state agency's ability to manage costs. For example:

- If a state agency cannot reliably identify competitive prices, it could authorize vendors with unreasonably high prices. This could inflate food costs.

- If a state agency sets price limitations so low that competitively priced stores cannot make a profit from WIC sales, they might stop participating in WIC. This could reduce participant access to WIC foods.

In order to ensure that vendor peer group systems operate as intended over time, WIC regulations require state agencies to periodically assess them for continued effectiveness and to make appropriate changes.[6]

In recent years, FNS reviewed state agencies' WIC vendor cost containment policies and procedures for compliance with federal requirements and to identify areas that could be improved. FNS found that many state agencies had not assessed the effectiveness of their vendor peer group systems, as required, in order to make appropriate changes. Because peer group systems serve as the foundation of other important vendor cost containment requirements, FNS determined that follow-up efforts must focus on improving state agency compliance with vendor peer group system requirements.

EVIDENCE AVAILABILITY

To initiate this effort, FNS held discussions with various state agencies to learn about the barriers to compliance with vendor peer group system requirements. These state agencies noted that the required assessment is a very resource-intensive process that requires complex statistical analyses. Many reported that they did not have access to the resources and staff necessary to complete statistical tasks, nor did they have the expertise to interpret or use statistical analyses conducted by outside entities. Even state agencies with access to relevant resources raised questions about how to conduct and use a vendor peer group assessment. Most importantly, state agencies reported that FNS guidance was outdated and ineffective because it did not reflect recent changes in the program and retail market, such as the transition to EBT.

Based on state input, FNS decided to update guidance on vendor peer group systems. FNS contracted with the Altarum Institute to study and outline a statistically sound methodology that state agencies could use to assess and develop effective vendor peer group systems. Researchers identified vendor characteristics that affect pricing and determined which characteristics most WIC state agencies could use. FNS required the researchers to consider state diversity in terms of geography, size, and the numbers and types of vendors authorized, and to develop a methodology that could be tailored to meet each state's needs.

The study identified three common and easily defined vendor characteristics that generally predict vendor prices: (1) store type, (2) geography, and (3) number of cash registers. The WIC peer group report describes these three characteristics, along with how state agencies can use them to assess and update their systems.[7]

While the study did not change vendor cost containment requirements, it provided an evidence-based methodology for conducting a vendor peer group assessment and using it to make system improvements. After reviewing the study, many state agencies reported that they understood and agreed with the approach as presented but still lacked the resources necessary to adopt the framework.

EVIDENCE USE

FNS set out to assist states with implementation and determined that the best approach was to develop additional resources for state agencies to use to adopt the methodology recommended in the study. FNS worked with the researchers to develop a practical vendor peer group toolkit, which clearly outlines the step-by-step process of conducting a vendor peer group assessment and developing vendor peer groups using data available to state agencies. The toolkit includes step-by-step lesson plans, sample datasets and Excel worksheets, and video demonstrations.

The toolkit empowers WIC state agencies with limited resources to independently conduct statistically sound vendor peer group assessments and to update their peer group systems so that they can be used, as intended, as the foundation of their vendor cost containment systems.

While a few states completed their most recent vendor peer group assessments in close collaboration with FNS staff, many others have reported that the materials made it possible for them to complete the assessment on their own. As a result, they are able to comply with vendor cost containment requirements and ensure that their vendor peer group systems are working as intended.

State agencies reported that the study's recommended peer group characteristics and methodology were helpful and greatly reduced the resources necessary to conduct an assessment and update their peer groups. One WIC state agency reported that, after using the toolkit as a starting point, it was inspired to tap into additional resources to complete an even more robust analysis. Notably, a few state agencies have reported using the toolkit to proactively assess their peer group systems ahead of schedule or to enhance vendor management after EBT implementation.

Additionally, the toolkit serves as a training resource to support FNS staff tasked with providing technical assistance to state agencies. In 2017, FNS headquarters used the toolkit to provide training to regional office staff who work directly with states. In turn, regional office staff used the toolkit to provide training and technical assistance to their respective states.

The robust data analysis included in the study led decision-makers at FNS to develop a toolkit to support effective policy implementation and, consequently, additional data analysis at the state level. This course of action demonstrates the range of possibilities available to entities seeking to support effective policy implementation.

LESSONS

- *Research does not have to end at finding the problems.* FNS' strategic use of research around WIC vendor peer group systems resulted in one of the most comprehensive pieces of WIC guidance to date. Through the development of the toolkit, FNS was able to illustrate *how* to accomplish a task, instead of simply outlining *what* must be done. The success of this effort can be a guide to ensure that future research projects have actionable results and create user-friendly tools for states to use as they work to implement eventide-based recommendations."

- *Policy implementation can be improved when using evidence and resources.* FNS continues to look for ways to leverage innovative ideas and data-driven strategies to improve WIC program integrity and provide excellent customer service to state partners. The combination of designing the study to offer practical solutions and the development of the vendor peer group toolkit has made a positive impact on WIC state agencies' ability to comply with related requirements and to build their vendor cost containment systems on a strong foundation. State agencies use of the toolkit has resulted in increased compliance with federal requirements, improved vendor peer group systems for several state agencies, and increased confidence in both FNS and state staff responsible for keeping WIC vendor cost containment systems up to date and working as intended. ■

1   WIC provides services to children until their 5th birthday.

2   U.S. Department of Agriculture. "About WIC-WIC's Mission." Last modified February 14, 2018. Available at: https://www.fns.usda.gov/wic/about-wic-wics-mission.

3   T. Andreyeva, et al. "Positive Influence of the Revised Special Supplemental Nutrition Program for Women, Infants, and Children Food Packages on Access to Healthy Foods." *Journal of the Academy of Nutrition and Dietetics*, 112(6): 850-858, 2012.

4   *Child Nutrition and WIC Reauthorization Act of 2004*. Public Law 108-265, 2004.

5   "Food Delivery Methods." 7 CFR 246.12.

6   7 CFR 246.12.(g)(4)(ii)(C).

7   Stacy Gleason, et al. *WIC Vendor Peer Group Study: Final Report and State Agency Guidelines*. Altarum Institute, May 2017. Available at: https://fns-prod.azureedge.net/sites/default/files/ops/WICVendorPeerGroup.pdf.

POLICYMAKING WITH EVIDENCE IN STATE AND LOCAL GOVERNMENTS
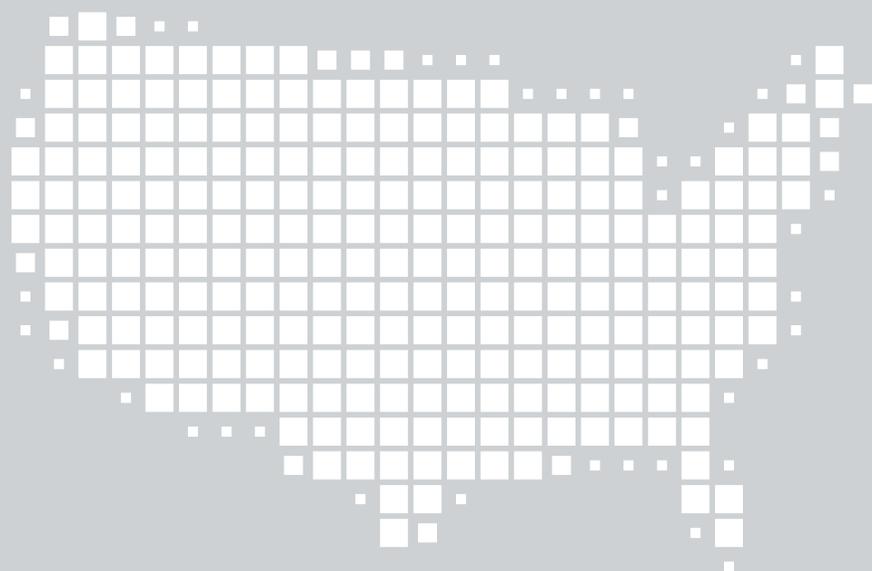
13. MAKING STREET-LEVEL BUREAUCRACY WORK

Safer Food in Seattle and King County

Daniel Ho is the William Benjamin Scott and Luna M. Scott Professor of Law, Professor (by courtesy) of Political Science, and Senior Fellow at the Institute for Economic Policy Research at Stanford University. He served as a co-project lead for the Food Safety Inspection for King County.

Becky Elias was the manager of the food program in the public health department in Seattle, King County. She also served as a co-project lead.

Every year foodborne pathogens make 15 percent of the American public ill and lead to 128,000 hospitalizations.[1] The Government Accountability Office classifies the U.S. food safety system as "high risk" due to "inconsistent oversight, ineffective coordination, and inefficient use of resources."[2] Food safety inspectors play a critical role in protecting the public and are housed at the Food Safety and Inspection Service, the Food and Drug Administration, and local health departments. While there has been a popular push for greater consumer disclosure of inspection results, there is also widespread concern that enforcement efforts are inconsistent across and within departments.

Improving the effectiveness of food safety enforcement is critical for protecting public health, including efforts to ensure permitted businesses comply with food safety regulations. This challenge of quality and consistency is also not unique to food safety. Agencies such as the Patent and Trademark Office, the Nuclear Regulatory Commission, the Social Security Administration, and the Board of Veterans Appeals all struggle with similar challenges of quality and consistency in applying, enforcing, and administering the law. Quality-improvement initiatives that work therefore have implications for many other areas of governance characterized by the decentralized administration of complex bodies of law.[3,4]

ISSUE BACKGROUND

In 2001, King County was the first county in Washington State to make food safety inspections available to the public on the internet, responding to demands for increased transparency. Twelve years later, county residents organized a petition with 2,000 signatures, calling for an improved food safety rating system that would include still-greater

transparency, this time at the point of purchase. With increased public and media scrutiny, the county took steps to require restaurants to publicly display a summary rating from food inspectors so that consumers can make more informed dining decisions.

Stakeholders offered a series of recommendations during public meetings about the King County food safety system:

1. The system should use more than a single inspection to determine ratings. This was perceived as fairer and more informative than existing systems, overcoming conventional critiques of inspections as providing only a single snapshot in time.[5,6]

2. Ratings should convey relative performance. King County only permits businesses to be open when they meet a minimum safety standard; stakeholders perceived the rating should convey how well a business performs beyond meeting the minimum standards and how businesses fare relative to one another.

3. The rating system should consider inspection consistency. Stakeholders in the community, including restaurant operators and inspection staff, focused on consistency in light of perceptions and evidence of variability in inspection style and stringency, both in King County and other jurisdictions.

EVIDENCE AVAILABILITY

When King County's placarding efforts began, there was limited evidence to inform regulatory design and implementation. One study examined grading systems in 17 large jurisdictions, finding substantial variability in the design of such systems. Collecting inspection data from 10 jurisdictions, the study identified (a) substantial inconsistencies in how the same establishment was scored over time, and (b) substantial evidence of grade inflation after the adoption of grading.[7]

Some observational studies found health benefits comparing jurisdictions before and after adoption of restaurant grading.[8,9,10] The leading study of Los Angeles was confounded by the state's largest salmonella outbreak in Southern California occurring before Los Angeles implemented restaurant grading.[11] There is also mixed evidence about whether food safety inspection scores are correlated with foodborne illness outbreaks.[12,13,14]

With limited evidence available, King County was the first jurisdiction in the state to implement restaurant grading with a rigorous evaluation plan in mind.

Developing a stronger evidence base

In 2014, the Public Health department in Seattle and King County engaged what is now the Regulation, Evaluation, and Governance Lab (RegLab) at Stanford University to support the evaluation plan. RegLab is a research laboratory whose mission is to promote evidence-based regulatory policy and administration through rigorous demonstration projects using

data science and technology. RegLab and King County formed a collaborative agreement to develop and improve the evidence base for food safety enforcement. King County's food program is responsible for health inspections of the county's 11,000 permitted establishments. Three projects emerged out of this collaboration.

Evaluation of peer-review pilot program

Building on information from existing academic literature, the county and RegLab designed a pilot program for health inspectors to engage in peer review process for conducting inspections. Existing research suggested reviewing each other could improve the accuracy and consistency of inspections.[15] Using an experimental evaluation design, researchers randomly assigned half the health inspectors into the four-month trial program. In the program, inspectors spent one day each week visiting establishments with peers, independently scoring health code violations, and comparing differences. Surprisingly, inspectors cited major code items differently in 60 percent of inspections, even when observing identical conditions and agreeing on food risks. For a food safety rating system, these citation differences could affect restaurant inspection scores and summary ratings. As a result of the variations in citations, the county also developed standards for high-prevalence and inconsistently scored infractions and trained inspectors to improve their overall consistency. The evaluation results concluded that the peer-review program caused (a) a 17 to 19 percent increase in violations detected, and (b) an improvement in inter-inspector consistency based on independent inspections.[16]

Design of rating system

While peer review improved the consistency of inspections, substantial inter-inspector variability remained. To further improve consistency as part of the rating design, the partners used results from the peer-review evaluation and retrospective observational data to inform the rating methodology.[17] First, inspectors were more likely to consistently cite businesses with "critical" violations—those more directly linked to foodborne illness—during peer-review inspections. Second, analysis of observational data suggested that repeat violations did not substantially predict future performance. Third, to determine the number of inspections of a business before assigning a grade, the partners analyzed how well historical information predicted performance. While each additional inspection provided better predictions, the gains were more limited after four inspections.

Policy evaluation of rating system

RegLab developed an experimental evaluation to assess the health impacts of the new rating system as part of its implementation. Public Health agreed to roll out the grading system in phases, which both facilitated operational implementation and the evaluation.[18] In addition, due to concerns about how inspectors would issue citations as a result of having a grading system in place, an independent team of contractor inspectors, who were trained in part by the Food and Drug Administration, observed risk factors in establishments without delivering any inspection report to operators. This allowed for direct observation for

AS A RESULT OF BROAD ENTHUSIASM FOR THE PEER-REVIEW PROGRAM, THE FOOD PROGRAM MANAGEMENT TEAM INSTITUTED PEER REVIEW AS PART OF AN ONGOING QUALITY-ASSURANCE PROGRAM … PEER REVIEW AND OBSERVATIONAL EVIDENCE ALSO INFORMED THE DEPARTMENT'S GRADING-SYSTEM DESIGN.

risk-factor prevalence in food-handling practices, independent of the rating system. Public Health also designed a qualitative assessment through community-based participatory research to receive feedback from stakeholders.

Although early in the evaluation process, no appreciable effects were identified from the rating system on foodborne illnesses, foodborne hospitalizations, or risk factors after eighteen months into the roll-out. The rating system may have increased public engagement in the form of submission of complaints in the short run. On the other hand, inspectors decreased citation of critical violations in a way not reflected by noncritical violations of the risk-factor study, suggesting that the rating system has a direct deleterious effect on inspector stringency.

EVIDENCE USE

As a result of broad enthusiasm for the peer-review program, the food program management team instituted peer review as part of an ongoing quality-assurance program for all food safety inspectors on a monthly or bimonthly basis. In other words, the control group from the evaluation was brought into the peer-review process, and management continues to use reports from peer-review inspections to target training materials.

Peer review and observational evidence also informed the department's grading-system design. Researchers calculated ratings based on the relative performance of an establishment of average critical violations over the past four routine inspections. Researchers placed no special weight on repeat violations given lack of predictive power. They assessed relative performance within a ZIP code area in recognition of the fact that most dining choices are local, which also had the collateral benefit of adjusting for inspector differences, as inspectors are assigned principally based on ZIP code. This choice proved contentious to elected officials, but there are few alternatives that would provide meaningful distinctions between restaurants and account for inter-inspector and regional variability.

Finally, how the county will use the experimental evaluation remains to be seen. Due to political constraints, such as the popular enthusiasm for and the King County executive support of restaurant grading, it is unlikely that the county will abandon grading. The management team is considering additional quality-assurance and performance-management efforts to ensure accurate signposting and to mitigate the impact grading systems have on inspectors, causing them to be more reluctant in writing violations. The trial results may be the most beneficial for jurisdictions facing the choice of whether to adopt a grading system, as it provides the first rigorous evidence of the effects of restaurant grading.

LESSONS

- *Clear benefits to government-academic collaborations exist.* Academic-agency partnerships are a critical way to develop the evidence base when agency research resources may be lacking, and researchers can craft evaluations to address broader scholarly questions. Both the peer-review trial and the evaluation of the grading system were first of a kind in this particular context.

- *Rigorous evaluations can inform policy and operations.* Policy implications are strongest from the experimental evaluation of peer review, with implications across many domains of regulatory enforcement and adjudication. The evaluation of the rating system at this point, however, does not sustain popular enthusiasm for restaurant grading as a way to lower foodborne illness, as there was no evidence of any health benefits.[19] If anything, the direct effect on citation behavior, which leads to well-documented forms of grade inflation, is something that public health agencies will need to take seriously in training and managing frontline staff when implementing rating systems. The citation effects, however, also corroborate the design of the rating system to mute "citation behavior" (using relative rankings within each ZIP code area).

- *Practical implementation barriers and public expectations must be acknowledged.* Although the rating system was grounded in evidence, the complexity of the design also posed practical challenges. RegLab developed open-source software to implement the system, but it was nonetheless challenging to integrate into information-technology systems. In addition, the complexity of the rating system, most notably the ZIP code adjustment, may be difficult to explain to the lay public, causing challenges for public messaging and outreach. The underlying difficulty stems from a public interest for a rating system based on inspections never intended for that purpose, a challenge not unique to food safety inspections. Ultimately, there may be considerable misunderstandings of what can and cannot be expected of food safety inspections.[20]

- *Decision-makers want positive results.* Substantial political pressures and the desire for immediate results created tension around the grading evaluation. Once a grading system is implemented, the champions of such a system desire studies to generate positive effects. A chief virtue of the grading experiment is a pre-commitment to an analysis plan (preventing "specification searching" to reach desired results), but how that evidence is used in prospective operations remains an open question.[21]

- *Continued engagement for building evidence is key.* The case study illustrates the need for active and ongoing stakeholder relations through the system design, implementation, and evaluation to support the integrity of the project throughout. Notwithstanding the challenges mentioned above, the case study illustrates the tremendous scientific and public-policy payoffs to academic-agency collaborations like the Stanford RegLab-Public Health model. ■

1 U.S. Centers for Disease Control and Prevention. "Estimates of Foodborne Illness in the United States: Findings." Last reviewed November 5, 2018.
Available at: https://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html.

2 U.S. Government Accountability Office. *High Risk Series: An Update*. GAO 15-290, 2015, 78.

3 Daniel Ho. "Does Peer Review Work. An Experiment of Experimentalism." *Stanford Law Review*, 69(1): 1-119, 2017.

4 Daniel Ho and Sam Sherman. "Managing Street-Level Arbitrariness." *Annual Review of Law and Social Science*, 13: 251-272, 2017.

5 Richard Boehnke and Catherine Graham. *International Survey on Public Posting of Restaurant Inspection Reports, and/or Grade Card Posting Schemes Based upon Health Inspections*. RHB Consulting & Associates, August 2000.

6 Chris Wiant. "Scores, Grades, and Communicating about Food Safety." *Journal of Environmental Health*, 61(9): 37-38, 1999.

7 Daniel Ho. "Fudging the Nudge: Information Disclosure and Restaurant Grading." *The Yale Law Journal*, 122(3): 522-851, 2012.

8 Ginger Zhe Jin and Phillip Leslie. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards" *The Quarterly Journal of Economics*, 118(2): 409-51, 2003.

9 T. Serapiglia, et al. "Association of Food Premises Inspection and Disclosure Program with Retail-Acquired Foodborne Illness and Operator Noncompliance in Toronto." *Journal of Environmental Health*, 70(1): 54-59, 2007.

10 Melissa Wong, et al. "Impact of a Letter-Grade Program on Restaurant Sanitary Conditions and Diner Behavior in New York City." *American Journal of Public Health*, 105(3): 81-87, 2015.

11 Daniel Ho, Zoe Ashwood, and Cassandra Handan-Nader. "New Evidence on Information Disclosure through Restaurant Hygiene Grading." *American Economic Journal: Economic Policy*, Forthcoming 2019.
Available at: https://www.aeaweb.org/articles?id=10.1257/pol.20180230.

12 M.A. Cruz,, D. J. Katz, and J. A. Suarez. "An Assessment of the Ability of Routine Restaurant Inspections to Predict Food-Borne Outbreaks in Miami-Dade County, Florida." *American Journal of Public Health*, 91(5): 821-823, 2001.

13 K. Irwin, et al. "Results of Routine Restaurant Inspections Can Predict Outbreaks of Foodborne Illness: the Seattle-King County Experience." *American Journal of Public Health*, 79(5): 586-590, 1989.

14 T.F. Jones, et al. "Restaurant Inspection Scores and Foodborne Disease." *Emerging Infectious Diseases*, 10(4): 688-692, 2004.

15 Kathleen Noonan, Charles Sabel, and William Simon. "Legal Accountability in the Service-Based Welfare State." *Law & Social Inquiry*, 34(3): 523-568, 2009.

16 Ho. "Does Peer Review Work."

17 Daniel Ho, Zoe Ashwood, and Becky Elias. "Improving the Reliability of Food Safety Disclosure: A Quantile Adjusted Restaurant Grading System for Seattle-King County." Forthcoming, 2019.

18 Cassandra Handan-Nader, Daniel Ho, and Becky Elias. *Feasible Policy Evaluation by Design: A Randomized Synthetic Stepped-Wedge Trial in King County*. Stanford Institute for Economic Policy Research, Working Paper 18-040, 2018.

19 Archon Fung, Mary Graham, and David Weil Fung. *Full Disclosure* (New York: Cambridge University Press, 2007).

20 Lauren Dundes. "Scores and Grades: A Sampling of How College Students and Food Safety Professionals Interpret Restaurant Inspection Results." *Journal of Environmental Health*, 64(5): 14-22, 2001.

21 Edward E. Leamer. "Let's Take the Con Out of Econometrics." *The American Economic Review*, 73(1): 31-43, 1983.

14. KEEPING KIDS SAFE AND HEALTHY

Changing Colorado Policies to Prevent Child Neglect

Jessica Corvinus is the manager for Research and Evidence-based Policy for the Colorado Governor's Office of State Planning and Budgeting.

Aaron Miller is the program manager for the Colorado Community Response within the Colorado Department of Human Services.
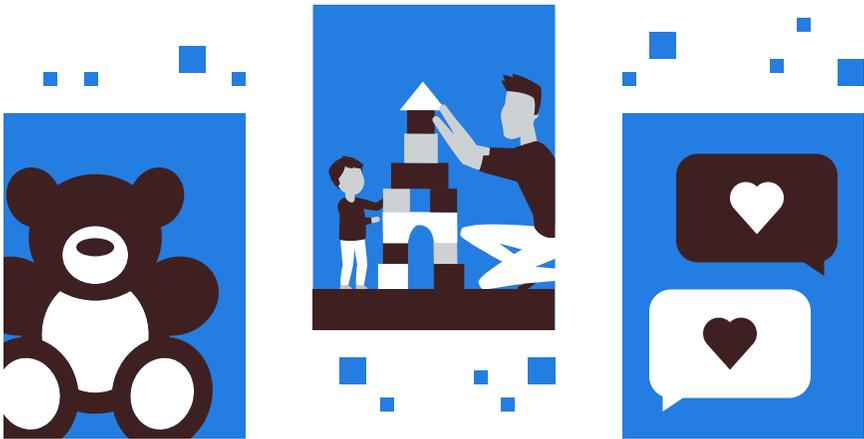
While there have been numerous initiatives in Colorado to demonstrate the state's commitment to evidence-based policymaking, one success was in ensuring that Colorado's children are safe and healthy. The Keeping Kids Safe and Healthy initiative created a strategic plan to improve and strengthen the state's child welfare system. The initiative launched in 2012 and initially focused on identifying best practices that could fit within Colorado's communities while improving outcomes for children.

Reforming a state child welfare system can be complex, time-consuming, and expensive. Colorado cautiously adopted reforms, including evidence-based prevention services to reduce the risk of child abuse and neglect. Specifically, the Colorado governor and the state Department of Health and Human Services' Office of Early Childhood wanted to increase prevention services for families who were once reported to Child Protective Services for child abuse or neglect but were eventually screened out. These children may not currently be in danger anymore, and their families may be stable enough for now, however, to ensure true stability and reduce chances of the children ending up back in the child welfare system, these families need support. The prevention-services program that addresses this need is Colorado Community Response, a voluntary family-focused program that provides a range of targeted services.

ISSUE BACKGROUND

When launching the Colorado Community Response program in 2014, decision-makers faced weighing how to address two related issues: child maltreatment and the screening procedures for families. Child neglect accounts for roughly three-quarters of confirmed cases of child maltreatment nationally and is the least clearly defined, understood, or

publicly recognized form of maltreatment.[1] Neglect occurs when a parent or caregiver fails to provide for a child's basic needs, potentially placing the child at risk for serious harm. Child maltreatment and a family's socioeconomic status are closely related.[2] Poverty increases the likelihood of child neglect, especially when combined with other risk factors, such as depression, substance abuse, mental health issues, and social isolation.

In Colorado, approximately 66 percent of reports for suspected child maltreatment are screened out or closed after assessment without the provision of any services from Child Protective Services. This happens when information received does not indicate that abuse or neglect is occurring. Families with closed assessments have a visit from a caseworker but do not have a safety concern requiring child welfare involvement. These determinations are not always perfect because caseworkers cannot predict the future.

Families receiving child welfare support can access family support and prevention services through their local county child welfare agencies. Agencies provide these services to prevent out-of-home placement or to return the children in placement to their own homes. But services are not available to families screened-out during the initial review by Child Protective Services. While an immediate safety risk may not exist for a child, many screened-out families still experience high levels of stress that can lead to future incidences of child maltreatment and re-referral to the child welfare system.

Recognizing the challenges Colorado faced in addressing and preventing child maltreatment, in 2012 Governor John Hickenlooper launched the Keeping Kids Safe and Healthy initiative. The Colorado legislature funded new prevention services, which allowed the Department of Human Services to prioritize the families screened-out from child welfare with an emphasis on families impacted by neglect and concerns around economic security. The new funding also required Colorado to evaluate the services to understand whether they achieved the intended goals. One of the major gaps identified during the review of the state's existing prevention services was the absence of an evidence-based program model that specifically focused on addressing and reducing neglect. Colorado looked to other states for potential models and identified the Community Response Program in Wisconsin as a promising approach for preventing child neglect.

Program development

Colorado designed its program to expand on the Wisconsin framework to prevent child abuse and neglect.[3] Colorado's program provides case management, direct services, resource referrals, home visits, and financial decision-making assistance. These services address characteristics that build on family strengths, buffer risk, and promote better outcomes.

In designing its program, Colorado opted to use an assessment tool for family advocates to assess key factors. One is family functioning. For this assessment, the state considers a range of characteristics, like finances, living arrangements, and social, economic, or health issues in a family. The second area is protective factors, which consider parental resilience, socioemotional issues, and child-development knowledge. Together these factors form the

basis for an assessment of areas of family strength or need. Caregivers along with support from their family advocate can then consider this information in making determinations about which services are most appropriate or necessary for a family to receive.

The assessments are a critical feature of the program design and for ensuring the Colorado Community Response program achieves its goals for addressing issues of child neglect related to economic insecurity and increase overall family functioning. Goals include providing comprehensive voluntary services, reducing referrals to child protective services, increasing protective capacity in families, addressing the link between poverty and neglect, promoting safe relationships, and increasing family economic security. The intent is that if the program achieves these goals, children will be safer and healthier.

EVIDENCE AVAILABILITY

While there is a vast amount of research about child welfare generally, there is a limited body of evidence on programs that target services to families screened-out from child welfare. In looking for relevant programs, Colorado sought programs that had a strong evidence base or were a promising practice for meeting the key goals of preventing child maltreatment and escalation of risk requiring child welfare system involvement.

Prior to even launching the Colorado Community Response program, staff from Colorado's Office of Early Childhood reviewed existing literature and even visited other programs.[4] Existing research further suggests that the Community Response Program is capable of engaging and successfully working with families around poverty-related neglect issues, income, employment, and other economic needs that could remove the likelihood of child abuse and neglect in the home. Especially when service goals included activities related to work, employment, and increased income, families were less likely to be re-reported to child-abuse hotlines after receiving community-based child abuse prevention services.[5]

Site visits, which collected observational information from the Wisconsin Community Response and the Milwaukee GAIN programs, allowed staff to get a better understanding of the programs in those states and to decide whether modifying policies in Colorado based on those experiences would lead to a relevant policy change. The observation of Wisconsin's program suggested the approach could yield promising results for families not previously involved with child protective services. It also filled a critical gap in the child-maltreatment prevention continuum.

Performance monitoring

Colorado established a series of performance measures and tracked trends and characteristics of program operations. The measures provide timely data and increase transparency for the program, while facilitating continuous improvement. The program established two measures identified by leadership in the department as important to monitor:

- **Economic self-sufficiency**: One measure includes the number of families who have an established goal to improve their economic self-sufficiency within 30 days of a program assessment. This goal provides insights to managers because, as a requirement for program participation, families must establish at least one goal.

- **Change in assessment scores**: Another reported measure monitors the difference between entry and exit scores in economic self-sufficiency. Colorado Department of Human Services expects an upward trend in this indicator, so a negative or reduced level may be cause for concern and potential program changes.

Together these program performance measures provide near real-time information for Colorado's Department of Human Services to consider how the program is operating.

Impact evaluation

Colorado Community Response was the first to conduct an impact evaluation to determine whether the program achieved intended outcomes. Using a nonexperimental design, the evaluation compared families who received services with those who did not.[6] Data collection for the Colorado Community Response program evaluation began in 2014 and concluded in 2017, with follow-up one year later. Researchers collected information at 21 sites that provide services to about 600 families in nearly half of Colorado's counties.

The evaluation examined "leading" indicators, such as protective factors, increased family functioning, and the provision of concrete services hypothesized to achieve a long-term goal. Researchers assess these indicators by asking caregivers to complete a standardized assessment as a pre- and a post-test for services. The evaluation found that the percentage of families below key thresholds decreased in all domains identified by caregivers as "readiness for change" areas between pre- and post-test.[7] As part of the assessment process, families also completed a protective factors survey. Protective factors are the strengths each family possess that help buffer and support them. The evaluation identified significant positive changes in each of these domains.[8]

The evaluation also considered "lagging" measures of output in child welfare, such as whether families had an assessment of child maltreatment 12 months after program completion. Lag indicators looked at child welfare re-involvement for families who completed program services versus those who were eligible but not referred. Initial evaluation findings in 2018 suggested families who completed the program had improved assessments, fewer substantiated reports of child maltreatment, and fewer out-of-home placements after one year compared with those who did not participate in the program.[9]

EVIDENCE USE

Colorado implemented the community response program after considering information available on Wisconsin's pilot initiative in 2012. Colorado launched the program as a

promising approach for increasing economic security for families, while reducing the prevalence of child neglect and further involvement in the child welfare system. The program used performance measures routinely. But it was not until the state conducted a more robust impact evaluation of the program that policymakers had access to the information they needed to determine whether to continue funding the project.

Using performance information

Because the program started collecting performance information early on, decision-makers are able to routinely monitor trends in key metrics. In addition to routine monitoring from program administrators, senior leadership in Colorado's Department of Human Services also monitors performance every month through C-Stat, a performance-based analysis strategy meeting, which allows the department's programs to meet and discuss performance trends. When indicators suggest issues that need attention, the C-Stat allows senior leaders to quickly be attuned, to develop a plan for improvement, and ultimately to contribute to improved performance. By identifying areas of focus, department leadership determines what is working and what areas may need improvement to ensure the best results for families. These regular executive meetings provide a strategy to assess the effectiveness of the program implementation and to support continuous quality improvement of service delivery.

The state made incremental changes as a result of the performance-monitoring data after discussing the issues. For example, performance measures helped identify variability in the delivery of services across multiple program sites. The state could quickly correct these variations through contracting requirements to service providers and by establishing new performance targets to help administrators prevent avoidable variations in implementation.

Use of the evaluation

Completion of the impact evaluation in 2018 allowed the Colorado Department of Human Services to better understand the constructs of what the program should measure routinely in performance measures related to child welfare outcomes and economic well-being for families.[10,11] The impact evaluation improved the state's understanding of whether families who completed program services had fewer negative outcomes than families who did not receive services.

The program is currently funded, and policymakers are interested in filling yet more gaps in existing knowledge moving forward. Recognizing the limits of the impact evaluation, Colorado planned a second evaluation to consider outcomes associated with families who return to the program within one year of completing services. Early indications suggest the program did prevent escalation of risk resulting in future child welfare involvement.

Additionally, the first impact evaluation did not examine economic well-being. This limited the information available for decision-makers about a key goal and attribute of the program, even though some descriptive analysis was available. In line with the program's

performance measures, a new evaluation will consider whether families with an economic self-sufficiency goal causally achieve better child welfare outcomes.

In this case, by completing an initial evaluation the department and state leaders recognized more information could be helpful in informing future decisions about the program and for achieving child welfare goals.

LESSONS

- *Innovation should always be paired with an evaluation.* Colorado looked to identify an evidence-based strategy that could meet their need of addressing child maltreatment and neglect outcomes through programming that targeted economic security, but the state recognized that evidence needed to come from current best practices. In order to move the Colorado-specific Community Response program along the evidence continuum toward evidence-based status, Colorado built on excellent work in other states, such as Wisconsin, and budgeted for a rigorous evaluation up front. One key lesson is that the state should pair innovation with an evaluation to understand the efficacy of programming. As part of the evaluation process, program leaders should refine activities and methods throughout and should focus on both short-term and long-term outcomes with meaningful performance metrics.

- *The first pilot may not answer all the questions.* Lessons from the completed impact evaluation highlight that program activities needed to be modified by eliminating variability in the model, program services needed a larger uptake to fully understand program impact, and the state agency has additional questions it wanted answered, such as on the impacts of economic self-sufficiency. The second evaluation will help answer additional questions and further demonstrate how and why building evidence takes time and careful consideration. A program cannot go from being promising or considered a best practice to becoming evidence-based without multiple types of evidence and evaluation.

- *Effective implementation is essential.* Effective implementation is necessary both in terms of budgeting and long-term sustainability of quality programming. Implementation resources such as training, coaching, and technical assistance should be ongoing to ensure the best delivery of programming and services. Colorado's program staff remain focused on learning directly from the Colorado Community Response study.

- *Confidentiality should always be a priority.* Maintaining confidentiality pledges is essential for county welfare agencies when sharing contact information about a family with community-based prevention programs. While some sharing is

essential, Colorado state law lacks clear guidance on the authority and process for doing so, leading to differing legal opinions about sharing information. Child welfare agencies could benefit from reviewing their state laws to plan for necessary data-sharing activities. The federal government could also provide improved guidance about expectations and how federal laws affect such activities.

- **Funding evaluations strengthens evidence-based programs.** The state legislature provided funding to conduct an evaluation while implementing Colorado Community Response. Funding for program evaluation was critical to strengthening the evidence about the program. However, programs would benefit from using implementation science as a process to identify barriers and reduce program variability across other agencies. State agencies will also take this approach to other innovations. For state policymakers to understand the impact of programming and make informed decisions, we need to ensure we are budgeting for, prioritizing, and advocating for the use of the evidence-based policy. ■

1 U.S. Department of Health and Human Services. *Child Maltreatment 2016*. Administration for Children and Families. Children's Bureau, 2016. Available at: https://www.acf.hhs.gov/sites/default/files/cb/cm2016.pdf.

2 Paul Bywaters, et al. *The Relationship Between Poverty, Child Abuse and Neglect: An Evidence Review*. Joseph Rowntree Foundation, 2016.

3 Jacqueline Counts, et al. "The Development and Validation of the Protective Factors Survey: A Self-Report Measure of Protective Factors Against Child Maltreatment." *Child Abuse & Neglect*, 34(10): 762-772, 2010.

4 National Quality Improvement Center. A literature review of *The Formal Public Child Welfare Responses to Screened-Out Reports of Alleged Maltreatment*: Wisconsin's Community Response Program for Families that Have Been Reported for Child Maltreatment, Project GAIN (Getting Access to Income Now), Parent Support Outreach Program, and the Colorado Family Resource Center Association Family Pathways.

5 Paul Bywaters, et al. *The Relationship Between Poverty*.

6 The comparison looked at families who received Colorado Community Response services against a comparison group who did not, matching on case characteristics and demographic factors, including referral type (screen out or closed assessment), number of children and adults living in the home, ages of children, prior referrals to child welfare, and type of allegation.

7 Paul Bywaters, et al. *The Relationship Between Poverty*.

8 Ibid.

9 Ibid.

10 Kristen Shook Slack, Lawrence M. Berger, and Kathryn Maguire Jack. *Community Response Program (CRP) Pilot Initiative: Final Implementation Report to the Wisconsin Children's Trust Fund*. University of Wisconsin-Madison, Institute for Research on Poverty, March 2012.
Available at: https://preventionboard.wi.gov/Documents/CRP_ImplemenationReport.pdf.

11 Colorado is for Kids. *Evaluation of Colorado Community Response Program Reveals Stronger Families*. October, 10, 2017. Available at: https://co4kids.org/community/evaluation-colorado-community-response-program-reveals-stronger-families.

15. IMPROVING LITERACY

Investing in Reading Coaching and Evaluation in Tennessee

Nate Schwartz is the chief research and strategy officer for the Tennessee Department of Education.

Laura Booker is the executive director of research for the Tennessee Department of Education.

Adam Rollins is a senior research analyst for the Tennessee Department of Education.

Between 2010 and 2015, Tennessee celebrated large gains across a wide variety of K-12 student outcomes. High school graduation rates increased steadily. More students enrolled in advanced coursework, and state test scores climbed. On the National Assessment of Education Progress (NAEP), Tennessee grew faster than any other state.[1]

Elementary reading was the exception. In 2011, the state deemed 46 percent of third-graders proficient on the state exam that measured reading comprehension, vocabulary, and language skills and knowledge.[2] Five years later, the percentage meeting the bar was the exact same. In contrast, over the same time period, for students in third through fifth grades, proficiency rates in mathematics soared from 39 percent to 59 percent and average ACT scores climbed by half a point.[3]

The lack of growth in early reading proficiency raised doubts about the sustainability of the state's academic gains. Research from Tennessee and across the nation has repeatedly demonstrated the tight connection between early reading ability and later outcomes. By the middle grades, these differences start to harden. In Tennessee, fewer than 10 percent of eighth-graders who are reading below grade level meet the college-readiness benchmark on the 11th grade ACT exam. Without improvements in reading proficiency, Tennessee's long-term academic progress seemed likely to grind to a halt.

ISSUE BACKGROUND

In 2007, the U.S. Chamber of Commerce awarded Tennessee a failing grade in the category of "truth-in-advertising" to highlight the misleading information coming from the high proficiency rates on Tennessee's statewide assessment (around 91 percent proficiency in English language arts) as compared with the more objective standard measured by the NAEP (28 percent proficiency).[4]

Supported by a favorable political climate and two governors who made education their primary focus, the state responded by doubling down on more rigorous standards and graduation requirements over the next several years. Between 2007 and 2012, Tennessee raised standards in core academic subjects and made substantial investments in teacher training and performance evaluation, funded by a $500 million Race to the Top grant.[5] Improvements in most student outcomes across the time period seemed to vindicate this approach, making the lack of growth in elementary reading proficiency all the more problematic.

Tennessee's static reading results in the early grades were directly at odds with the level of effort districts were devoting to the issue. Each year, the vast majority of Tennessee districts placed elementary reading improvement as one of their highest priorities in the required annual plans they submitted to the state, and they supported these priorities through investments in trainings and resources.

When a new commissioner of education took office in January 2015, she announced that early grades literacy would be a major area of action for the Tennessee Department of Education. Her commitment to the issue was driven by evidence that being a proficient reader by third grade is foundational to future success and by the data showing Tennessee's lagging scores in this area.

EVIDENCE AVAILABILITY

Making research and data use a priority at the outset, the department launched its early reading initiative with a study of the landscape of teaching and learning in Tennessee K-3 classrooms. This needs assessment served to ground improvement efforts in a deeper understanding of the quality and variation in classroom instruction across the state.

The department outlined its dive into classroom practice in *Setting the Foundation*, the first in a series of reading reports that highlighted critical classroom needs.[6] Classroom observations suggested that kindergarten through third-grade reading teachers across the state were overly focused on skills-based competencies like alphabet knowledge, fluency, spelling, and print concepts, with not enough time spent on the deeper comprehension and vocabulary—teaching practices that decades of literacy research suggest matter most for long-term student gains.[7] Moreover, teachers rarely exposed students to high-quality texts that built their knowledge as well as their decoding and comprehension skills.

The *Setting the Foundation* report also suggested, as a statewide goal, that 75 percent of Tennessee third-graders be proficient in reading by 2025. Yet it wasn't clear how to shift teachers' practice in reaction to these needs in order to meet this goal. While professional-development programs tend to be popular among teachers and administrators, numerous experimental evaluations of teacher training programs have suggested that the programs often do not meet expectations for impacts on student learning or on teacher practice.[8]

Studies of teacher coaching are something of an outlier in this literature. As one meta-analysis including only randomized control trials and quasi-experimental evaluations concluded: "Coaching works. With coaching, the quality of teachers' instruction improves by as much as—or more than—the difference in effectiveness between a novice and teacher with five to 10 years of experience, a more positive estimated effect than traditional [professional development] and most other school-based interventions."[9]

Based on reviews of this research and the increasing prevalence of instructional coaches in Tennessee schools, the Tennessee Department of Education started to envision a strategy based around supporting the work of literacy instructional coaches across the state.[10] Many districts and schools in the state had invested in instructional coaches as a lever for improving instruction, but coaches often reported little training on how to actually be a coach.

Nevertheless, this initial direction only served as the starting point for a series of far more difficult, practical decisions about the program itself. The challenge here was practice, not policy. While the evidence broadly suggested that coaching could work, these impacts tended to emerge in studies of smaller, more-intensive programs, and large-scale programs tended to be far less reliable.

The department did not have the option of replicating and scaling up a small-scale model. Funding literacy coaches in a large contingent of schools across the state was financially out of reach, even with a committed legislature. However, legislators were looking for a program that could reach the majority of districts in the state. As is the case with many such programs, state leaders faced a tension between the political need to create a program with a large enough reach to justify its expense and the reality that there was not a blueprint for success on a large scale.

In order to meaningfully shift classroom practice toward higher-quality literacy instruction across around 15,000 K-3 classrooms and 146 school districts, the department needed to use existing structures with innovative, evidence-based approaches. The department would also need to use ongoing evidence collection to inform continuous improvement in implementation.

EVIDENCE USE

The coaching strategy proposed by the Tennessee Department of Education to the legislature relied on a small group of centralized leaders—15 reading coach consultants across eight regions—who would build relationships with a far larger group of district-employed instructional coaches, who would then work with individual teachers across the state to build a shared understanding of early grades literacy practice. Each of the regional coaches would work with about 13 district coaches who in turn focused on supporting 15 teachers in their district. The district coaches would convene twice per year for focused training on reading instruction and coaching best practices.

This immediate setup was hurried along by legislative and budget timelines that freed up money in the late spring for a program directed to launch at the beginning of the following school year. It led to a program with the broad goal of transforming teaching across hundreds of classrooms but with only a rough outline of how to get there.

At the same time, program leaders across the department shared a long-term and genuine commitment to continuous improvement and openness to research and evaluation that is relatively rare in the political world of state agencies. Even as program leaders first began to piece together the outline of the coaching initiative, they worked with members of the department's embedded research and strategy team to lay out a framework for ongoing monitoring of short-term and long-term outcomes. This included thinking through a logic model for the initiative tied to concrete ways of assessing the stated outcomes. A member of the research team also accompanied program leaders to information sessions around the state for district leaders to share the data driving the program and to communicate about the evaluation plan. This partnership meant that ongoing evaluation became a primary element of program design, with discussions about data and evidence built directly into the planning and implementation process.

Embedded evaluation

As part of an embedded evaluation, the department committed to classroom observations that would allow monitoring of actual classroom practices over the three years of the coaching network, rather than relying on after-the-fact analyses of test scores to measure program effects. Through a stratified sampling plan to ensure representation on a variety of crucial school and district characteristics, the department identified 18 schools that experts could visit over time to track changes in teaching practices.

The department also built a system of continuous feedback from teachers and coaches, drawing on processes that were already in place across the state, specifically annual surveys of educators and district leaders, but adapting questions to meet evaluation needs. The team also developed new pre- and post-surveys for each of the six semesters of initiative. Making use of the statewide annual educator survey allowed for comparisons between teachers receiving coaching and a control group of uncoached teachers, and the supplemental surveys for participating coaches and teachers meant deeper information related to the program content. Tracking intermediate measures like changes to teacher and coach knowledge and practice was critical to knowing whether the program was on track to achieve its ultimate goal of moving the needle on student reading proficiency. These intermediate measures were especially important because statewide testing of students in Tennessee begins in third grade, meaning there was limited achievement data for assessing the ultimate outcome.

At the close of the first year, classroom observations suggested little to no positive movement across a variety of metrics of teacher practice. As program leaders studied these results, it became increasingly clear that the initial program aimed to cover so many elements of strong instructional practice that it left coaches and coach consultants with little focus on concrete classroom outcomes.

Meanwhile, the feedback from surveys and observations zeroed in on reading texts and curriculum as a crucial input to the success of the coaching program. On the early literacy branch of the state's annual educator survey, K-3 English language-arts teachers reported that misalignment between instructional standards and district curricula meant that they were spending an average of four and a half hours per week just sourcing the right texts and materials to use in their classrooms, leaving little to none of their crucial prep time for building a deeper understanding of their craft. Also teachers used coaches mostly as purveyors of instructional materials rather than as teaching experts. Moreover, surveys of coach time showed that many coaches struggled to actually spend time with teachers or to focus on observing and providing feedback, often because they were pulled into other duties.

Programmatic changes

Through ongoing conversations that were originally meant to ensure that classroom observers collected the right evidence, the program team ended up redefining the program itself.

Using the observational data-collection tool as a guide, program leaders narrowed in on the changes they'd like to see teachers make across the state, and modified the program to focus on (a) the quality of classroom texts; (b) question sequences and tasks selected to build conceptual knowledge; and (c) systematic foundational skills instruction that incorporated opportunities to practice through reading and writing. Interactions between regional coach consultants and district instructional coaches in year two aimed to solidify teachers' understanding of these key outcomes. District coaches also began to focus their interactions with teachers more solidly on these outcomes. The research team supported this effort by developing one-pagers that unpacked each of these three priority areas.[11]

In response to findings from observations and surveys that the quality of instructional materials was a barrier to increasing the quality of instructional practice, in year two department leaders expanded the scope of the initiative to include a strategy focused on the materials and resources available to K-3 teachers. This included a short-term strategy of developing short units with tasks and question sequences built around a series of highly recommended text sets. It also included a long-term strategy to modify the textbook adoption process to ensure stronger alignment between the teaching techniques built into the coaching model and the curricula that the state recommended for adoption. Finally, in an attempt to bring school leaders directly into the work and thereby create more time and space for coaches to carry out their duties, the department held a series of regional principal meetings to invest principals in the vision.

Program impact

By the start of year three, observation results suggested that these program modifications had begun to yield meaningful changes in classroom practice. Observed teachers were increasingly introducing students to foundational skills within the context of genuine reading and comprehension exercises rather than creating an artificial divide between skills practice and the act of reading. Text quality in classrooms shifted, with more teachers using demanding texts that would introduce students to meaningful ideas and vocabulary. In

classrooms using the newly developed short units, almost all teachers were now using high-quality texts and one-third were using high-quality question sequences compared with only 10 percent before implementation of the units.

All the same, several crucial metrics of classroom practice, including the quality of student tasks had not moved. More problematic still, there were no changes yet in student test scores that evaluators could tie directly to the program. Student growth was difficult to measure since tests only began in grade three and the program focused primarily on grades K-2. Still, this remained the primary outcome that department leaders hoped to move.

The third and final year that the program had guaranteed legislative funding brought a new governor to the statehouse and a new commissioner to the department. While the program had achieved substantial popularity among district superintendents, it did not surface as a policy priority during the governor's election campaign. Instead, the governor, a fiscal conservative, had trained his sights on funding to improve career and technical education in high school and to increase school choice. As of this writing, it remains unclear whether the initiative will receive continued legislative funding. Yet despite the uncertainty, schools and districts across the state remain committed to the instructional improvements that the statewide coaching initiative spurred, and the work to increase the quality and rigor of instructional materials continues to be a priority of the new commissioner.

LESSONS

- *Partnerships can benefit program quality.* The partnership between the early literacy team and the research team at the department demonstrated the value of long-term, embedded evaluation. When the research team first formed in the Tennessee Department of Education, officials often called on team members to evaluate programs after the fact. The research team would do whatever was possible with existing data, which often led to evaluations of outcomes not tightly linked to programmatic goals. By involving the research team from the very beginning, we were able to create a joint partnership that simultaneously improved the quality of both the research and the program.

- *Tye ongoing evaluation to the initiative.* Program leaders within the department had reason to continually review their theory of action, since they were working to align evaluation metrics to that theory. By developing pre- and post-knowledge surveys for each of the in-person trainings offered to the district coaches, the program team had to could solidify their expectations. The research team also gave presentations of the ongoing findings to the regional coaches so that those implementing the program had the opportunity to grapple with the data.

- *Learn to be okay with not getting positive results.* By forcing difficult conversations, the continuous flow of sometimes negative evidence helped to give direction to year-by-year adjustments that resulted in a stronger program over time.

- *Evidence use doesn't guarantee funding.* Tennessee continues to grapple with the ways that the political realities of funding make it difficult to sustain a truly iterative process of continuous improvement over the long term.

- *Internal and external evaluations are both valuable and serve different purposes.* It is also important to note that there are benefits to having an internal research team like the one described in this case study, but there are also real reasons to have many aspects of the state's long-term evaluation work carried out by an independent and external organization. In Tennessee's case, the state greatly benefited from the Tennessee Education Research Alliance at Vanderbilt University, which works in partnership with the department to build knowledge around central priorities.[12] ∎

1   State Collaborative on Reforming Education. *Priorities for Progress: 2018-2019 State of Education in Tennessee.* Available at: http://tnscore.org/wp-content/uploads/2019/02/2018-19-Annual-Report_18P1959_digital-FINAL.pdf.

2   Tennessee Department of Education. *Setting the Foundation: A Report on Elementary Grades Reading in Tennessee.* March 2016. Available at: https://www.tn.gov/content/dam/tn/readready/documents/coaching_network_docs/Setting_the_Foundation_Reading_Report.pdf.

3   Ibid.

4   Sheryll Poe. *Does Your State Make the Grade?* July 2014. Available at: https://www.uschamber.com/does-your-state-make-grade.

5   U.S. Department of Education. *Race to the Top: Tennessee Report Year 1: School Year 2010-2011.* January 2012. Available at: https://www2.ed.gov/programs/racetothetop/performance/tennessee-year-1.pdf.

6   Tennessee Department of Education. *Setting the Foundation: A Report on Elementary Grades Reading in Tennessee.* March 2016. Available at: https://www.tn.gov/content/dam/tn/readready/documents/coaching_network_docs/Setting_the_Foundation_Reading_Report.pdf.

7   Harvard University Lead for Literacy Initiative. "Literacy Unpacked: What Do We Mean By Literacy." Memo No. 4, 2012, 1. Available at: https://langlit.gse.harvard.edu/files/lesaux/files/04literacy_unpacked_langlit.pdf.

8   *See, for example*: M. Garet, et al. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement.* Institute of Education Sciences, September 2008. Available at: https://ies.ed.gov/ncee/pdf/20084030.pdf.

9   Matthew A. Kraft, D. Blazar, and D. Hogan. "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence." *Review of Educational Research*, 88(4): 547-588, 2018.

10 Among Tennessee districts, 80 percent reported having instructional coaches, according to a 2015 survey.

11 Tennessee Department of Education. *First Steps: A Report on Elementary Grades Reading in Tennessee.* 2018.

12 Vanderbilt University. "Tennessee Education Research Alliance." Available at: https://peabody.vanderbilt.edu/TERA/index.php.

16. GREAT EXPECTATIONS, RE-CALIBRATED: EVALUATING DC'S POLICE BODY WORN CAMERA PROGRAM

David Yokum is director of The Policy Lab at Brown University and an adjunct associate professor at the Watson Institute for International and Public Affairs. He was previously director of The Lab @ DC in the Government of the District of Columbia, where he remains as senior advisor.

Anita Ravishankar is a research scientist at the Metropolitan Police Department in Washington, D.C. and a fellow with The Lab @ DC.
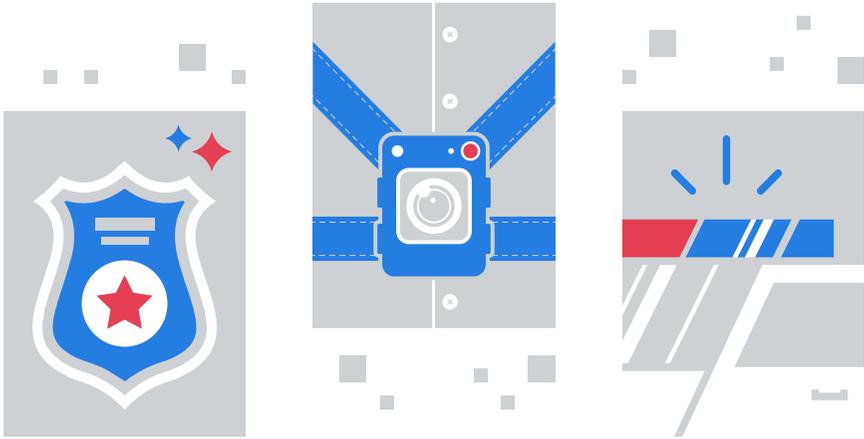
Body worn cameras (BWCs) are purportedly a tool to improve policing and enhance the public legitimacy of a police department. The technology is expensive and, when widely adopted, creates a notable expansion of state surveillance. To what extent do the hoped-for benefits occur, and how does a community balance those benefits against the monetary and privacy costs when deciding whether and how much to invest in a BWC program? What can facilitate informed thinking in this decision, especially amid the emotionally charged backdrop of a nation roiled by a series of high-profile use-of-force incidents, many involving minority residents?

We tell here the story of how Washington, D.C.'s police department handled its BWC program, and in particular, how randomly assigning some but not all officers to wear body cameras and comparing the outcomes helped build the department's capacity to use evidence while generating insights that can inform the national debate.[1]

ISSUE BACKGROUND

It's March 2015 in the Metropolitan Police Department command center, and the senior police leadership are gathered to make an important decision on the deployment of body-worn cameras in the nation's capital. Since October 2014, the department has been carefully planning the potential deployment of a BWC program to its 3,800-member police force. Pilot work assessed equipment options and settled on Taser International's AXON Body 1; policymakers drafted debated, and re-drafted regulations over access to video footage; and the city appropriated funds to the program, with a District of Columbia City Council mandate to outfit all officers with BWCs before the end of 2016. The decision point—after

weeks of advance discussion, draft pitches, informal brainstorms over beers, and now a formal proposal in front of those flickering screens—is whether and how to evaluate the program across the 68 square miles of the District of Columbia.

The national movement was remarkable. A series of high-profile, controversial deaths of unarmed African Americans at the hands of police officers, Eric Garner, Michael Brown, Laquan McDonald, Tamir Rice, Walter Scott, sparked a robust national conversation on police accountability and transparency. The role of video footage, either by bystanders and/or dashcam, in bringing many of these incidents to light led directly to a widespread embrace of police body-worn cameras as a technological solution. President Barack Obama proposed a $75 million, three-year investment to help purchase 50,000 BWCs; the U.S. Department of Justice awarded $23 million for BWC in 2015 for the initiative.[2,3] A nationwide survey at the time found that 95 percent of large police departments planned to have a BWC program. The central hope was that the watchful lens of a camera would moderate behavior on the streets (either immediately or via a disciplinary and training feedback loop), which in turn would reduce the likelihood that encounters between officers and residents would escalate into violence.

EVIDENCE AVAILABILITY

Yet the evidence base in early 2015—although promising—was surprisingly thin, especially relative to how rapidly BWC programs were beginning to spread across the country. The primary empirical justification was citation to a single study from Rialto, California.[4] Fifty-four police officers were randomly assigned to wear a camera on some shifts but not others, for a total of 988 shifts over 12 months. Officers were less likely to use force on shifts when wearing a BWC than on shifts when not wearing a camera. Officers were less likely to use force on shifts when wearing a BWC than on shifts when not wearing a camera. No difference in complaints was detected between shift-types. Only an after-the-fact time-series analysis identified a reduction in complaints, a result which was commonly misinterpreted in the media as a finding from the more rigorous evaluation. The authors cautioned against too quickly deciding that BWCs merit expansion, noting in the paper that the results were from "but one experiment."[5]

Most justification was therefore theoretical, based on evidence from other domains. A wide range of research, dating back to the classic experiments at Hawthorne Works, reported that people may act differently when watched.[6] This literature is not without controversy, but studies have regularly reported—and a common belief is—that we're more likely to work harder, give to charity, tell the truth, recycle, vote, and so forth if being observed than if alone. If this behavioral effect carried over into the on-street policing environment, then both officers and residents should be more likely to behave in line with community norms when a camera is present. With both parties more closely regulating their behavior, there should be a lower risk of events escalating into violence.

EVIDENCE USE

This case is about the use of evidence, but in many respects, the most important lessons are about the *generation* of evidence for public consumption. The same evidence may be more or less likely to be used, we believe, as a function of how it is generated and presented in relation to the political decision-making process. But first, it is important to note that the most immediate use of evidence and theory was, of course, in justifying the deployment of the BWC program in the first place. For example, in his testimony before the Committee on the Judiciary, Deputy Mayor for Public Safety and Justice Kevin Donahue relied on the report of an advisory group summarizing the Rialto study and the underlying theory; the Committee Report that informed the City Council's vote to create and fund the program also used this report.[7,8]

Perhaps the most inspiring part of this project was the political desire and will to learn what happened after the initial evidence-informed decision to begin a BWC program. Noting the limitations of prior evidence, Mayor Muriel Bowser, Chief Lanier (and later Chief Peter Newsham), Deputy Mayor Donahue, and City Administrator Rashad Young all recognized the responsibility to evaluate the program as rigorously as possible. A full deployment was inevitable by legislative mandate, so the intention was not to inform a decision to scale; rather, the learning would inform other law enforcement agencies considering BWC programs as well as provide a baseline understanding of what was happening in the District. Such a baseline would provide an evidentiary platform to inform further optimization of the program and, more generally, other activities of the police department.

A decision was made to conduct an experimental evaluation, with each of 2,224 Metropolitan Police Department (MPD) officers randomly assigned—imagine flipping a coin—to either wear or not wear a camera. (The full details of the study, which are a touch more nuanced, are described in a paper under publication in the *Proceedings of the National Academy of Sciences*.) Because of the experimental design with random assignment, outcomes (for example, the likelihood of uses of force or civilian complaints) should be the same on average for the group of officers with BWCs as for the group of officers without cameras, unless the one thing controlled to be different—namely, the assignment of BWCs—*causes* a difference.

The study began in June 2015 with a pilot in two of the seven MPD police districts. This period was to ensure proper implementation of the program and to monitor fidelity to the study protocol. It also, importantly, generated early evidence of potential effect sizes, which researchers used to inform research design options. With immense public pressure to fully deploy the program immediately, the selection of the research design—including how long to run the study and at what level of randomization—was not an easy or simple one. Yet our pilot results indicated that the precision of the effect estimates would increase with the duration of the study. Although there are rule-of-thumb thresholds in the sciences for determining whether an experiment is precise enough to yield meaningful results, the decision of precise to be is ultimately a value judgment. We were able to inform that value judgment—that political decision—by providing power analyses of the minimal detectable effect, assuming the study

lasted three, six, nine, or 12 months. The infusion of this technical information bolstered the political buy-in to wait until the latest legislatively permissible endpoint, December 2016, before distributing BWCs to the group that had not received them initially. Mayor Bowser was even questioned on MSNBC's Morning Joe as to why the department had implemented this delay; she smoothly whipped back that "there was no delay at all," but rather we needed a "control group" to learn how the program was working.[9]

On October 20, 2017, we released a working paper with interactive results on a website dedicated to the study.[10,11] The study found that BWCs had no statistically significant average effects on any of the measured outcomes. To get a feel for the results, imagine groups of officers over the course of a year: a group of 1,000 officers with BWCs was estimated to document 74 more uses of force in a year than a group of 1,000 officers without BWCs, yet the data were also consistent with the real effect of BWCs being anywhere from a decrease of 97 documented uses of force to an increase of 244 documented uses of force per 1,000 officers, per year. A null result obtained for complaints too, with anywhere from a decrease of 24 complaints to an increase of 138 complaints per officers, per year, consistent with the data. We also did not find any downstream courtroom effects in how complaints or cases were adjudicated, although the statistical power was much weaker on this front.

News of the study quickly went global. The study's release coordinated with a front-page article in *The New York Times* and an NPR segment for *All Things Considered*.[12,13] NBC Nightly News interviewed Chief Newsham. Dozens more articles and interviews proliferated.[14,15] Researchers presented at the International Association of Chiefs of Police and fielded numerous phone calls with researchers and leadership teams from police agencies across the country and internationally to educate people about the findings. We also held two community conversations in D.C., to engage with local residents about the study findings and their implications for the city.

We spent an inordinate amount of effort ensuring that no one misunderstood or exaggerated the results or their implications. Indeed, the working paper, website, and associated coverage are perhaps unique in the quantity of pages and airtime given to explaining what the null results do *not* mean, and to highlighting alternative explanations for why we might have found what we did. We noted the risk of spillover, questions about generalizability beyond the D.C. context, limitations of administrative data, the possibility of small or distributional effects outside of what the study methods can capture, and the fact that no survey data was collected about public perceptions of police legitimacy. We also vetted our work, before any public releases, with a series of quality control measures and reviews. All analyses were conducted by two independent statistical teams, for example, which helped avoid coding errors and confirmed a convergence of results. We shared drafts and presentations with peer experts for advance review. Our code and replication data publicly will also be made publicly available.

Our overall recommendation was that we should recalibrate our expectations about the impacts of BWC programs, especially with regard to the possibility of large reductions in the average use of force or complaint rates. We deliberately did not advocate for or against the

adoption of a BWC program. The language was instead carefully crafted to prompt Bayesian thinking, that is challenging pre-existing beliefs based on the evidence. We considered how this study should be incorporated into the broader research and policy debate: the study does not definitively prove that BWCs have no impacts—no single study could accomplish such a feat—but it does recommend updating expectations, recognizing that the likelihood of positive effects on the measured outcomes is lower than previously believed.

It remains a political judgment as to whether a mayor or chief of police wants to implement a BWC program, and there may be justifications beyond seeking large average reductions in use-of-force or complaint rates. Perhaps the main objective is to enhance the public perception of legitimacy (which the study did not measure), or perhaps a jurisdiction would be willing to absorb the considerable financial and privacy costs for the possibility of capturing even a single wrongful act on video (an effect size too small to be practically measured). In the District, for instance, there is ambition to use the video footage for innovative officer training, to which end the city passed a unique legislative provision empowering access to the footage for research purposes.

If other jurisdictions use evidence from the District study to prompt more careful deliberation about these justifications and tradeoffs, then it will be a successful use of evidence, no matter which side of the decision a department ultimately lands on. For example, some people occasionally question the generalizability of the results by noting that the MPD has a unique history of reform and officer training protocols. BWC may be more impactful in police agencies without that history. Maybe indeed. But if people use the study's evidence successfully, it will prompt consideration of whether to prioritize investments in the technology or in implementing reform and training efforts similar to those in the District of Columbia.

LESSONS

- *Use a pre-analysis plan.* A remarkable feature of this project was the relative lack of after-the-fact attempts to discredit the study's findings on methodological grounds, to attack the independence of the research team (The Lab @ DC is housed out of the Executive Office of the Mayor), or to politically attack the mayor and chief of police for the program results. Discussion was instead, refreshingly, almost entirely about the implications of the evidence and what to do next. We attribute this success to a unique campaign of community engagement, anchored around public registration of a pre-analysis plan. A pre-analysis plan documents the questions a study will ask and the methodologies it will use in answering those questions—all written down before looking at or even collecting the data. Different methodological choices can generate different answers, and research has uncovered that scientists, consciously or more often unconsciously, have a bias toward choices that generate a desired result (so-

called "*p*-hacking"). A related bias is the tendency to tell stories after results are known, as if the result were theoretically expected from the onset (so-called "HARKing," or hypothesizing-after-results-known).[16] Both biases are short-circuited if methodological choices are not made in advance of conducting data analysis.

- ***Lean into the reality that science involves politics.*** The pre-analysis plan is also a unique vehicle for enhancing the *political* integrity of a project.[17] Science involves value judgments: how large of an effect size, measured at what precision, justifies a decision to take one action over another? Objective methodologies can provide reliable statistics, but weighing the risks and tradeoffs of any given intervention is always a political act. Lean into this reality. In particular, we used the drafting of the pre-analysis plan to work closely with key decision-makers to facilitate advance thinking about what the most important questions are, how much effort should be expended on answers, and what methods are most suitable. We publicly registered a draft pre-analysis plan on the Open Science Framework in October 2016 to facilitate community feedback.[18] Of course, not many people casually browse the Open Science Framework archives, so a critical step was hitting the streets: between October 2016 and June 2017, we held 13 public events to discuss the pre-analysis plan, ranging from professional audiences, to non-profit and advocacy groups, to everyday audiences gathered at libraries and schools. We circulated the pre-analysis plan widely and invited comments. We worked closely with press offices to ensure that officials understood and could confidently discuss the details when the opportunity surfaced—from council hearings to *Morning Joe.* Our proactive and open approach bolstered trust in the research. When the results came out, there was not the usual arguing over methods because the methodology, whatever its lingering shortcomings, had already been extensively discussed and accepted. When people scrutinized our independence, we could point to the publicly registered pre-analysis plan and explain again how the transparency curbed against *p*-hacking and HARKing.

- ***Co-production of research increased the likelihood of use.*** From the onset, we embedded closely with MPD. We joined internal meetings, conducted ride-alongs, learned to use the equipment, and visited the training academy. The department even hired one researcher as a staff member. And likewise, the department was authentically part of the research enterprise: Chief of Staff Matthew Bromeland, Commander Ralph Ennis, Special Assistant Heidi Fieselmann, and BWC Program Coordinator Derek Meeks worked closely with us to operationalize the field experiment and joined community discussions. This integrated approach empowered us to design a rigorous methodology meshed into the operational realities of what it takes to deploy thousands of pieces of physical equipment across an entire police force. Perhaps the most heartening outcome is how the project elevated the overall capacity of the police department to use and generate evidence into the future. We mapped and documented the administrative data of the department, a resource which is now being leveraged in new projects. MPD also created three permanent positions—a research scientist, a data scientist, and a management analyst—to boost its internal capacity to do research and

evaluation work, and has since initiated a variety of projects to improve recruiting, training, and violence reduction efforts. Terms like "counterfactual," "confidence interval," and "effect size" are commonly used in the hallways and meeting agendas, with the question of "how are we going to know if this works?" a common follow-up to any new proposal. ■

1   David Yokum, Anita Ravishankar, and Alex Coppock. *Proceedings of the National Academy of Sciences*, forthcoming 2019.

2   Andrea Peterson, "President Obama wants to spend $75 million to buy police bodycam." *The Washington Post*, December 1, 2014.

3   U.S. Department of Justice. "Justice Department Awards over $23 Million in Funding for Body Worn Camera Pilot Program to Support Law Enforcement Agencies in 32 States." September 21, 2015. Available at: https:// www.justice. gov/opa/pr/justice-department-awards-over-23-million-funding-body-worn-camera-pilotprogram-support-law.

4   Barak Ariel et al. "The Effect of Police Body-Worn Cameras on Use of Force and Citizens' Complaints Against the Police: A Randomized Controlled Trial." *Journal of Quantitative Criminology*, 31(3): 509-535, 2015.

5   Ibid.

6   For a recent review of research on the Hawthorne effect, *see*: Jim McCambridge, John Witton, and Diana R. Elbourne. "Systematic Review of the Hawthorne Effect: New Concepts Are Needed to Study Research Participation Effects." *Journal of clinical epidemiology*, 67(3): 267-277, 2014.

7   B21-351, the "*Body-Worn Camera Program Regulations Amendment Act of 2015*", B21-356, the "*Public Access to Body-Worn Camera Video Amendment Act of 2015*", and PR 21-327, the "*Metropolitan Police Department Body-Worn Cameras Approval Resolution of 2015*": *Hearings Before the Committee on the Judiciary*, 2015 Leg, (Washington, DC. 2015), statement of Kevin Donahue, Deputy City Administrator, Deputy Mayor for Public Safety and Justice. Available at: https://mpdc.dc.gov/ sites/default/files/dc/sites/mpdc/publication/attachments/FINAL%20Donahue%20Testimony_10%2021%2015.pdf.

8   Council of the District of Columbia, Committee on the Judiciary, *Report on Bill 21-0351, the "Body-Worn Camera Program Amendment Act of 2015"* Council Period 21, 2015.
Available at: http://lims.dccouncil.us/Download/34469/B21-0351-CommitteeReport1.pdf.

9   Muriel Bowser, interview by Joe Scarborough and Mike Brzezinski, *Morning Joe*, MSNBC, July 14, 2016.

10 David Yokum, Anita Ravishankar, and Alexander Coppock. "Evaluating the Effects of Police Body-Worn Cameras." Working Paper, October 2017.
Available at: https://bwc.thelab.dc.gov/TheLabDC_MPD_BWC_Working_Paper_10.20.17.pdf.

11 The Lab @ DC, "Randomized Controlled Trial of the Metropolitan Police Department Body-Worn Camera Program," 2017. Available at: https://bwc.thelab.dc.gov/#home.

12 Amanda Ripley, "A Big Test of Police Body Cameras Defies Expectations," *The New York Times*, October 20, 2017.

13 Nell Greenfield, "Body Cam Study Shows No Effect on Police Use of Force or Citizen Complaints," *All Things Considered*, National Public Radio, October 20, 2017.

14 Peter Herman. "Police Officers with Body Cameras Are as Likely to Use Force as Those Who Don't Have Them." *The Washington Post*, October 20,2017.

15 Amanda Ripley. "New Study Finds Little Change in D.C. Officers' Behavior with Body Cameras." *The Seattle Time*, October 20, 2017.

16 For an accessible and interactive introduction to issues such as p-hacking and HARKing, *see*: Christie Aschwanden. "Science Isn't Broken: It's Just a Hell of a Lot Harder Than We Give It Credit For." *FiveThirtyEight*, August 19, 2015. Available at: https://fivethirtyeight.com/features/science-isnt-broken/#part1.

17 David Yokum. "Psychology, Open Science, and Government: The Opportunity." *Association for Psychological Science Observer*, 29(4), 2016. Available at: https://www.psychologicalscience.org/observer/psychology-open-science-and-government-the-opportunity.

18 David Yokum. "Pre-Analysis Plan: A Randomized Controlled Trial of the Police Body-Worn Camera Program in the District of Columbia." October 7, 2016. Available at: https://osf.io/yjyng/.

DESIGNING POLICYMAKING PROCESSES FOR EVIDENCE USE
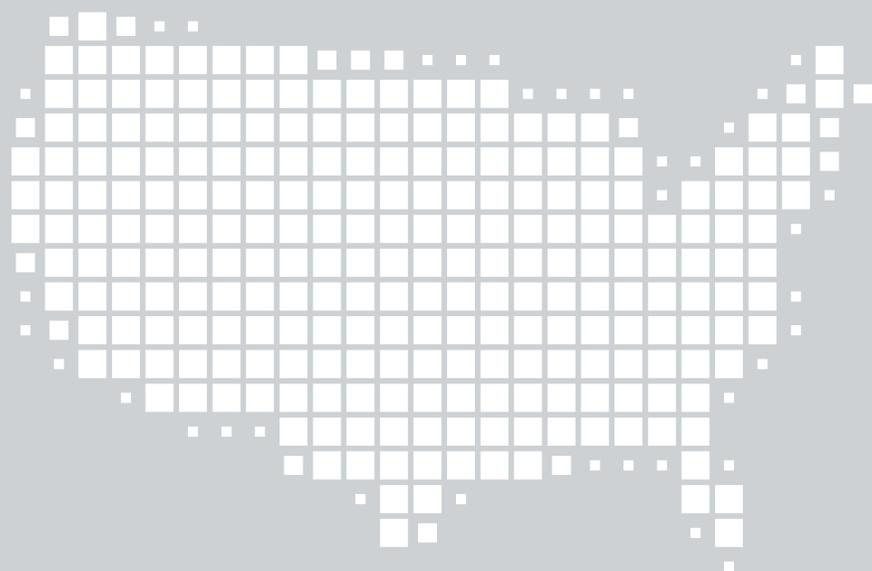
17. TEEN PREGNANCY PREVENTION PROGRAMS

A Tale of Persistence

Andrea Kane is the vice president for policy and strategic partnerships at Power to Decide.

Rachel Fey is the policy director at Power to Decide.

G. William Hoagland is a senior vice president for the Bipartisan Policy Center. He is currently a council member at Power to Decide and previously worked as the Republican director of the Senate Budget Committee.
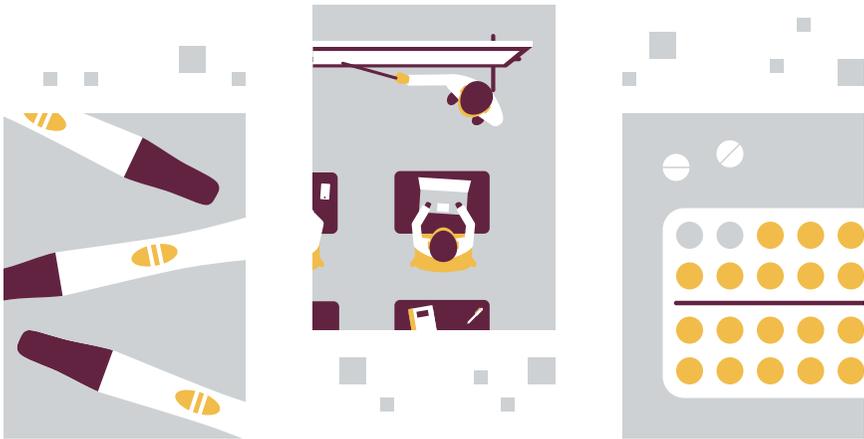
Teen pregnancy can have a wide range of health, social, educational, and economic consequences for young people. Policymakers long ago decided that reducing these preventable consequences for adolescents was a priority. But despite considerable progress, teen pregnancy and birth rates in the United States remain higher than most other developed countries, and racial/ethnic, geographic, and socioeconomic disparities persist within the United States, suggesting there is room for more progress.[1,2,3]

Beginning in 2009, federal policymakers began a new evidence-backed effort to invest in educational programs related to teen pregnancy prevention. The federal government created two complementary programs: the Teen Pregnancy Prevention (TPP) Program and the Personal Responsibility Education Program (PREP).[4] The programs have a shared goal—to reduce teen pregnancies—but operate with different grant-making structures and strategic coordination to achieve public-policy objectives. Creating the programs was a major milestone. Previously, the only dedicated federal funding for reducing teen pregnancy was not based on criteria about effectiveness; it was based exclusively on content, namely abstinence-only approaches.

While Congress has introduced bipartisan legislation since 1999 to fund evidence-based programs and build knowledge about effective strategies, TPP and PREP were the first to gain traction.[5] They were propelled by the broader movement toward evidence-based policymaking.

Nearly a decade after the launch of TPP and PREP, these two programs provide vital information and skills to some of the most marginalized youth in the country. They also contribute to producing knowledge about effective teen pregnancy prevention efforts. These programs offer important lessons both for the power of using evidence to inform policymaking and the potential obstacles that can derail these programs and others informed by evidence.

ISSUE BACKGROUND

Both TPP and PREP are examples of tiered evidence-based programs. Tiered evidence is a framework typically applied to grant-making where policymakers allot a share of program funds to replicate models based on evidence of effectiveness in past studies. Policymakers allocate other funds to further strengthen evidence about other potentially effective models (to distinguish from the overall TPP and PREP programs or funding streams, we use "model" to refer to a specific curricula or educational program that has been or can be evaluated to assess impacts). As a result, most funding supports replication of successful models, while the broader program continues to build the overall evidence base through high-quality evaluation, continuous improvement, and innovation.

Teen Pregnancy Prevention Program

TPP is administered by the U.S. Department of Health and Human Services (HHS) Office of Adolescent Health. In TPP's first year, the program received a $110 million appropriation. In the subsequent nine years, Congress appropriated between $98 million and $101 million for the program.

In order to support strong TPP implementation and assessment, up to 10 percent of program funding can be used for training and technical assistance, evaluation, and program support. Of the remaining funds, 75 percent is available for "replicating programs that have been proven effective through rigorous evaluation" to reduce teenage pregnancy and other relevant behavioral risk factors. This funding is awarded competitively to what are called "Tier 1 grants." Tier 1 grantees can select from a broad array of previously evaluated models that best meet the needs of their communities. As part of the selection process, some grantees choose models with a focus on abstinence, others choose models that teach about both abstinence and contraception. Still others focus on youth development, parent-teen communication, or healthy relationships. What these have in common is an objective, high standard of evidence.

Recognizing the continued need for innovation as the characteristics and behaviors of adolescents change over time, an additional 25 percent of TPP funding supports competitive grants to develop, replicate, refine, and test additional strategies (called "Tier 2 grants").[6] A wide variety of organizations, including universities, local health departments, faith-based organizations, and community organizations, administer both the Tier 1 and Tier 2 competitive grants.

Personal Responsibility Education Program

The HHS Administration on Children, Youth, and Families administers PREP. The program receives $75 million annually to educate adolescents on both abstinence and contraception—with the goal of preventing pregnancy and sexually transmitted infections. The program also supports activities that prepare adolescents for adulthood, including topics such as maintaining healthy relationships, communicating with parents, and developing financial literacy.[7]

PREP focuses on youth at greatest risk of teen pregnancy (including foster youth, homeless youth, youth with HIV/AIDS, youth who are victims of human trafficking, and youth under 21 years of age who are pregnant or parenting) and geographic areas with high teen birth rates.[8] Most funding is formulaically allocated to states to replicate programs with existing research proving they are effective at changing behaviors (that is, delaying sexual activity, improving use of birth control, or reducing pregnancies). While the evidence criteria for PREP is slightly more flexible than for TPP, states take the commitment to evidence seriously, and they dedicated over 90 percent of their grant funding for models from the HHS evidence review.

Consistent with a tiered framework, in order to support innovation, HHS awards $10 million for organizations to compete to develop and test innovative strategies (PREP's version of Tier 2).[9] In addition, HHS sets aside $3.5 million annually for grants to Indian tribes or tribal organizations. Similar to TPP, PREP reserves 10 percent of its funding for training and technical assistance, evaluation, research, and dissemination of promising practices to reduce teen pregnancy.[10]

Together, the complementary TPP Program and PREP—with a combined investment of $185 million—serve nearly 350,000 youth annually.[11,12] The programs strengthen the capacity of states, tribes, and communities to replicate, implement, and evaluate evidence-based programs. Importantly, the programs generate useful results about what works to reduce teen pregnancy for different youth and in different settings.

Given the nature of these programs, and a historical focus on abstinence-only approaches, policymakers sometimes view these activities with an ideological perspective, rather than predominantly through a lens of funding the most effective approaches.

EVIDENCE AVAILABILITY

Before 2010, Congress did not appropriate federal funding to evidence-based approaches to reduce teen pregnancy. Evidence building began with private, philanthropic, and academic commitments.

Systematic reviews

In 1997, the National Campaign to Prevent Teen Pregnancy (now Power to Decide) commissioned an evidence review of effective strategies for preventing teen pregnancy.[13] The review found that, with a few exceptions, most studies assessing the impact of programs to reduce teen sexual risk-taking failed either to measure or to find sustained long-term impacts on behavior. A subsequent research review in 2001 identified additional research about programs that sustained positive behavioral effects.[14] The review offered stakeholders a few identified effective strategies; but finding funding for such programs was a challenge. By 2007, 15 strategies demonstrated evidence of positive and meaningful impacts on reducing adolescent sexual behavior, pregnancy, or sexually transmitted disease (STD).[15]

To support implementation of TPP and PREP, HHS commissioned an independent, systematic, and rigorous review of evaluation studies to establish the HHS Teen Pregnancy

Prevention Evidence Review.[16] The initial HHS effort in 2010 reviewed hundreds of studies of individual models and identified 28 that changed teens' sexual behavior or teen pregnancy. Because the HHS Evidence Review identified the models, they were eligible for use by grantees as part of TPP's Tier 1.

HHS regularly updates the Evidence Review to incorporate new research, including from TPP and PREP grantees as well as from other sources. The menu of effective models continues to grow and reached 48 in April 2018.[17] This expanding menu gives communities more options to best meet the needs of particular young people and implementation sites. For example, some communities want to provide longer-term strategies, while others are looking for relatively quick interventions. Communities may need models that are developmentally appropriate for youth in middle school, high school, or out of school. Or, they may want to focus on underserved or high-risk populations, such as youth in foster care or the juvenile justice system.[18] To assist program grantees and others in the field to select appropriate models that fit their needs, HHS offers a searchable database that categorizes models by populations, settings, and other characteristics, along with training and technical assistance for grantees, including online resources.[19,20]

Changing human behavior is hard, and so is conducting rigorous evaluation that detects such changes. Against this backdrop, the evidence generated from TPP and PREP to date is impressive in its scope, process, and results. In 2017, the bipartisan U.S. Commission on Evidence-Based Policymaking highlighted TPP as an example of a federal program developing increasingly rigorous portfolios of evidence.[21]

Evidence generated by TPP and PREP grantees

It is ambitious and can even be risky for a government-funded social program to make research about its effectiveness publicly available. Releasing the results of what worked, and what did not, then using that knowledge to inform current projects could lead to perceptions about program failure, as well as concern among program developers and grantees about exposing their efforts to rigorous evaluation when evaluations do not produce the expected results. These programs illustrate both the benefits and risks, especially in an environment where ideological views about program content run strong, despite the focus on evidence.

Grantees conducted the first round of 41 grantee evaluations of 32 individual models from 2010 through 2015. All of the evaluations used experimental research designs to determine program impacts. HHS released the final results in 2016. Of the models evaluated, one-third were part of TPP's Tier 1. Through Tier 1 evaluations, evidence-based teen pregnancy prevention models could be replicated and evaluated in different settings or with different populations than prior evaluations. Four of the 10 models positively impacted behavior in Tier 1.

Grantees rigorously tested another 22 models as part of TPP's and PREP's Tier 2 grants, including evaluating new and innovative approaches.[22] Eight of the 22 models tested in Tier 2 were effective. As a result of the number of evaluations, several academic journals dedicated volumes to discussing both TPP and PREP. These included articles about program results as well as implementation and evaluation lessons.[23,24,25]

These conclusions must also be met with caution about how effective grantees were overall. Of all the models evaluated as part of the first-round research, 12 out of 32 (38 percent) showed positive impacts. This is well above the 10 to 20 percent of rigorous evaluations (typically experimental evaluations with random assignment) that experts say typically demonstrate positive results. In addition, 13 of the 41 individual studies were inconclusive because of implementation challenges or methodological limitations (note: Grantees evaluated some of the 32 models in more than one place, which is why there were 41 studies of 32 models).[26] This illustrates a challenge for conducting impact evaluations, especially with smaller or underserved populations, such as rural or Native American teens, where getting adequate sample sizes can be difficult. But the grantees collectively contributed to developing a robust new body of research and evaluation to support future implementation of the tiered evidence framework in TPP and PREP.

EVIDENCE USE

In keeping with the tiered evidence framework for the programs, lessons and insights gained from existing evaluations immediately began to feed back into the future cycles of funding for TPP and PREP. HHS uses evidence as a criterion to award funding, and in turn, HHS expects the programs to continue generating new evidence. After completing the grantee evaluations, HHS shared the results and the Office of Adolescent Health actively encouraged the second round of TPP Tier 1 grantees awarded in 2015 to shift toward replicating the most effective models based on the latest knowledge about what works best in different contexts. Grantees adjusted their plans, which resulted in a stronger mix of models than grantees used in the first round.[27] In addition, the Tier 2 evaluations from the first round contributed to eight new models that met Tier 1 standards, and current Tier 1 grantees are now replicating some of those models.

The misuse of the evidence

Despite longstanding bipartisan support for TPP and PREP and recognition by experts that these programs are strong examples of a tiered evidence framework that have contributed to evidence building, the Donald Trump administration has tried to make dramatic changes to TPP. While they have had mixed success in doing so, an unfortunate theme in the administration's actions is a misuse of evidence—including mischaracterizing results, using evidence as a weapon rather than a tool, and using lower standards of evidence than Congress intended. First, the administration proposed cutting the program funding, though Congress did not agree and continued to fund the program. Second, HHS abruptly cut short the final two years of the second round of five-year grants, without providing grantees notice or explanation. The reaction from Congress and program advocates was swift, raising questions about the rationale for the decision.[28] In addition, four federal judges ruled that shortening the grants was illegal and funding was ultimately restored for the last two years of the grants.[29] Third, HHS took steps through a new funding opportunity announcement to redirect TPP funding to grants with lower standards of evidence and evaluation than Congress intended and that the past administration had used. This too raised objections from some members of Congress, and two courts found the Tier 1 funding opportunity announcement illegal.[30]

In trying to justify shortening the second round of grants, the Trump administration responded by citing results from the first round of grantee evaluations, claiming that TPP failed or may have even harmed young people. In fact, the decisions had been based on longstanding pressure from advocates in outside groups—some of whom ended up in the administration—who called for defunding TPP and instead funding what was previously called abstinence-only programs, now rebranded as "sexual risk avoidance," or SRA (typically described in statutes as programs that urge young people to "refrain from non-marital sexual activity").[31]

Stakeholders noted that the Trump administration mischaracterized the results and lacked context for what researchers typically expect to find when replicating rigorous evaluations.[32] So while "using" some of the existing evidence in their assertions, the uses did not reflect widely accepted stakeholder and evidence experts' perspectives of the overall TPP success or design to improve and expand use of effective models. Moreover, they misused results from individual evaluations to claim the overall TPP Program failed, when in fact learning what does and does not work is part of an evidence-based program's goal. In addition, they misused results from prior evaluations as a weapon to justify cutting current grants that had been strengthened by learning from those results and were in the process of generating new evidence.

Meanwhile, PREP has not faced the same kind of challenges. In contrast to TPP, the administration supports continued funding,[33] and Congress continues to fund it as part of a package that includes parallel funding for abstinence or SRA grants to states. So far, the administration has not made visible efforts to redirect the program focus or use of evidence.

LESSONS

- *Incorporating evidence early on can deter partisanship.* The fact that funding levels and legislative language governing both TPP and PREP remained fundamentally consistent for nine years, with bipartisan support in Congress—and continued despite efforts to undermine TPP—is a testament to the power of an approach that incorporates evidence into the policymaking and practices of grantees. At the same time, TPP offers a cautionary tale of how ideology can potentially derail efforts to use evidence during program implementation and how evidence can be misused in the pursuit of ideological objectives.

- *Programs cannot let the fear of potential findings stop them from evaluating in the first place.* A commitment to rigorous evidence standards and evaluation is hard and involves risk for programs, model developers, and researchers. If those involved fear that the result of any rigorous evaluation that does not demonstrate a positive impact will be weaponized against that model or even an entire program, they will be far less likely to subject their models and programs to this rigorous evaluation in the first place. The evidence-based policy movement has more work to do to

manage expectations and put results in context—even before evaluations begin. Doing so will help to build a safe space for evidence and evaluation. TPP was the first of the federal tiered-evidence programs to do a large-scale release of evaluation results. Experts and advocates had to work hard to put the results in context after completing the evaluations. Managing expectations from the start about the meaning of the results might have mitigated some of the misuses of available evidence. Policymakers, the press, and the public do not always understand that a lack of positive impact is common in rigorous evaluations and that what appears to be failure can actually be part of a success story. Indeed, people frequently misconstrue the difference between null results (no statistically significant positive or negative impact) and negative results. In the case of TPP, several studies had null results, but only a handful had negative results. Evidence-based policymaking is about learning from what does and does not work, and then using that data to inform the focus of future efforts, again and again. It is an iterative process. Similarly, it is important to clearly differentiate between the success of overall programs such as TPP and PREP—which are clearly examples of high-quality, evidence-based policymaking—and individual models that are evaluated within these programs.

- *Instilling a culture of continuous learning.* A steadfast commitment to continuous improvement and learning from results is essential if the American public is to see the full benefit of evidence-based policymaking. It requires patience on the part of policymakers and other stakeholders to ensure programs have the time and sustained investments to learn from results and build on them, rather than cutting off research midstream or declaring programs a failure before results from ongoing replications and research are available. For example, the past Tier 2 grants have already yielded impressive new evidence to help meet the changing needs of young people, and those underway have the potential to yield even more evidence if program disruptions do not compromise the research.

- *Building evidence is not cheap, but it can be worth it.* Supporting and scaling a tiered evidence program such as TPP or PREP in multiple sites and with a diverse array of grantees with different levels of evaluation expertise cannot be done on the cheap. It requires an investment in training, technical assistance, and capacity building; it also benefits from opportunities for grantees to share best practices and to learn from each other.

- *Evidence can solve problems but not change beliefs.* Even evidence-based programs are affected by ideology and values in political decision-making processes. When a favorite model is subjected to rigorous evaluation, you might not like the results. For example, some popular models that teach both abstinence and contraception to young people did not replicate the strong results shown in earlier evaluations. Asking those tough questions, and not always liking the results, is critically important. The stakeholders of the program have to continually support the process of learning and adjust accordingly. Keeping the focus on evidence and results, rather

than on a particular ideology or approach to content, helped broaden political buy-in for TPP. For example, some conservative policymakers found it extremely helpful to know that some programs that taught about both abstinence and contraception helped young people wait longer to have sex. In other words, abstinence was a behavioral result of programs that were not abstinence-only. Even if both sides do not love everything about the program, when it is framed as an evidence-based program rather than one defined by content, policymakers on both sides of the aisle have more cover to support the program—or to at least take a neutral stance.

- ***Teaching evidence-based policymaking is just as important as doing it.*** Educating key audiences about what evidence-based policymaking is and what it means—everyone from policymakers, press, and the public—helps explain a program's story and impact better. TPP and PREP grantees had to learn how to explain the "what" of their work and tell that story to policymakers, the press, and the public in a credible way. ■

1   U.S. Centers for Disease Control and Prevention. "Reproductive Health: Social Determinants and Eliminating Disparities in Teen Pregnancy." Last reviewed January 28, 2019.
    Available at: https://www.cdc.gov/teenpregnancy/about/social-determinants-disparities-teen-pregnancy.htm.

2   International Comparisons. "International Teen Pregnancy Statistics."
    Available at: http://internationalcomparisons.org/social/teen-pregnancy.html.

3   United Nations Statistics Division. *Demographic Yearbook 2003* (New York: United Nations, 2015), 331-353.

4   Adrienne L. Fernandes-Alcantara. *Teen Pregnancy: Federal Prevention Programs*. U.S. Library of Congress. Congressional Research Service, R45183, 2018. Available at: https://fas.org/sgp/crs/misc/R45183.pdf.

5   *See for example:* Teenage Pregnancy Reduction Act of 1999 (H.R. 1636) in the 106th Congress; Preventing Teen Pregnancy Act (H.R. 3581) in the 110th Congress; and Putting Prevention First Act (S. 2336 and H.R. 4192) in the 108th Congress.

6   *Department of Defense and Labor, Health and Human Services, and Education Appropriations Act, 2019 and Continuing Appropriations Act, 2019.* Public Law 115-245, 2019.

7   Personal Responsibility Education, U.S. Code 42 §§ 713, and *Bipartisan Budget Act of 2018.* Public Law 115-123, 2018.

8   While TPP funding announcements and guidance also focus on high need populations and communities, the PREP statute specifically mentions these groups.

9   Administration for Children and Families. "Personal Responsibility Education Innovative Strategies Program (PREIS)." Family & Youth Services Bureau. Available at: https://www.acf.hhs.gov/fysb/programs/adolescent-pregnancy-prevention/programs/preis.

10 For a thorough review of the history of TPP Program and PREP, see CRS report below as well as the funding for abstinence and sexual-risk avoidance programs. Adrienne L. Fernandes-Alcantara. *Teen Pregnancy: Federal Prevention Programs.* U.S. Library of Congress, Congressional Research Service, R45183, 2018.

11 For the latest numbers from TPP, see: U.S. Department of Health and Human Services. "Performance Measures Snapshot: The Teen Pregnancy Prevention Program: Performance in Fiscal Year 2017 (Year 2)." Office of Adolescent Health, October 2017.
    Available at: https://www.hhs.gov/ash/oah/sites/default/files/tpp-performance-measures-year-2-brief.pdf.

12 For the latest numbers from PREP, see: Adrienne L. Fernandes-Alcantara. *Teen Pregnancy: Federal Prevention Programs.* U.S. Library of Congress, Congressional Research Service, R45183, 2018, 26.

13 Doug Kirby. *No Easy Answers: Research Findings on Programs to Reduce Teen Pregnancy.* National Campaign to Prevent Teen and Unplanned Pregnancy, 1997.

14 Doug Kirby. *Emerging Answers: Research Findings on Programs to Reduce Teen Pregnancy.* National Campaign to Prevent Teen Pregnancy, 2001.

15 Doug Kirby. *Emerging Answers 2007: Research Findings on Programs to Reduce Teen Pregnancy and Sexually Transmitted Diseases*. National Campaign to Prevent Teen and Unplanned Pregnancy, 2007.

16 U.S. Department of Health and Human Services. "Welcome to the Teen Pregnancy Prevention Evidence Review." Available at: https://tppevidencereview.aspe.hhs.gov/. The Office of the Assistant Secretary for Planning and Evaluation manages this evidence review; it is consistent with best practices in the evidence field and is included in the Pew-MacArthur Results First Clearinghouse Database.

17 Julieta Lugo-Gil, et al. *Updated Findings from the HHS Teen Pregnancy Prevention Evidence Review: August 2015 through October 2016*. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, 2018.
    Available at: https://tppevidencereview.aspe.hhs.gov/pdfs/Summary_of_findings_2016-2017.pdf.

18 The profiles of grantees published by the Office of Adolescent Health reflects the diversity of approaches. *See*: U.S. Department of Health and Human Services. "TPP Successful Strategies." Office of Adolescent Health. Last reviewed December 15, 2016. Available at: https://www.hhs.gov/ash/oah/grant-programs/teen-pregnancy-prevention-program-tpp/successful-strategies.

19 U.S. Department of Health and Human Services. "Teen Pregnancy Prevention Evidence Review: Programs." Available at: https://tppevidencereview.aspe.hhs.gov/EvidencePrograms.aspx.

20 U.S. Department of Health and Human Services. "Using Evidence-based Programs." Available at: https://www.hhs.gov/ash/oah/resources-and-training/tpp-and-paf-resources/using-evidence-based-programs/index.html.

21 U.S. Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking*. U.S. Commission on Evidence-Based Policymaking, 2017.
    Available at: http://cep.gov/content/dam/cep/report/cep-final-report.pdf.

22 Amy Feldman Farb and Amy L. Margolis. "The Teen Pregnancy Prevention Program (2010-2015): Synthesis of Impact Findings." *American Journal of Public Health*, 106(S1): S9-S15, 2016.

23 Ibid.

24 Andrea Kane. "Adolescent Pregnancy Prevention in Undeserved Populations: The Way Forward." *American Journal of Public Health*, February 27, 2018. Available at: https://ajph.aphapublications.org/toc/ajph/108/S1.

25 Evelyn Kappeler. "Building the Evidence to Prevent Adolescent Pregnancy." *American Journal of Public Health*, 106(S1), September 2016.

26 Ron Haskins and Andrea Kane. *The Trump Administration Deals a Blow to Evidence-Based Policy*. Evidence-Based Policymaking Collaborative, October 2018.

27 Ibid, 7-8, 15-16. The mix of programs shifted in three ways. First, one model (the Teen Outreach Program), which the first cohort widely replicated without successfully replicating results in most new settings, was used only on a very limited basis in the second cohort. Second, the evidence review added a number of new models based on emerging evidence from Tier 2 evaluations in the first cohort that were then selected for replication as Tier 1 projects in the second cohort. (For example, at least nine TPP sites are replicating a new Tier 1 healthy relationship model called Love Notes.) Third, OAH disallowed replication in the second cohort of one model (Cuidate) that had produced negative results in the first cohort and allowed one other (It's Your Game: Keep It Real) to be replicated only under narrow circumstances that were different than the first cohort evaluation.

28 Ibid.

29 Megan Uzzell and Charisma Troiano. "One Year of Successful Battles to Protect the TPP Program Against Trump Administration Unlawful Actions; Fights Remain as Administration Continues its Assault on Evidence." *Democracy Forward*, April 11, 2019. Available at: https://democracyforward.org/press/trump-administration-continues-unlawful-effort-to-dismantle-the-evidence-based-teen-pregnancy-prevention-program/.

30 Ibid. While the government dropped its appeals in two of the Tier 1 funding opportunity announcement (FOA) cases, two legal appeals are still pending—one challenging both the Tier 1 and Tier 2 FOAs, and another challenging Tier 2.

31 Heidi Pryzbyla. "Notes, Emails Reveal Trump Appointees' War to End HHS Teen Pregnancy Program." *NBC News*, March 20, 2018. Available at: https://www.nbcnews.com/politics/politics-news/notes-emails-reveal-trump-appointees-war-end-hhs-teen-pregnancy-n857686.

32 Robert Gordon and Ron Haskins. "Trump Team Doesn't Understand Evidence-based Policies Regarding Social Problems." *The Hill: Pundit's Blog*, July 26, 2017. Available at: https://thehill.com/blogs/pundits-blog/the-administration/343908-trump-team-doesnt-understand-evidence-based-policies.

33 The president's fiscal year 2020 budget proposes one year of funding; Congress most recently extended it for two years, through fiscal year 2019, in the *Bipartisan Budget Act*. Public Law 115-123.

18. THERE'S NO PLACE LIKE HOME

Improving Outcomes One Home Visit at a Time

Lauren Supplee is the deputy chief operating officer of early childhood development at Child Trends. She is a former director of division of family strengthening at the Administration for Children and Families in the Office of Planning, Research and Evaluation.

Ted McCann is a fellow at the Bipartisan Policy Center. He was previously assistant to the speaker for policy, managing welfare, education, workforce, pension, telecom, and Social Security issues for former House Speaker Paul Ryan.
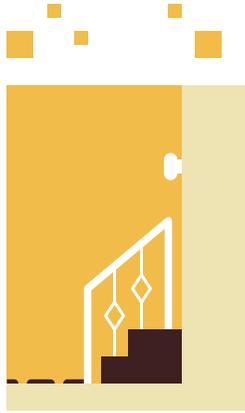
Home visiting is a service delivery strategy that aims to improve a range of outcomes for families with children, including school readiness, maternal and child health, child maltreatment prevention, family economic self-sufficiency, and others.[1] Typically, the model involves sending a trained professional or paraprofessional to visit a household to provide education, support, and referrals to needed community services in order to improve outcomes for children and families. There are numerous models for deploying home visiting approaches that guide structure, procedures, and substantive content of the services for local programs.[2]

The home visiting system in the United States is largely decentralized and supported with funding from multiple levels of government. In 2009, a range of federal, state, and philanthropic sources spent an estimated $500 to $750 million in the United States on early childhood home visiting.[3]

Congress and the White House established a new approach for home visiting in 2009, as part of the Patient Protection and Affordable Care Act. The law provided $1.5 billion in federal funding for a new strategy for conducting home visiting: a tiered evidence approach through the new Maternal Infant and Early Childhood Home Visiting Program (MIECHV). This case study focuses on how evidence was built into the design of the program and then used to justify the program's reauthorization in 2018, extending the program for another five years.

ISSUE BACKGROUND

Congress and the White House designed MIECHV to encourage both the production of evidence and the allocation of program funds predominantly toward strategies that were validated and supported by existing research. The approach, called tiered evidence, was also applied to other grant programs during the Obama administration.

The basic concept of tiered evidence in MIECHV is that the law set up the executive branch as a user of the existing evidence to determine which home visiting models qualified for which proportion of funding. The federal government required 75 percent of the grants it distributed to providers to use models that demonstrated evidence of effectiveness. Providers could use another quarter of program funding for promising models that had not yet been subjected to an impact evaluation, but the grantee had to simultaneously conduct an evaluation to determine if the promising approach could be funded as evidence-based going forward.

Beyond the funding allocation, other strategies encouraged the executive branch staff to integrate multiple forms of evidence into MIECHV. In 2017, the White House's Office of Management and Budget encouraged executive branch agencies to use this type of design as a strategy to support continuous learning and improvement.[4] Even in advance of the reauthorization debates for MIECHV, the program was ready to provide a range of evidence to inform program administration and legislative reauthorization discussion.

EVIDENCE AVAILABILITY

For the reauthorization debates in 2018, policymakers had access to performance information, impact evaluations, and numerous implementation studies. All of the evidence is available through publicly accessible websites, newsletters, social media, presentations at research conferences, and congressional or executive branch staff briefings. MIECHV had such a strong emphasis on data, research, and evaluation that producing relevant evidence was not just part of the program design; it was embedded in the culture of the program for managers, grantees, and participants throughout the country.

Program performance statistics

MIECHV required grantees to report on a series of performance measures on a wide range of outcomes: maternal and child health, child maltreatment, school readiness, crime and domestic violence, family economic self-sufficiency, and improved referrals. MIECHV expected grantees to make improvements in four out of six of these domains in the first three years. If grantees did not accomplish this, they would receive technical assistance to correct identified deficiencies. The information collected was instrumental in providing basic program characteristics and knowledge about grantee trends and conditions. In general, the performance measures suggested the program was meeting the statutory goal of expanding the reach of home visiting to the nation's most vulnerable children and families living in at-risk communities. The data also indicated the majority of grantees demonstrated improvement across the domains over time.[5] The performance measures informed how technical assistance was targeted to areas grantees found challenging, such as dealing with intimate partner violence or family enrollment and retention. These areas also became the focus of continuous quality improvement efforts.

National impact evaluation

Congress required MIECHV to conduct a national impact evaluation using an experimental design. The national evaluation—called the Mother and Infant Home Visiting Program Evaluation—was an ambitious endeavor to measure impacts for programs using consistent measures at the same points in time and, for some, was the first independent evaluation of their model. The impact evaluation, which launched in 2011, only had interim results available by the time Congress considered the 2018 reauthorization.[6] Impact results when the children were 15 months old were made available in 2019, documenting significant positive impacts in parenting, parent mental health, and child behavior.[7] The evaluation is ongoing and will have additional impact results available when the children enter kindergarten.

MIECHV grantee evaluations

As part of the tiered evidence model, MIECHV required grantees to conduct evaluations that addressed policy-relevant questions. HHS provided technical assistance to ensure those evaluations were rigorous. These evaluations were important as they signaled to state and tribal grant administrators that the ongoing investment in evaluation activities was central to this program. It created ownership of evaluation findings because grantees designed evaluations to answer critical implementation and policy questions facing their state or tribe versus priorities of the federal government. For example, these evaluations answered states' questions about building systems to better connect families to needed programs, workforce development, and evaluating the effectiveness of promising approaches.[8]

Implementation science

The law also drew indirectly from research on effective implementation of evidence-based programs mandating features such as a well-trained workforce, reflective supervision, data systems, and continuous quality-improvement methods. HHS directly leveraged the language related to continuous quality improvement. More specifically, all grantees must establish continuous quality-improvement plans, which establishes the use of data for improvement as a core feature of the program. In 2013, the Home Visiting Collaborative Innovation and Improvement Network (HV CoIIN) launched; it uses methods of quality improvement from the health care sector. Grantees analyzed data to identify areas for program improvement, set goals for improvement, and rapidly and iteratively tested innovations to identify service improvements. State administrators reported that the experience of participating in the HV CoIIN embedded the use of data into daily practice and highlighted areas for program and policy improvement.[9]

Systematic review

In order to assess the availability and quality of the research on the home visiting models in a fair and impartial manner, HHS conducted an independent systematic review of the existing evidence. During the first year of the Home Visiting Evidence of Effectiveness research review, HHS identified over 10,000 research articles and over 250 potential home visiting models.[10]

The studies included a mix of evaluation designs. The review could not assess all of the available literature at the same time. Because of the breadth of information, HHS developed a prioritization scheme to ensure the models with the most evidence would be reviewed first.

In the first year, seven models met the review criteria for quality-of-study design and execution on relevant outcomes. In 2018, 20 models of the 250 identified met the criteria for evidence of effectiveness. The Home Visiting Evidence of Effectiveness is still ongoing with many more models to review.

Partnerships for building evidence

One of the primary, likely unintentional, consequences of the statutory requirements for evidence in MIECHV was the forced close partnership between the research office and the programmatic office. Before MIECHV, it was more typical than not that research and program executive branch staff operated in silos that sometimes came together, but it was often intermittent and often focused on a specific study. The depth and breadth of the evidence requirements in MIECHV meant research and practice worked together throughout.

The Administration for Children and Families and the Health Resources Services Administration worked jointly on many aspects of MIECHV implementation that involve research and data. For example, staff from both offices designed (and re-designed) the performance measures over time, and representatives from both offices attended calls with the contractors for projects such as the systematic review and the national MIECHV impact evaluation to ensure the research and practice voices were included.

In another example, staff from both offices developed a learning agenda for building evidence over time for the program as a whole The learning agenda includes a diverse portfolio of evidence-building activities that answer questions for the program both in the near term and the long term, ranging from descriptive data and research, to capacity building in research, to impact research, to answering the program's pressing questions.

EVIDENCE USE

With an impressive array of existing research, evaluation, and performance information, it would have been a mistake to continue to authorize a program based only on the evidence available. The actions of Congress to encourage a robust portfolio of evidence into MIECHV allowed the program to be designed with learning from evaluation at its core, which will continue to improve the program and its outcomes into the future.

Importantly, and without a doubt, the broad portfolio of evidence available to decision-makers about MIECHV was instrumental in securing the program reauthorization, based on allocation of funding toward interventions that work as well as positive performance information and an ever-growing body of research to validate and improve the models used in the field.

During congressional consideration of the MIECHV program, there was strong bipartisan support from members of Congress for the program. In fact, as other reauthorizations were under consideration (such as the *Families First Prevention Services Act*), congressional staff used MIECHV as an example of how to construct an evidence-based program. As a result of this strong support, when negotiating the addition of the MIECHV program to the Bipartisan Budget Act of 2018, the "four corners" of leadership—that is, then-Speaker Paul Ryan, Senator Majority Leader Mitch McConnell, then-Minority Leader Nancy Pelosi, and Senate Minority Leader Chuck Schumer —all agreed to include the MIECHV program without asking for anything in return.

The key controversy surrounding the reauthorization of MIECHV in Congress was how to pay for the program's extension. A five-year extension would cost $400 million per year, and offsetting that cost became the largest source of controversy. Republicans generally wanted to see offsets come from the welfare space; Democrats wanted to see offsets come from outside of the welfare space.

Support for MIECHV was so strong, however, that the offset mix eventually included "pay fors" from two different areas: health and welfare programs. Even within the welfare offsets, evidence played a key role. Congressional negotiators settled on expanding the Reemployment Services and Eligibility Assessments program. The Congressional Budget Office viewed the evidence as strong enough that it even noted the budgetary savings that occurred in other, related programs. This means that the evidence available provided a strong basis for the various congressional actors to perceive it both as worthwhile and able to produce meaningful cost reductions to government expenses in other areas of the budget.

Congressional staff found the additional funding for the program through an offset from the health industry. The health industry has one of the most effective lobbying operations in D.C., and so the support for using offsets within this sector to pay for MIECHV shows the broad, deep, bipartisan support for the MIECHV program.

LESSONS

After implementing MIECHV, the executive branch researchers and program staff learned many lessons about how evidence use and how it impacts program changes. Below is a small handful of lessons learned.

- *Legislative requirements should balance specificity with flexibility.* To support a portfolio of evidence and learning, legislative language needs to both provide parameters for quality and flexibility to adapt to the evidence needs over time. Legislative language can both set a bar for study quality and also, unintentionally, limit how evidence is built over time, which may limit the utility of the evidence

for policymakers. The statutory language related to the evidence criteria needs to balance specificity of research designs while allowing evidence standards to evolve and grow to follow the state of the field and best practices. Examples include:

1. Legislative language that pre-specifies methods or thresholds for statistical significance does not align with professional best practices.[11]

2. The requirement to use "peer-reviewed" research, which implies that peer review is a proxy for high-quality research. Unfortunately, much of the peer-reviewed literature may still not meet the quality of the study design and execution required by HomeVEE. In addition, a peer-review requirement can inadvertently prohibit the inclusion of government reports, which can be just as rigorous, or more, than peer-reviewed papers.

3. The MIECHV language uses the word "models," which constrains the systematic review from including the evidence on other elements of the program that have emerged over time, such as program enhancements, targeted content regarding intimate partner violence, or services using telehealth models. These enhancements also contribute to the overall effectiveness of the program, and the field should know about this evidence base.

- *Interpreting a body of evidence.* Executing a portfolio of evidence and learning agenda means building the capacity of legislative and executive branch employees to understand how to interpret a body of evidence that may not be straightforward. Evaluations often provide mixed findings about outcomes, making definitive declarations a challenge for users. Evaluations often conclude that outcomes and impacts are mixed, including within the same study or across studies. Evidence from a single evaluation—positive, negative, or null—is not sufficient to universally deem a policy or intervention evidence-based. Often multiple studies of models are testing an approach with different populations, outcomes, or locations. A model may have impacts in one study but not in another. The reason for the discrepancy may be due to the model, but it also may be due to other factors. The complexity of the evidence reinforces the importance of a strong partnership between the research and practice staff to make decisions about the use of evidence.

- *Evolving evidence.* The current set of models for home visiting are complex and expensive, which adds to the challenge for decision-makers in deploying the models. This may change over time with further development and innovation. Once deemed evidence-based, models continue to evolve and improve and are not stagnant or set in stone. Federal programs need the flexibility to fund and use a portfolio of evidence to support a learning agenda, including performance measures, descriptive studies, evidence reviews, impact studies, and innovative research. For example, the Health Resources and Services Administration funds the Home Visiting Applied Collaborative to examine whether the methods behind

precision medicine can be applied to home visiting to make models more efficient and effective. While overall home visiting can achieve impacts, researchers need to identify ways to make impacts stronger for the complex set of families served by the programs. This research needs to leverage research designs that get answers to these questions more quickly.

- **Well-executed evidence-based policy can improve outcomes.** Evidence-based policy is a lot more complicated than it seems. Implementing evidence-based policy is not straightforward or easy, and there are many areas of nuance and continued learning. It is like Winston Churchill's famous quote about democracy being the worst form of government except for all the other forms of government. Evidence-based policy is not perfect, but the more we can start from what's known, build close partnerships between research and practice, and adopt a culture of curiosity and learning in social services, the better chance we will achieve the outcomes we truly want for children and families. ◼

1   Charles Michalopoulos, et al. *Design Options for the Maternal, Infant, and Early Childhood Home Visiting Evaluation: Draft Final Report.* U.S. Department of Health and Human Services, April 2011.
     Available at: https://www.acf.hhs.gov/sites/default/files/opre/dohve_design.pdf.

2   Ibid.

3   E. Stoltzfus and K. Lynch. *Home Visitation for Families with Young Children.* Congressional Research Service, 2009.

4   White House Office of Management and Budget. "Building and Using Evidence to Improve Government Effectiveness." In: *Analytical Perspectives: Budget of the U.S. Government Fiscal Year 2018* (Washington, D.C.: Government Printing Office, 2017), 55-58.
     Available at: https://www.govinfo.gov/content/pkg/BUDGET-2018-PER/pdf/BUDGET-2018-PER-4-2.pdf.

5   U.S. Department of Health and Human Services. *Demonstrating Improvement in the Maternal, Infant, and Early Childhood Home Visiting Program: A Report to Congress.* March 2011. Available at: https://mchb.hrsa.gov/sites/default/files/mchb/MaternalChildHealthInitiatives/HomeVisiting/pdf/reportcongress-homevisiting.pdf.

6   Charles Michalopoulos, et al. *The Mother and Infant Home Visiting Program Evaluation: Early Findings on the Maternal, Infant, and Early Childhood Home Visiting Program.* Office of Planning, Research and Evaluation, OPRE Report 2015-11, January 2015.
     Available at: https://www.mdrc.org/publication/mother-and-infant-home-visiting-program-evaluation.

7   Charles Michalopoulos, et al. *Impacts on Family Outcomes of Evidence Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Program Evaluation.* Office of Planning, Research and Evaluation, OPRE Report 2019-07, January 2019. Available at: https://www.mdrc.org/publication/impacts-family-outcomes-evidence-based-early-childhood-home-visiting.

8   Susan Zaid and Lance Till. *Overview of Grantee-Led Evaluations: The Maternal, Infant, and Early Childhood Home Visiting Program.* James Bell Associates, OPRE Report 2016-78, October 2016. Available at: https://www.acf.hhs.gov/sites/default/files/opre/overviewofgranteeledevaluationsmiechv102816_b508.pdf.

9   Education Development Center. *Home Visiting Collaborative Improvement and Innovation Network 1.0 (HV CoIIN).* 2017. Available at: http://hv-coiin.edc.org/sites/hv-coiin.edc.org/files/HV%20CoIIN%20Information%20Resource%202017_0.pdf.

10   D. Paulsell, S. Avellar, E. Sama Martin, and P. Del Grosso. *Home Visiting Evidence of Effectiveness Review: Executive Summary.* U.S. Department of Health and Human Services, 2017.

11   American Evaluation Association. "An Evaluation Roadmap for a More Effective Government."
     Available at: https://www.eval.org/evaluationroadmap.

19. SCHOOLHOUSE ROCK

Enhancing Educational Opportunities for Students

Jessica Ramakis is a senior advisor in the Office of Planning, Evaluation and Policy Development at the U.S. Department of Education.

Matthew Soldner is the commissioner of the National Center for Education Evaluation at the U.S. Department of Education's Institute of Education Sciences.
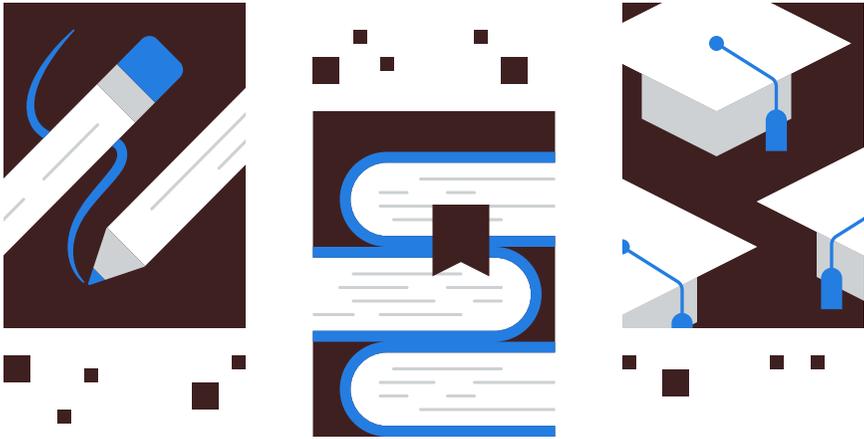
Research only improves education when its lessons are put into practice. For 40 years, the U.S. Department of Education has sought to better understand what works in education and to encourage the use of evidence in educational practice and policymaking. As calls to improve education systems have quickened, the department has begun to explore new ways to accelerate the adoption of evidence-based practices and augment its existing approaches to build evidence. This case study explores a critical lever in that effort: the department's grant-making portfolio.

Each year, the Education Department invests in education through formula and competitive grant programs that support states, school districts, institutions of higher education, and other entities around the country. Nearly $50 billion is available for education each year from the federal government, excluding student financial aid. That is, on average, about 9 percent of all education funding in the country. Of the amount from the federal government, the department awards about $4 billion per year competitively. These competitively awarded funds are where the Education Department can readily shape how program designs through strategically developed policies that encourage the use of evidence.

ISSUE BACKGROUND

Although how best to improve the education system is the subject of considerable national debate, most agree about the importance of doing so. To be sure, there are signs of progress. High school graduation rates are at an all-time high, for example, and many groups of students are going to college at near-historic rates.[1,2] But there is still work to do. When comparing internationally, students in the United States continue to lag behind their peers abroad in science, reading, and math literacy.[3]

Despite any disagreements about how to best support educational improvement, most policymakers can agree on one point: the funding, personnel, and time to address the challenge are limited. Because the available resources are finite, policymakers and educators need to know as much as possible about what works for students—and then need to use what they've learned—to use resources efficiently. Identifying what works in education is no easy feat. Learning is a complex process, as is defining what students ought to know and how-to best measure learning. The wide array of contexts in which teaching and learning happen further complicates the process of building evidence, making it hard to know if what is successful with one type of student, one type of educator, or one type of school will apply elsewhere. As a result, studying educational programs, policies, and practices with the most rigorous research and evaluation methods available is critical to the process of discovery and to increase the stock of evidence-based practices that educators can use to meet the needs of today's diverse learners.

EVIDENCE AVAILABILITY

In 2002, Congress passed the No Child Left Behind Act, with the goal that all students would be proficient in math and reading by the end of the 2013-2014 school year.[4] Weeks later, Congress passed the Education Sciences Reform Act of 2002, which established that the department would best meet its mission—improving student achievement and ensuring equal access to educational opportunity—by using research and evaluation to improve educational policy and practice. This law created the Institute of Education Sciences, including the National Center for Education Statistics, the National Centers for Education and Special Education Research, and the National Center for Education Evaluation and Regional Assistance (NCEE). Together, the centers within the Institute of Education Sciences support the creation and use of high-quality, independent, and unbiased evidence that is available to inform practices in education.

NCEE has a unique role in encouraging policymakers, educators, and other stakeholders to use evidence about policies, programs, and practices that improve student achievement. NCEE's What Works Clearinghouse reviews and synthesizes research about education policies, programs, and practices. The Clearinghouse provides reviews of nearly 11,000 individual studies and 600 distinct interventions. Collectively, these reviews offer assessments of interventions' potential impact on a range of important educational outcomes as well as the strength of evidence underlying the assessments. Summaries and syntheses are widely available to prospective grantees, district and building leaders, and other stakeholders via the Clearinghouse's website. These data also are foundational to many of the technical-assistance efforts that NCEE and other offices throughout the department lead.

EVIDENCE USE

Today, three distinct efforts come together to support improved education practice and policymaking at the state and district levels. These include the Education Department's competitive grant program portfolio as a whole, the subset of that portfolio that includes

grants that use tiered-evidence standards to support innovation, and targeted technical assistance to encourage the use of evidence-based practices.

Competitive grants

When considering how best to use evidence in the department's grant competitions, an internal collaboration of experts—program, policy, and budget experts supported by a team of researchers and evaluators—discusses the evidence that supports policy or practice inside a given program area. Consider the case of a program that supports services for students in postsecondary institutions. Is there, for example, a strong sense of effective policies and practices? Does that include a small set of policies and practices, or are there many options for policymakers or educators to consider? Alternatively, are there no, or perhaps only a few, practices to choose from? To address program evidence considerations, expert collaboration over a series of meetings is essential to better understanding the goals of the program and its key mechanisms, relevant policy or budget constraints, what has been learned through prior research, and what questions remain. The result of that discussion informs the parameters of the next year's grant competition, including program priorities, grantee evidence requirements, and evaluation criteria.

Conversations such as these lead to a greater focus on evidence among both program staff and grantees. For example, almost a third of the department's new fiscal year 2017 grant competitions included evidence in their design.[5] In our work with grantees, we are exploring how best to support evidence-building, the implementation of evidence-based practices, and taking what is learned from grantees to the broader education community to maximize the department's impact. Notable examples of this work include collaborations between program offices and NCEE in helping grantees (1) better understand ESSA's evidence requirements and how to use the What Works Clearinghouse to identify interventions supported by strong and moderate evidence; and (2) develop rigorous plans to evaluate their work through quasi-experimental or experimental methods.

Increasingly, lessons learned from the competitive grant programs influence the department's work with programs that receive federal grants by formula. States and districts have the flexibility to use federal education funds in support of the evidence-based policies and practices that best meet local needs. To support those efforts and as a result of identifying the need for clearer guidance, the department issued guidance to explain the appropriate use of evidence in education as a resource.[6]

Tiered evidence grants

Both the i3 program and the newer Education Innovation and Research program are examples of tiered-evidence grant programs within the department. The purpose of the Education Innovation Research is to support the creation and testing of innovative educational practices and to scale-up those practices that demonstrate evidence of impact through field-initiated grants. Prospective grantees apply for early phase, mid-phase, or expansion grants based on the level of evidence that supports the policy or practice they propose to support with the funding.

Results of the national evaluation of i3 suggest the program largely met its goals of advancing evidence use, scaling effective practice, and building evidence.[7] About three-quarters of the 67 i3 evaluations reviewed by the program's national evaluator conducted a high-quality, independent impact evaluation. Practices taken to scale were also more likely to be effective (50 percent yielding positive impacts) than those in earlier phases, including validation (40 percent yielding positive impacts) and development (8 percent yielding positive impacts). Taken together, 12 of 67 grants (18 percent) yielded statistically significant positive impacts on student academic achievement. The Education Innovation and Research program and i3 offered the department a series of important lessons, including demonstrating how tiered-evidence programs can be used to support the discovery of promising practices, highlighting interventions for which there is a growing body of evidence of efficacy, and showing how to responsibly use federal resources to support scaling of practices with evidence of effectiveness.

Technical assistance

Finally, the department invests in a variety of direct technical-assistance activities to further amplify what it has learned across its portfolio of evidence-building activities. Two notable examples of this work include the Regional Educational Laboratories and the Comprehensive Centers.[8,9] Each year, these organizations work with thousands of education stakeholders across the United States, its territories, and freely associated states to bring evidence to bear as leaders work to solve local problems of policy and practice. The work of the Regional Educational Laboratories program emphasizes applied research and building educators' and policymakers' capacity to understand and make use of data and evidence. In fiscal year 2018, the Regional Educational Laboratories hosted hundreds of public events, trainings, and coaching activities for state and local educators and policymakers focused on the design, evaluation, or implementation of evidence-based practices spanning prekindergarten to postsecondary education. The Comprehensive Centers emphasize building the human and organizational capacity of state education agencies to design and implement policy related to ESSA through training and technical assistance.

LESSONS

As the department has increasingly used its grant portfolio to encourage evidence-building and use, it has identified several lessons that may be relevant to those who seek to do the same.

- *Promotion of collaboration is critical.* It is critical to support staff across the organization and promote collaboration. Evaluators and researchers benefit from learning more about the contexts and constraints that programs face. Similarly, program staff members benefit when they learn more about how evidence can inform program design and get more engaged with the evidence-building process.

The department's work to develop and use evidence through its discretionary grant portfolio depends on experts with diverse perspectives. The type of competition planning we describe above regularly include program staff, policy leadership, program attorneys, and budget analysts, in addition to advance consultation with evaluation experts.

- *Partnerships for technical expertise and capacity can yield benefits.* The Institute of Education Sciences, an independent agency within the department, is a unique resource that provides a critical perspective as the department sets policy, designs programs, and makes meaning of available data and research. Partnerships between the Institute of Education Sciences and the policy and program offices in the Education Department hold significant potential benefits for the field. Evaluators have an important role to play in building evidence through grantmaking. IES staff advises colleagues from across the department on how rigorous, project-level evaluations can be conducted and supported.

- *Learning from grantees' efforts to use and build evidence is ongoing.* Understanding and supporting the most important step—how evidence becomes a part of the approaches used by teachers, school leaders, non-profits, and institutions—is complex. Staff across the department work together to learn from grantees that have committed to using evidence in their programs and to building evidence through their activities. The department values the partnership of the teachers, administrators, and innovators who are asking tough questions, striving for results, and serving students in ways that have the best chance of improving their lives. ■

1   U.S. Census Bureau. *Current Population Survey (CPS). October 1970 through 2017.* Bureau of Labor Statistics. Available at: https://www.bls.gov/cps/.

2   U.S. Department of Education. *Consolidated State Performance Report, 2010-11 through 2015-16.* Available at: https://www2.ed.gov/admins/lead/account/consolidated/index.html.

3   David Kastberg, Jessica Ying Chan, Gordon Murray, and Patrick Gonzales. *Performance of U.S. 15-Year-Old Students in Science, Reading, and Mathematics Literacy in an International Context: First Look at PISA 2015.* U.S. Department of Education, National Center for Education Statistics, December 2016. Available at: https://nces.ed.gov/pubs2017/2017048.pdf.

4   *No Child Left Behind Act of 2001.* Public Law 107-110, 2002.

5   U.S. Department of Education. *Fiscal Year 2017 Annual Performance Report and Fiscal Year 2019 Annual Performance Plan.* February 12, 2018, 68. Available at: https://www2.ed.gov/about/reports/annual/2019plan/fy17apr-fy19app.pdf.

6   U.S. Department of Education. *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments.* September 2016. Available at: https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf.

7   Beth Boulay, et al. *The Investing in Innovation Fund: Summary of 67 Evaluations.* U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, June 2018. Available at: https://ies.ed.gov/ncee/pubs/20184013/pdf/20184013.pdf.

8   U.S. Department of Education. "Programs: Comprehensive Centers Program." Last modified April 11, 2019. Available at: https://www2.ed.gov/programs/newccp/index.html.

9   Institute of Education Sciences. "Regional Educational Laboratory Program." U.S. Department of Education. Available at: https://ies.ed.gov/ncee/edlabs/.

20. ENTERING THE EVIDENCE PROMISED LAND

Making the Evidence Act a Law

Nick Hart is a fellow at the Bipartisan Policy Center and chief executive officer of the Data Coalition. Previously, he served as the director of BPC's Evidence Project. He was also the policy and research director for the U.S. Commission on Evidence-Based Policymaking.
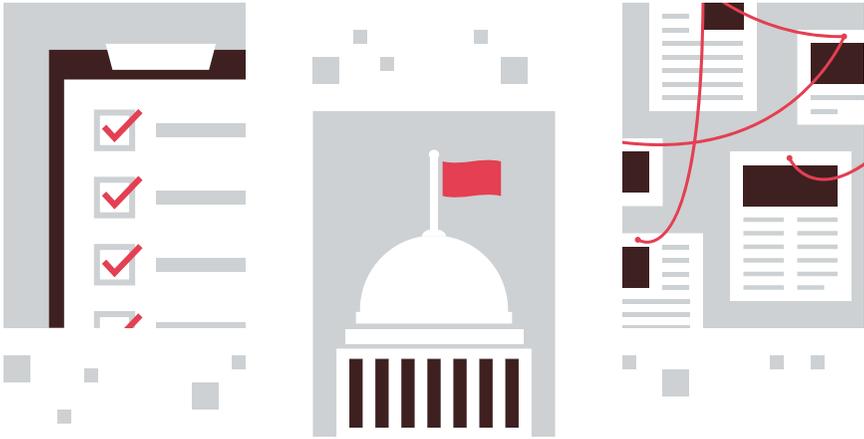
The evidence ecosystem related to government activities involves multiple actors and many individuals. The American public—including individuals and organizations—often willingly provide data to government to support analyses. Researchers, statisticians, and evaluators produce evidence that is relevant and useful for policymakers. Policy analysts and knowledge intermediaries configure information to make it useful and timely for decisions at hand. Public administrators, government decision-makers, appointed officials, and elected leaders all serve in capacities to potentially use evidence.

But there are points at which the ecosystem functions less efficiently—leading government decision-makers to not have timely, relevant, credible evidence available when making decisions. This recognition by two members of Congress—Paul Ryan and Patty Murray —led to the creation of the U.S. Commission on Evidence-Based Policymaking. The unanimous recommendations from the commission addressed themes related to the supply of evidence as well as processes to facilitate meaningful use.

The commission's work itself was an example of how the demand for evidence directly relates to the supply of evidence, leading to use. The commission collected data, analyzed the information, made decisions, then presented the information to Congress and the president in *The Promise of Evidence-Based Policymaking.* The result: a monumental law called the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act).[1]

ISSUE BACKGROUND

Amid a government shutdown in 2012, while Rep. Ryan and Sen. Murray were serving as the lead budget negotiators in the House of Representatives and U.S. Senate, respectively, the duo came to a bipartisan agreement on many issues. One key issue was their interest in knowing

more about how programs operate, what strategies are most effective in implementing government operations, and how to go about assuring the American public that the government is using their taxpayer funds prudently.

Ryan and Murray jointly recognized that government collects a lot of data, largely through the course of implementing activities but also through survey collections and other instruments. However, these data are not always brought to bear in meeting the needs of decision-makers. Ryan and Murray agreed there was a need to develop a strategy to more effectively use government data. Together, over the next several years, they refined the idea to create a bipartisan commission of experts to figure out a better way. The legislation that the two filed in 2014 took several years to work through Congress, but then-President Barack Obama signed it into law in 2016.[2]

The very creation of the Evidence Commission was a clear signal about the growing prominence of the evidence movement in the United States. The law was a declaration from members of Congress about the presence of modern barriers to data access and use that inhibit evidence-building activities. But when the commission was established, the law gave it explicit direction about how to undertake its work as well as a very detailed charge about questions that Congress and the president wanted addressed.

The law established that the 15-member commission would include political appointees from Republicans and Democrats in Congress, as well as the president. It also required that five of these appointees have expertise in privacy issues. The final membership of the commission represented a wide range of backgrounds, including former heads of the Census Bureau and the Bureau of Labor Statistics, the chief statistician of the United States, former program administrators, past members of the White House Council of Economic Advisors, government privacy experts, and former congressional staffers. The commission members came from a range of academic disciplines, including economics, law, statistics, evaluation, psychology, accounting, and computer science.

President Obama tapped Katharine Abraham as chair. She joined with incredible experience as a former commissioner of the Bureau of Labor Statistics and a member of the Council of Economic Advisors. Ryan, then the House Speaker, selected Ron Haskins as co-chair. He had just completed a book on evidence-based policymaking and was a long-time champion of better evaluating government programs. Together Abraham and Haskins hired a support staff that offered additional expertise about government operations, including the disciplines of program evaluation, economics, statistics, survey methodology, privacy, library science, and public administration. The commission's co-chairs laid out a masterful strategy for rapidly addressing their charge, while developing a new body of evidence and applying that evidence to the commission's decision process. The 15-month process from the commission's first meeting to final report culminated in a set of unanimous findings and recommendations, which Congress then had to determine how to act upon.

EVIDENCE AVAILABILITY

When the commission initiated its work, it had limited resources and limited accessible compilations of useful evidence on which to frame decisions. The co-chairs and staff quickly laid out a fact-finding and research strategy to address the 16 overarching questions that Congress and the president had requested answers to.

As a starting point, the staff of the White House's Office of Management and Budget (OMB) coordinated with federal agencies to produce a series of five framing memoranda to set the stage for issues the White House hoped the commission would address. The memos broadly explained OMB's views about the state of evidence issues in government, examples of using administrative records for building evidence, a list of barriers to accessing and using government data, perspectives about privacy and confidentiality issues, and a partial inventory of government datasets.[3]

Even with the 80 pages of detailed memoranda from OMB, the commission members had much to learn. The commissioners and staff embarked on a fact-finding process to ensure evidence was available to inform key decisions. The fact-finding process involved (1) a survey of federal agencies; (2) qualitative information gathered from public hearings, meetings with expert testimony, and solicitation of written public comments; and (3) additional agency, commissioner, and staff research.[4]

Commission survey of federal offices

With the support of a staff survey methodologist and experts from among the commission members, the commission developed, tested, and launched a survey of 209 federal agency units involved in evidence-building activities. They designed the organizational-level survey to gather a range of information about activities underway in government, perceived barriers and limitations to engaging in the work more fully, resource allocations, and capabilities for using evidence in decision-making processes.[5] The survey identified that more than half of responding units noted that legal limits pose substantial barriers to using data for their activities, including nearly all principal statistical agencies and evaluation units that responded. Nearly one-quarter of respondents noted that income and earnings data were especially challenging to access because of limits on data sharing.

The survey also provided insights about whether and how government agencies and offices provide data access to researchers. The survey identified that all principal statistical agencies allow external researchers to access data, though a relatively low share of the rest of responding units do the same. This suggested that agencies with administrative records often did not have formal processes for managing or providing access. The survey also presented a host of other issues, including constraints around resources and capacity to manage data-sharing activities, limits about documentation for existing data, funding allocations for evidence-building activities, challenges hiring a skilled workforce, and negative implications of existing processes for procurement and certain data-collection approvals.

Evidence from expert witnesses and public input

In addition to the survey, the commission hosted seven public meetings with nearly 50 invited expert witnesses on topics including privacy, international approaches to data management, legal standards for data security, and evaluation. The commission also hosted three public hearings with nearly 40 other witnesses and received more than 350 public responses to a request published in the *Federal Register*.

The commission transformed 2,000 pages of information gathered throughout the fact-finding and learning processes into evidence, as commission staff coded and analyzed the qualitative information to determine central themes. This body of evidence led to identifying issues for the commission related to government's capacity to generate and use evidence, including for formal program evaluation, as well as about lessons and models that could generalize broadly to the federal government.

Other research evidence

The commission members also had access to a range of information compiled as it proceeded in fact-finding. For example, Speaker Ryan's office asked the Congressional Research Service to prepare a study of recommendations from prior relevant commissions and to study various statistical laws. Commission staff also studied findings from related committees in the past. Commission staff explored additional data access and privacy issues on the commissioners' behalf, including strategies for encouraging collaboration and co-production of research that might facilitate use. They identified relevant research and considered studies suggested by expert witnesses.

As the fact-gathering phase concluded, the commission members and staff needed to make sense of the vast amount of information collected. The process of sense-making involved hours of dialogues between commissioners and staff. Notably, commissioners also had access to information and evidence from their various disciplines, careers, and experiences to inform, frame, and add context to understanding and using the compiled body of evidence. The process was iterative and fluid, sometimes leading to requests for new research from staff or federal agencies.

EVIDENCE USE

As the process of sense-making unfolded over the summer of 2017, the commission staff developed a series of 12 decision memoranda for commissioners. While staff conducted the initial drafting and compilation of materials, the individual voting members of the commission also provided input on all the memos.

Developing the commission's recommendations

Subgroups of commissioners formed to contribute to the synthesis of materials in the complex, multifaceted process. Each group focused on a different set of issues related to privacy, data access, or government's capacity to use data. These subgroups played an important role in the

commission's ultimate use of the compiled information, as each decision memoranda included suggested recommendations the commission could offer in its final report. The sub-groups of commissioners, applying their perspectives and expertise in conjunction with the collected information and evidence, framed and modified recommendations accordingly to address the questions Congress posed to the commission.

After the sub-group reached an agreement or provided input, it presented and discussed each decision memorandum with all commission members during formal meetings. During these meetings, staff and commissioners both presented information. In several critical instances, the commissioners who had served on sub-groups also served in the role of presenting to other members, relying on their expertise and credibility in offering the recommendations to the group. Each of these presentations and dialogues allowed commissioners to probe, reflect, and even redirect when necessary about interpreting the available evidence as well as translating that information into recommendations.

The decision memoranda process and formal meetings of commissioners produced more than 100 potential recommendations for consideration in a final report. The commissioners collectively decided during one meeting that this would simply be too much material to provide publicly if the goal was to elicit meaningful change. The commission grouped, consolidated, and modified the recommendations to reduce the list to 20 recommendations, which served as the foundation for writing the commission's final report and summarizing key findings.

Through the process of deliberating on the text of the findings, recommendations, and final report chapters, commissioners added additional recommendations. And as the commission members achieved more clarity in message and goals, they further modified other recommendations. A collaborative decision-making process was underway. The co-chairs set an objective to reach unanimity on the report, even though the commission was not required to do so. This goal necessitated last-minute edits and agreement among commissioners to reach complete unanimity. Clearly it was a step in which commissioners negotiated, conceded, and agreed on various points down to every single word in the recommendations—including two new recommendations added in the late stages of the deliberations.

At the end of the process, the commissioners were unanimous in their findings and recommendations. The evidence developed throughout the commission's fact-finding process was front and center in the final report, presented to Congress and President Donald Trump in September 2017. The recommendations from the Evidence Commission were, in fact, based on evidence. They focused on strategies to improve access to data, to strengthen privacy protections, and to enhance government's capacity for evidence-based policymaking.

The commission recommendations leaned in on enabling the production of valid and reliable evidence, promoting a new statistical agency to securely and temporarily link data, changing laws to enable certain types of data uses, and building mechanisms to ensure confidentiality of data subjects when data files or results are made public. It also offered strategies for encouraging the use of evidence by decision-makers. One recommendation included developing a chief evaluation officer position to provide a senior leader to promote production

and use of evaluations, as well as creation of learning agendas in agencies to provide signals to senior leaders and researchers about what knowledge gaps are most critical to address in future decisions. The commission also recommended the establishment of senior leadership positions to focus on data policy, a role that would later be called chief data officers.

But the commission only had the legal authority to make recommendations. It could not change laws or policies, just make suggestions to Congress and the president. Thus, any resulting change to laws or regulations would require action from Congress or the executive branch agencies. This reality led to the creation of the Bipartisan Policy Center's (BPC) Evidence-Based Policymaking Initiative, an effort to continue the discourse about the commission's recommendations even after the commission no longer existed.[7,8]

Creating the Evidence Act

The Evidence Commission's whole focus had been on enabling the supply of useful evidence for decision-makers to use. While most of the recommendations focused on production, there were also recommendations that encouraged the use of evidence.

House Speaker Ryan and Senator Murray participated in an event announcing the commission's unanimous recommendations. They personally applauded the unanimity of the report as well as the reasonable solutions presented to the identified challenges. Practically speaking, the unanimity offered credibility for the recommendations and a means for guarding against political backlash when advancing the recommendations.

The House Committee on Oversight and Government Reform invited four commissioners, including the co-chairs, to testify about the recommendations. Nearly all of the committee members attended the hours-long dialogue about what the commission found and what the findings meant. The hearing had the effect of building trust in the commissioners and the collective product they developed, as well as raising public awareness of the issues addressed by the commission. Members of Congress asked questions and had a dialogue directly about the substance of the report and gauged the expertise and credibility of those who had prepared it. Bringing the evidence from the commission and the commissioners' expertise to the public limelight also offered members of Congress the ability to gauge reactions from constituents, assessing the political viability of the recommendations offered but in a nonthreatening way and ahead of any formal legislation or votes. Both before and after the hearing, commission staff, and later BPC staff, briefed numerous congressional committees and member offices on a bipartisan basis.

Behind the scenes, congressional staff were in frequent contact with commissioners, the commission's staff, and BPC staff to develop a legislative proposal that incorporated some of the recommendations. Ryan and Murray wanted to file legislation quickly, which meant the recommendations included in the legislation needed to be carefully selected based on political goals and consensus-based processes. Through discussions with commission staff, Ryan and Murray staff, and committee staff, a set of recommendations was selected and prioritized for inclusion—recommendations that offered immediate progress on the commission's vision while

also avoiding difficult political choices. Thus, even while the recommendations were unanimous across party lines in the commission, political feasibility was still a factor in determining which recommendations to incorporate in legislation. The timeline also necessitated choosing recommendations perceived as less controversial and easier to do, recognizing the commission had provided a menu of potential recommendations to choose from.

The commission also worked closely with congressional staff to confidentially develop a legislative proposal, including translating the commission recommendations into statutory language. In some cases, this required creativity and in other cases legal consultation because the commission optimized the recommendations for lay accessibility but not legalese.

As a result of this process and collaboration between the congressional staff and the commission staff, and then BPC's evidence team, in October 2017 Ryan and Murray jointly filed the Foundations for Evidence-Based Policymaking in the House and Senate.[9] Their announcement of the proposed legislation directly referenced the evidence compiled by the commission.

Enactment of the Evidence Act

Within weeks of filing the legislation in the House and Senate, the congressional committee voted unanimously to advance the legislation to the full House. A committee report explaining the nuances of the proposed bill and its substance leaned almost exclusively on the evidence from the commission, as well as supporting references from the Government Accountability Office.[10] Shortly after the committee approved the bill, the full House of Representatives approved the legislation without a single opposition vote.

After rapidly passing the House, the legislation moved on a slower track through the Senate. In many ways, the plan to continue the commission's activities at BPC was most critical for this stage as it offered continued expertise on the legislation and ensured key commissioners remained involved in weighing potential changes to the law. One goal for the work at BPC was to ensure the legislation retained fidelity to the commission's recommendations; ensuring ongoing participation of the former commission co-chairs and a former commissioner was an essential aspect of this stage. This was particularly the case because policymakers premised the legislation itself on using the Evidence Commission's report as its key rationale. Thus, alignment was essential for political viability as well as garnering support from the data, science, evaluation, and privacy communities.

Maintaining involvement of the commission perspective was a key part of BPC's evidence team throughout 2018. Expertise and advice on the legislation emerged in various forms. Publicly, a series of editorials in newspapers aimed to keep attention on the importance and bipartisan nature of the proposed legislation.[11,12,13,14] BPC's staff coordinated a statement from former heads of federal statistical agencies to also lend additional credibility to the commission's recommendations and report, particularly with regard to confidentiality protections.[15] BPC staff also offered informal assistance to countless other stakeholders in educating constituencies about the commission's report and the legislation through briefings, statements, and events.[16]

BPC staff and former commissioners also frequently participated in dialogues about the legislation with congressional members and staff. The effects of some of these discussions could also be seen publicly, as some lines of inquiry resulted in the production of public technical papers to explain core concepts and how they applied in certain circumstances. For example, related to executive performance management activities, the relationship between evaluation officers and data officers, modern confidentiality protections related to data sharing, and a detailed understanding of data-sharing barriers.[17,18,19,20] Behind the scenes, BPC's evidence team offered extensive technical assistance related to the sponsors' goal of fidelity to the commission recommendations.

Countless other organizations also played a role in either advocating for the commission recommendations or advancing the legislation. The American Evaluation Association and Results for America encouraged attention on program evaluation (Title 1), the Data Coalition championed the OPEN Government Data Act (Title 2), the American Statistical Association encouraged passage of the Confidential Information Protection and Statistical Efficiency Act (Title 3), just to name a few.

In parallel with Senate consideration of the Evidence Act, the Trump administration announced various activities to advance some of the commission recommendations without waiting for congressional action. The president's reorganization proposal from 2018 announced an intent to have agencies create evaluation officers and develop learning agendas.[21] The President's Management Agenda also announced the creation of a new Federal Data Strategy to incorporate yet still other Evidence Commission recommendations.[22] Both proposals referenced the commission's work and report.

One political issue that arose during the deliberations on the Evidence Act in the Senate was a question of whether the administration would support the legislation as drafted. Recognizing that the administration had already announced some policies consistent with the bill, BPC staff coordinated educational briefings on evaluation and the Federal Data Strategy with congressional staff to maintain alignment to the concepts—including to highlight common uses of evidence related to a reliance on the findings from the Evidence Commission.

In late 2018, as the session of Congress quickly neared a conclusion and with Speaker Ryan preparing to retire from the House, renewed political urgency emerged to prod dialogues about enactment of the Evidence Act. A series of rapid, nonpublic deliberations and negotiations occurred among a handful of key members of Congress and their staffs, resulting in modest revisions to the proposed legislation. But, in December 2018, following dozens of briefings, numerous versions of the legislative text, and lively discussions about scope, framing, and intent, the legislation unanimously passed the Senate and received final passage in the House with overwhelming bipartisan support. In the weeks that followed, President Trump signed the Evidence Act, enacting the bill into law.

In addition to establishing new leadership positions to encourage evaluation activities and the use of data, the law directs agencies to make their data open by default. This means that the expectation is now that to the extent possible agencies create publicly accessible datasets.

Agencies also must document what data they collect and manage, improve privacy protections by better managing risks, and take steps to protect public trust in data and statistics. One particularly valuable provision of the law enables improved access to administrative, operational data for generating statistics in privacy-protective ways. The law also directs agencies to establish many of the core features of basic program evaluation capacity, including written policies and a supporting workforce for conducting evaluations of programs and policies.

As the work proceeds to implement the Evidence Act, it will persist as not only a set of changes to federal law that promote evidence-based policymaking but also as a model process for how evidence can be formulated and used in decision-making. In the months following enactment, the Trump administration continued to promote effective implementation of the new law in its budget proposal to Congress.[23] While the law exists, and the administration actions suggest support, the real test of its effectiveness will come in the years ahead—and that too should be subject to evaluation.

LESSONS

- *Decision-making processes were nonlinear.* While public policy and political science training often teaches students about a linear model for public policy and decision-making, the commission's decision framework to develop its findings and recommendations followed a different trajectory. The commission's decision-making process was iterative, multifaceted, dynamic, and nonlinear. But commission members were nonetheless committed to the process and motivated by a recognized need to have reliable evidence on which to base their decisions. Similarly, the decision-making process for advancing the legislation did not follow the classic model.

- *Unanimous recommendations set the stage for improved credibility.* The process undertaken by the commission in fact-finding and developing its recommendations maximized political credibility for future action by striving for unanimity, even though the commission was not required to do so. Those recommendations, paired with the technical credentials of the members of the commission, provided policymakers firm ground to stand on when looking for a defensible basis for proposing and supporting subsequent legislation, the Evidence Act. Without unanimity, while unclear what would have happened, it is likely political support for the law would have been more challenging to achieve on both sides of the aisle.

- ***Evidence only provided part of the answer for policymakers.*** In developing the legislation, framed by the Evidence Commission's report, lawmakers and their staffs still had to determine how to transfer the recommendations into statutory language. This required insights into how the recommendations could translate and typically required the support of intermediaries to assess the fidelity to the commission's intent. Commissioners, commission staff, and BPC staff served as intermediaries to help convey intent, meaning, and purpose to congressional staff.

- ***Motivated leadership was key.*** The presence of motivated and engaged leaders allowed the commission to come to fruition, led to the unanimous recommendations, and culminated in how the final legislation emerged as enacted law. Paul Ryan and Patty Murray provided senior political leadership throughout the process and stayed engaged as the commission undertook its work, as did their staffs. The commission co-chairs, Katharine Abraham and Ron Haskins, offered technical leadership to produce a report useful to policymakers, but they also ensured the commission's process was itself a model for evidence-based policymaking activities. ◼

1   *Foundations for Evidence-Based Policymaking Act of 2018*. Public Law 115-435, 2019. Available at: https://www.congress.gov/bill/115th-congress/house-bill/4174.

2   *Evidence-Based Policymaking Commission Act of 2016*. Public Law 114-140, 2016.

3   White House Office of Management and Budget. "Commission on Evidence-Based Policymaking." Available at: https://obamawhitehouse.archives.gov/omb/management/commission_evidence.

4   Appendix C of the commission's final report describes the process; supplemental online materials are also available in Appendices E-H. See: U.S. Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking*. U.S. Commission on Evidence-Based Policymaking, 2017.

5   U.S. Commission on Evidence-Based Policymaking. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking Appendices E-H*. U.S. Commission on Evidence-Based Policymaking, 2017, Appendix E.

6   Ibid, Appendix H.

7   Bipartisan Policy Center. "Evidence Project." Available at: www.bipartisanpolicy.org/evidence.

8   Bipartisan Policy Center. "Ryan-Murray Evidence-Based Commission Efforts Moving to BPC." September 7, 2017. Available at: https://bipartisanpolicy.org/press-release/ryan-murray-evidence-based-commission-efforts-moving-to-bpc/.

9   U.S. Congress. House. *Foundations for Evidence-Based Policymaking Act of 2018*. HR 4174. 115th Cong., 2nd Sess. Became Public Law No. 115-435, January 14th, 2019. Available at: https://www.congress.gov/bill/115th-congress/house-bill/4174.

10 U.S. Congress. House. Committee on Oversight and Government Reform. *Foundations for Evidence-Based Policymaking Act of 2017: Report (to Accompany H.R. 4174)*. 115th Cong., 1st Sess, 2017. H. Rep. 115-411. Available at: https://www.congress.gov/congressional-report/115th-congress/house-report/411/1?overview=closed.

11 Nick Hart and Robert Shea. "Opinion: Is there Room for Science and Evidence in Trump's Budget?" *Roll Call*, February 5, 2018. Available at: https://www.rollcall.com/news/opinion/opinion-room-science-evidence-trumps-budget.

12   Andrew Feldman and Nick Hart. "OMB Wants to Strengthen a Learning Culture in Government." *Government Executive*, May 21, 2018. Available at: https://www.govexec.com/excellence/promising-practices/2018/05/omb-wants-strengthen-learning-culture-government/148350/?oref=river.

13   Katharine G. Abraham and Ron Haskins. "Time for Action to Improve Government Data Analysis." *The Hill*, September 4, 2018. Available at: https://thehill.com/opinion/finance/404928-time-for-action-to-improve-government-data-analysis.

14   Ted McCann and Nick Hart. "'Evidence' is the Word of the Year. Neither Party Owns It." *Roll Call*, October 1, 2018. Available at: https://www.rollcall.com/news/opinion/evidence-word-year-neither-party-owns.

15   Bipartisan Policy Center. "Former Government Statistical Agency Heads Call for Better Use of Data in Policymaking." January 24, 2018. Available at: https://bipartisanpolicy.org/press-release/former-government-statistical-agency-heads-call-for-better-use-of-data-in-policymaking/.

16   Bipartisan Policy Center. *Evidence: A Year of Progress on Evidence-Based Policymaking.* Evidence-Based Policymaking Initiative, September 2018. Available at: https://bipartisanpolicy.org/wp-content/uploads/2018/09/A-Year-of-Progress-on-Evidence-Based-Policymaking.pdf.

17   Nick Hart and Kathryn Newcomer. "Presidential Evidence Initiatives: Lessons from the Bush and Obama Administrations' Efforts to Improve Government Performance." Bipartisan Policy Center, February 2018. Available at: https://bipartisanpolicy.org/wp-content/uploads/2018/03/Presidential-Evidence-Initiatives.pdf.

18   Nick Hart. "Improved Data Coordination and Use Required for Evidence-Based Policymaking." Bipartisan Policy Center, November 9, 2017. Available at: https://bipartisanpolicy.org/blog/improved-data-coordination-and-use-required-for-evidence-based-policymaking/.

19   Nick Hart and Katherine Wallman. *Transparency, Accountability and Consent in Evidence Building: How Government Ethically and Legally Uses Administrative Data for Statistical Activities.* Bipartisan Policy Center, July 2018. Available at: https://bipartisanpolicy.org/wp-content/uploads/2018/07/Transparency_accountablility_and_consent_in_evidence_building.pdf.

20   Nick Hart and Kody Carmody. *Barriers to Using Government Data: Extended Analysis of the Commission on Evidence-Based Policymaking's Survey of Federal Agencies and Offices.* Bipartisan Policy Center, October 2018. Available at: https://bipartisanpolicy.org/library/barriers-to-using-government-data-extended-analysis-of-the-u-s-commission-on-evidence-based-policymakings-survey-of-federal-agencies-and-offices/.

21   Nick Hart. "President's Government Reform Plan Includes Evidence Proposals." Bipartisan Policy Center, June 21, 2018. Available at: https://bipartisanpolicy.org/press-release/presidents-government-reform-plan-includes-evidence-proposals/.

22   Nick Hart. "Trump's Management Agenda Proposes Data as a Strategic Asset." Bipartisan Policy Center, March 26, 2018. Available at: https://bipartisanpolicy.org/blog/fact-sheet-trumps-management-agenda-proposes-data-as-a-strategic-asset/.

23   Nick Hart. "Yes, Trump's Budget Really Does Promote Evidence-based Policies." *Roll Call*, March 25, 2019. Available at: https://www.rollcall.com/news/opinion/yes-trumps-budget-really-promote-evidence-based-policies.

CONCLUSION
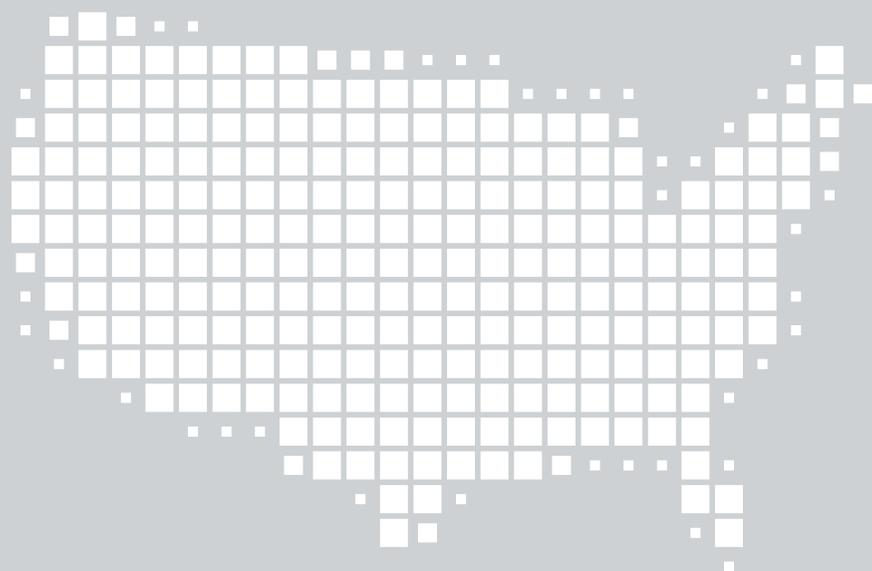
USING EVIDENCE IN THE REAL WORLD

What We Learned

Nick Hart is a fellow at the Bipartisan Policy Center and chief executive officer of the Data Coalition. Previously, he served as the director of BPC's Evidence Project.

Meron Yohannes is a research analyst at the Bipartisan Policy Center.

We set out to develop a series of cases to explore how evidence informed policymaking and to explain how those uses occurred. *Evidence Works* includes 20 success stories that each offer a unique narrative. Taken together, the cases reveal a great deal about how society and decision-makers go about using evidence in real-world policy settings.

As alluded to in the introduction, the case studies suggest use is complex, multifaceted, iterative, and constantly evolving. There is no single approach or solution that universally encourages or facilitates use for all types of issues or questions. In some situations, uses were confounding, if not bewildering.

Perhaps the most common realization in producing *Evidence Works* is the frequency and tendency to focus on describing the available research evidence, followed by explaining what the policy action was without explicating how use occurred. This sometimes happened when describing the organization as the user, not the individual actors. In other cases, the explanation of how use occurred lacked the context or integration with other factors that affected a decision.

It turns out, the heart of evidence use was not only the hardest component to explain in the case studies but also a characteristic of the narrative that many overlook when retelling stories about policymaking. There are likely many reasons for this, including (1) a lack of general awareness about what use is and how to do it, even in the policymaking community; (2) the variety of perspectives about the different types of uses; and (3) the various perspectives about what constitutes evidence available for use. In developing the case studies, these reasons periodically emerged, suggesting a need for continued dialogue and education about different aspects of how evidence-based policymaking occurs.

Describing the challenges of evidence use (and especially how use manifests) is not a critique of those who generously offered to share their perspectives for this volume. However, it is a reality. In order to encourage others to tell their stories and better build theories or research about use, researchers and policymakers must overcome this critical challenge.[1] Indeed, the authors who contributed to this volume overcame it, in some cases with extended dialogue, consultation, and reflection.

INSIGHTS ABOUT EVIDENCE USE

Beyond the overarching challenge of talking about evidence use, we also learned an incredible amount in the course of compiling, developing, and delving into the cases. Here are the key lessons:

1. Leadership comes in many forms; its presence is essential.

Motivated leadership was instrumental in facilitating use in every single case. But how that leadership presented itself was deeply varied, as were the incentives and knowledge about research approaches for key leaders. Sometimes leadership was both obvious and intuitive,

such as for the administrative and regulatory actions described in Cases 5 through 12. These examples of actions undertaken in the fields of child support, early childhood development, energy assistance, and other areas likely offered clearer indications of leadership due to a hierarchical infrastructure in the executive branch with apparent choice points and decision architecture. In the legislative context, however, leadership was complicated by the number of decision-makers and the brokering process for reaching decisions. Cases 1 through 4 and Case 20 offer multiple instances where congressional leaders emerged as the champions or sponsors of legislation, though the uses of evidence often fell to support staff and others to provide advice or counsel to voting members of Congress.

2. Evidence use happens at both an organizational and individual level; both affect culture.

There are multiple dimensions to policymaking and, thus, multiple levels at which evidence use happens. One dimension is the individual actors. Behind every decision is a person or group of people who must interpret, construe, and judge the information and evidence presented on given issues. These decisions also reflect organizational interests and may, as is the case of Congress, reflect a more collaborative decision-making process that results in an organizational action.

Acknowledging the role of both levels for decision-making offers important implications for ensuring the capacity to facilitate effective uses. For example, experts can train and educate individuals about how to be effective users with knowledge about judging validity, credibility, and reliability of evidence as time and resources allow. Organizations, in contrast, can embody processes and a larger culture that can induce or promote effective use.

3. Partnerships, co-production, and collaboration present clear benefits.

Countless cases expressly involved evidence produced and used through a strategic partnership. Food safety improvements in King County, Seattle, occurred after a government agency brought in researchers to work alongside staff to develop more insights about a policy. The District of Columbia used a research lab to facilitate a stakeholder-informed process to develop insights about key aspects of the body camera program in the police force. In the disability insurance reforms, agency staff worked with congressional and White House staff to provide requested information. These cases, and others, highlight that when researchers generate evidence alongside the potential users, the evidence can be calibrated to address the questions that will be most useful to answer.

Other partnerships emerged in the form of interagency collaboration. The work between the Housing Department and the National Center for Health Statistics offers insights about the benefits of combining data from different sources, recognizing the potential power of consolidated insights and analytical capabilities.

These partnerships can also offer another key benefit, fostering consensus and agreement about what the evidence means. The Commission on Evidence-Based Policymaking's unanimous decisions were instrumental in securing passage and bipartisan support for the eventual legislative action. The broad support for the Earned Income Tax Credit and a host of proposals for potential expansions reflect a consensus about the effects of the policy in reducing poverty in the country. The collaborative process with stakeholders and state governments to develop insights about environmental contamination from old gas stations contributed to relatively rapid change in longstanding environmental policies. The list goes on through many more cases, suggesting that partnerships offer a potential path for productively creating useful evidence that is then used.

4. A supply of evidence is necessary, but not sufficient to facilitate use.

While using evidence in policymaking is the theme of this volume, it is also apparent that policymakers cannot use evidence that does not exist. Multiple cases identified the need for more and better evidence. But even in the cases where robust evidence did exist, the presence alone was not enough to facilitate use. In practice, mere availability is insufficient for use to occur, which might raise questions about the practical implications of "dissemination" of evaluation reports and policy research on government clearinghouses or websites. Cases 17 through 19 highlight this point well: the teen pregnancy prevention, education, and home visiting initiatives all included components of providing assistance and sharing knowledge about relevant and useful policies and practices. Similarly, the story about the Foundations for Evidence-Based Policymaking Act in Case 20 explains the role of intermediaries and knowledge brokers in relaying and translating key concepts into policy actions through briefings and technical advice. In other words, the activities to enable use did not cease when a report providing evidence was published.

In addition, multiple cases suggest that uses proliferate when a range of evidence is available to answer varied questions—or when new insights develop rapidly to address evolving questions from policymakers. In the disability insurance reforms, a federal agency was on tap to rapidly provide new analyses to answer evolving questions. In the reforms about energy assistance, policymakers applied a broad range of descriptive statistics, trend analysis, and survey analyses to other policy-relevant information in order to develop program reforms. For the workforce reforms, cost information about interventions and implementation knowledge contributed to modifying workforce training. Rarely did the evidence come from a single study. This suggests that users must pay attention to resources and apply a range of information types and analytical methods to develop useful insights that truly respond to the full spectrum of informational needs in policymaking.

5. Ease of production can address perceived barriers to use.

There is no doubt based on the case studies that the ease of evidence production does matter for whether and how evidence is used. The cases highlight areas where policymakers could, in fact, successfully encourage generating evidence using a range of methods and

approaches. Some cases used confidential administrative records, like for Social Security reforms, child welfare, and health in public housing analyses. Other cases relied on publicly available open data, such as the energy assistance reforms, or non-sensitive yet restricted data, such as about the nutrition cost efficiencies.

In all of these cases, data were readily accessible and available for use by analysts and presented to decision-makers. If policymakers or analysts had perceived data and ensuing analyses could not be available within the time frame and context needed, it is likely they would not have waited or asked about the evidence.

This insight suggests that attention to reducing barriers to data collection and data access can offer meaningful benefits to the availability of evidence and to perceptions about barriers to evidence use. It also suggests that developing effective mechanisms for policymakers and data experts to have a dialogue about potential questions, needs, and goals could result in ensuring the evidence needed is available. This is particularly evident when data issues underlie the delay in producing evidence.

6. Politics is an inevitable and necessary attribute of use.

While some may perceive the use of evidence to be a technocratic or apolitical activity, in reality, it is inherently political. In the cases here, politics was often a necessary characteristic for use to occur. Politics appeared most clearly around congressional decision-making, though it also related to the Earned Income Tax Credit, workforce training reforms, child support regulation, energy assistance programs, and even to how policymakers in Washington, D.C., interpreted and used information from the body camera study.

Using evidence in these political circumstances required the actors to weigh multiple factors about use, including political goals, potential critiques, and funding availability. The disability insurance reforms highlight the role of political values and philosophies in framing how policymakers reach decisions. One strategy, as in the body camera case, was to plan ahead for critiques, including politically charged critiques, which researchers can mitigate by promoting transparency and involving stakeholders early in the process of building evidence.

Alleged evidence "misuse" is also a concept that emerged in several cases, typically described as a result of political motivations. But misuse in one person's views may be a legitimate use for another. Some experts characterized the teen pregnancy prevention funding decisions as a misuse of evidence because an executive decision allegedly had a political philosophy that did not value the program or reach the same interpretation of the body of evidence. These types of claims demonstrate the reality of how evidence use invariably depends on values, training, and interpretation of the information available. They also suggest a political use aimed at either defending or undermining program activities, based on other values and inputs that decision-makers may normatively apply on top of evidence.

7. Success leads to more success.

Numerous cases showed that successful evidence use led to other engagements and uses for individuals participating in the program. The food safety case identified county officials who wanted further evaluations to learn more. The same occurred in Tennessee's education programs and the Community Response Program in Colorado, where initial evaluations provided further insights. This may, in part, be a realization from decision-makers that when the information presented does not conform to initial expectations, they may have still more questions to address. It may also be a reflection of policymakers' practical goals to reduce uncertainty for decisions, even planning for future reductions of uncertainty at the point in time when the first decision is made based on emergent realizations that available evidence cannot be perfect or address all decision needs.

The cases do clearly suggest that decisions will likely move forward, even in the face of uncertainty or if ideal evidence is not available. Given this, as policymakers experience using more evidence in the decision framework, they can learn and adapt their decision approaches, promote organizational changes for continuous improvement, encourage staff infrastructure for evidence-based policymaking, and ultimately usher in the cultural changes that accommodate improved evidence use.

One natural suggestion is to design programs with evidence use in mind, as was the case for teen pregnancy prevention, home visiting, education programs, and Tennessee's literacy program, among others. These program designs motivate culture change within organizations and among stakeholders or grantees to plan for more effective evidence use.

MOVING FORWARD

Taking these lessons together, it's clear that using evidence is a challenging endeavor. But it is also an endeavor that is within reach. At the federal level, reforms enacted in early 2019 suggest some policies and practices in government may be successful in inciting more widespread and effective evidence use in the future. Notably, none of the case studies described in this volume benefited from the law, because each offers a policy narrative before enactment.

The Foundations for Evidence-Based Policymaking Act requires agencies to establish certain leadership positions, specifically to encourage data use and conduct program evaluation. If both the chief data officers and evaluation officers are successful in leveraging the full potential of the law, leaders in federal agencies could have natural champions for prompting organizational culture change with the goal of making more useful evidence available for use. These leaders will have to develop evaluation policies, make data more accessible, and encourage improved data analytics about government policies.

One promising approach that builds on the lessons learned and that is included in the Evidence Act is the development of multiyear learning agendas. By policymakers and stakeholders working collaboratively to identify informational needs and gaps, co-production

of evidence is possible. If successful, the learning agendas could promote improved perceptions of credibility, relevance, and, ultimately, the usefulness of research evidence in federal agencies. Only after implementation of the law will policymakers be able to assess whether this happened as intended, a fitting follow-up about a law on evidence-based policymaking. In order to effectively evaluate the approach, it's important to monitor implementation of these approaches to learn and adapt in the future.

The cases also offer valuable insights about targeted strategies for promoting and encouraging the use of evidence. Policymakers should identify and motivate leaders to both produce and use evidence. This may require the creation of incentives, such as recognizing leaders affirmatively who achieve laudable uses or seeking commitments when political appointees assume their posts about their support for using research evidence. Agencies in government and non-governmental actors can also pursue partnerships that facilitate knowledge sharing and the co-production of relevant and timely insights. Finally, members of the evidence community must not let gaps or limits in existing capacity and infrastructure discourage them. Changing cultures takes time and repeated efforts, but with continued interest and enthusiasm leaders can continue to build policies that use evidence in the real world.

While we learned a great deal developing *Evidence Works*, the included cases are not intended to offer the perfect strategy or approaches that will apply across-the-board. But the cases do highlight ways that evidence is useful to policymaking. Hopefully, the cases can be applied, discussed, and expanded in the future.

We encourage decision-makers to share their stories of meaningful evidence use in policymaking. The field needs more cases to extend existing theories and to better understand how to facilitate and encourage use. Please help us—and the evidence-based policymaking movement—by sharing your successes so we can continue to learn and continuously improve.

When evidence informs actions, it can produce improved decisions, reasoned policies, and fulfilled goals. When it comes to a common-sense strategy for making government more effective and creating a better society, evidence works.

1   D.H. Gitomer and K. Crouse. *Studying the Use of Research Evidence: A Review of Methods*. William T. Grant Foundation, 2019.
     Available at: http://wtgrantfoundation.org/studying-the-use-of-research-evidence-a-review-of-methods.
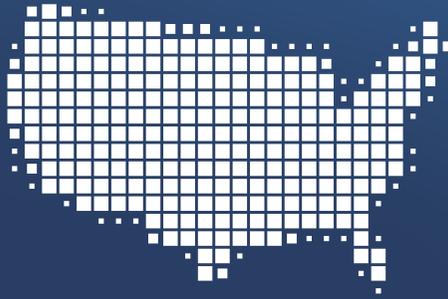
All across society people use information to make good decisions. Policymakers and elected officials must be held to the same standard and engage in evidence-based policymaking. Today, there are a great many success stories about how evidence informs policy. Evidence Works presents a collection of case studies that highlight the many approaches to using evidence, the different types of information that can be relevant, and the challenges faced in the real world. But one thing is clear—society benefits when policymakers use research evidence to make good decisions.