Using Data to More Rapidly Address Difficult U.S. Social Problems

By
JEFFREY B. LIEBMAN

This article argues that the evidence-based-policy movement needs to supplement its current emphasis on program evaluations with an approach that uses data at a much higher frequency to improve the administration and impact of government-funded social service programs. Doing so offers the best chance of making significant progress in ameliorating challenging social problems. I describe how an idealized government social service agency could use data and data analysis to improve its results, review the barriers that prevent agencies from operating in this way, and outline how targeted resources and technical assistance can help to overcome these barriers. Finally, I discuss strategies for moving beyond the effective administration of siloed service programs to the improvement of population-wide outcomes, especially among individuals and families who need multiple services.

Keywords: social innovation; evidence-based policy-making; performance management; human services; what works; active contract management

Suppose our nation's goal is to demonstrate significant progress within 5 years in ameliorating a large number of social problems in individual communities and, within 10 years, to have spread the successful practices nationwide. Based on the experience of the Harvard Kennedy School Government Performance Lab (GPL), which to

Jeffrey B. Liebman is the Malcolm Wiener Professor of Public Policy at the Harvard Kennedy School, where he directs the Taubman Center for State and Local Government. He founded the Harvard Kennedy School Government Performance Lab, which has provided pro bono technical assistance to more than fifty state and local governments to improve their social programs and contracting.

Correspondence: jeffrey_liebman@harvard.edu

date has worked with fifty jurisdictions spanning twenty-four states, two strategies offer the best chance of achieving this goal:

- First, we need to help the state and local government agencies administering social programs use data and data analysis more effectively as a management tool to generate innovation, systems reengineering, and continuous improvement.
- Second, we need to launch, structure, and fund purposeful data-driven, community-level efforts to tackle difficult social problems in a way that breaks down funding silos and introduces accountability for population-wide outcomes.

In this article, I elaborate on these strategies in three sections. In the first section, I argue that the evidence-based-policy movement needs to supplement its current budget-oriented approach—which focuses primarily on evaluating which interventions work and encouraging governments to allocate budget resources to "proven" interventions—with an approach that uses data at a much higher frequency to improve the administration and impact of government-funded social service programs. In the second section, I present a concrete example of how an idealized government social service agency (specifically, a child welfare agency) could use data to improve the results it achieves for the population it serves. I also describe the barriers that prevent agencies from operating in this way and outline how targeted resources and technical assistance can help to overcome these barriers. In the third section, I discuss the need to move beyond effective administration of siloed social service programs to focus on achieving improved population-wide outcomes, especially among individuals and families who need multiple services. I sketch an approach that philanthropic funders could take to encourage communities to experiment with solutions to this challenge, so that we can develop models that can be adopted nationwide.

The Need for Purposeful Attempts to Achieve Better Outcomes for Target Populations

Much of the rhetoric around the use of evidence in policymaking suggests that government-funded social programs can be divided into two categories: those that work and those that do not. Under this perspective, the main point of increasing access to government data is to perform more impact evaluations so that we know which interventions to expand and which ones to defund.

The infrastructure that has been built up around the "what works" framework—the Coalition for Evidence-Based Policy Top Tier Evidence initiative, the U.S. Department of Education What Works Clearinghouse, the Washington State Institute for Public Policy Benefit-Cost Results, and the Poverty Action Lab's evaluation database, among many other examples—has been quite successful in

spreading information on effective interventions. In our GPL work, we have found that in every state and local government social service agency with which we have worked, there are multiple officials who understand that some interventions in their field are "evidence based" and others are not. Moreover, although I am not aware of any comprehensive time series on the number of rigorous impact evaluations of U.S. social policy interventions completed per year, it certainly appears that the pace at which evaluation evidence is being developed is increasing and that this increase is resulting from a combination of demand-side factors (e.g., governments allocating resources based on tiered evidence standards, the philanthropic community making funding available for randomized control trials) and supply side factors (e.g., reduced costs of working with administrative data, the development of causal impact statistical frameworks that have increased researcher interest in randomization-based research strategies).

Despite all this momentum, we are still not making rapid enough progress on challenging social problems. Rates of disconnected youth, obesity, and prisoner recidivism remain high. There are still more than half a million homeless in the United States, and 30 percent of fourth graders score "below basic" in reading on the National Assessment of Educational Progress. Part of the problem is that we need a lot more innovation, experimentation, and evidence—at least twenty times what we are currently producing.[1] Most evaluations of social programs find disappointing results,[2] and a large portion of programs that look successful in an initial evaluation fail in replication. Therefore, we need to innovate and test at a much more rapid pace. Another part of the problem is that even when successful interventions are discovered, governments do not fund them at scale. Yet another part of the problem is that evidence becomes stale very quickly.[3] For example, randomized experiments in the 1970s and 1980s found that home-visiting services for low-income first-time mothers provided by the Nurse-Family Partnership (NFP) increased the spacing between first births and subsequent births. But birth control technology and Medicaid coverage for birth control has changed substantially since the original experiments were done, making them of little use in predicting the impact of those services today.[4] Finally, even the best models can fail when delivered on a large scale if staff quality and other implementation details are not sustained at the level of the original experiment.

But there is a broader issue as well. *Impact evaluations, while extremely valuable, are a relatively small portion of the hard work that needs to be done with data and data analysis if we are going to move the dial on difficult social problems.* Human service agencies need to be making greater use of data and analysis throughout their operations. Moreover, the rhetoric about using evidence to find out "what works" orients policy-makers incorrectly toward thinking that program effectiveness is a static concept and that the budget process is the primary way to achieve greater effectiveness. Instead, political leaders should be reviewing data on whether programs are doing better this month than last month (or this year than last year) and holding agencies accountable for reengineering their processes and those of their contractors to produce continually rising performance trends over time.[5]

Consider the problem of improving outcomes for a vulnerable population. Let us assume, for example, that we are leading a state health department in a jurisdiction that has one of the highest rates of infant mortality in the nation, and we have decided to focus on the objective of reducing infant mortality and the number of low-birthweight births statewide. This happens to be a policy area where there is a significant amount of evaluation evidence, as well as intervention models that have been certified as "evidence based."

There are at least four types of data analysis that are needed to tackle this problem:

First, the state needs to analyze infant deaths and low-birthweight births to *identify the entire target population* and formulate hypotheses about policy interventions that could potentially affect the problem. Where are the geographic hot spots? What fraction of the mothers are teens? What portion of infant deaths follow upon low-birthweight or premature births, and what portion result from child maltreatment? Are the mothers smokers, drug or alcohol abusers, overweight, or undernourished? What, if any, preventive services were provided to these mothers? Some of this analysis could be done with data from birth records, another portion from Medicaid claims data, another portion from health department and child welfare agency case files, and some might require pulling the medical charts of a random sample of births with bad outcomes.

Second, the state needs to use data on risk levels and intervention cost-effectiveness for specific subpopulations to *refer the right people to the right services*. Which low-income pregnant mothers should be referred to intensive evidence-based home-visiting providers, which to lower-intensity home visiting, and which to no home visiting at all? Which communities need to be targeted for public health campaigns around nutrition, exercise, and healthy infant sleep practices? Where do more resources need to be invested in drug treatment programs and teen pregnancy prevention? How can target mothers be identified early enough in pregnancy to impact birth outcomes?

Third, the state needs to *track service receipt in real time and then collaborate with service providers to minimize the portion of the target population that falls through the cracks*. Each month the state program lead should review data on what percentage of pregnant mothers referred to home visiting received services, and state program staff should meet with providers to review case files to identify reasons that target individuals failed to receive services and make changes in processes to improve the fraction of the target population that is reached. The state should also review cases with bad outcomes that were not referred to services and analyze what can be done through better outreach and targeting of services to reach the highest-risk population. And it is not just the initial receipt of services that should be tracked; progress toward program completion should be monitored as well.

At the GPL, we refer to this combination of high frequency review of data and regular collaborative meetings between government agency staff and service providers to identify opportunities for systems reengineering as *active contract management*. We contrast it with the more typical relationship between contract officers and providers that focuses on invoice processing and compliance reviews.

In the projects with which we have been involved, we have seen active contract management increase the fraction of recently released prisoners who make it to job training in New York State, improve the targeting of permanent assisted housing slots to homeless individuals in Massachusetts, reduce the lag between when a child welfare agency refers a family to emergency services and when those services begin in Rhode Island, and improve the coordination of services for the homeless in Seattle.[6]

Fourth, the state needs to *annually compare outcomes for individuals referred to different services to make decisions about how to allocate resources and adjust referral protocols going forward*. If multiple service providers are serving the same population, their results should be compared and the state should either reallocate slots to the most effective providers or convene meetings at which the higher achievers can share best practices with the others. These sorts of comparisons are not always straightforward. Results need to be adjusted to account for differences in the populations being served by different providers. Otherwise, providers who target the most difficult cases will be penalized. And short of randomization, there is no way to adjust for differences that are not measured in available data. But quite often there are opportunities to use regression-discontinuity strategies to compare outcomes for people just above and below thresholds for referral to services, and there are opportunities to replace idiosyncratic referral processes with deliberate ones that involve randomization to facilitate comparisons of relative effectiveness. Moreover, even when only unadjusted outcomes by service type can be calculated, they can be quite revealing. For example, if we observe that only 10 percent of TANF recipients referred to job training are both employed and earning more than $10,000 per year three years after training is completed, we would know that the state needs to rethink its strategy for helping this population achieve economic self-sufficiency. If we never identify self-sufficiency as the goal and never measure medium-term earnings outcomes, the state might plod along, funding slots with the same service providers, without any realization that the strategy needs to be rethought.

The overall point is that if we want to achieve better outcomes for vulnerable populations, we need to make a purposeful effort to do so. Defining the population that we are trying to reach, measuring the outcomes that we are trying to improve, and using data and analysis throughout the policymaking and service delivery chain drives the systems reengineering and continuous improvement efforts necessary to achieve better outcomes. Evaluating the impact of particular intervention models is only one part of what needs to be done.

Overcoming Obstacles to Effective Use of Data by Government Human Service Agencies

Every one of the three dozen state or local government social service agencies the GPL has worked with produces a large volume of performance data, much of them in quarterly or monthly reports mandated by the legislature or the federal

government. Most of the agencies have internal performance dashboards that they review in regular meetings with senior leadership. But many of the performance metrics are highly imperfect (for example, a state labor department that focuses on the percentage of people completing a job training program who are employed immediately after completing the program, ignoring those who failed to complete the program and also failing to measure the longer-term impact on employment and earnings). And agencies find it very challenging to go beyond performance reporting to use data to drive performance improvement. This is not to say it does not happen. We have observed remarkable examples of agencies improving processes in a relatively short time when leadership and technical capacity have been aligned, but these have mostly been isolated examples of a single agency improving a single process rather than agencies that use data effectively throughout their operations to constantly improve results.

Let me illustrate some of the kinds of analyses that agencies find difficult to do. To make the examples concrete, I discuss these issues in the context of a child welfare agency; most could be illustrated in any other human services agency as well.

Improving outcomes for the entire target population, rather than focusing solely on those who arrive at the agency's front door

In annual reports of child welfare agencies, the first table often shows trends over time in reports of child maltreatment. If this trend is downward, it is interpreted as progress. But reports of maltreatment can decline either because maltreatment declines or because reporting rates decline. While all states now have processes to review child deaths, in some states these reviews continue to focus primarily on cases that occur in families that were previously known to the state human services agencies. Agencies should regularly measure all preventable child deaths in their jurisdiction as well as child injuries. If the ratio of maltreatment reports to deaths and injuries is declining, this is a sign that reporting needs to be improved. Case reviews of preventable deaths and serious injuries in families that were not previously known to the state human service agencies should ask why these families were not previously identified as needing services and what services might have prevented the incidents. Analysis should also identify geographic areas and demographic groups that are most likely to be missed by the system and target outreach to those areas and groups. The reason that these sorts of analyses are rare today is that agency program leads are in charge of managing a siloed spending program that delivers services to people identified by that program, but typically no one is responsible for whether the system as a whole is achieving the desired outcomes for the entire target population.

Evaluating the quality of decision-making on whether to do an investigation, whether to open a case, whether to refer a family to particular services, and whether to close a case

Some state child welfare agencies have very clear "structured decision-making" protocols linked to validated risk assessments that guide staff in

determining when to do an investigation and when to remove a child for place-
ment into foster care. Other jurisdictions give staff considerable latitude in mak-
ing ad hoc judgments based on experience and expertise. But either way, state
agencies should regularly evaluate whether the lines are being drawn at the right
place by comparing subsequent safety outcomes (new reports of maltreatment,
child deaths, child injuries) for cases just below the thresholds for doing an inves-
tigation or opening a case to those just above the thresholds. If subsequent safety
outcomes are poor for those just below the thresholds, then the threshold should
be lowered. If subsequent safety outcomes are very good for those above the
threshold, it may be worth experimenting with raising the threshold to see if
those families do similarly well even with a more limited set of in-home services
but without a formal case being opened. To facilitate such analysis, it may be
necessary to collect data on the subjective assessments of risk levels by agency
staff making the decisions so that those cases that are near the threshold can be
identified.

Focusing on outcomes rather than volume of service metrics

While all child welfare agencies that GPL has worked with have strategic plans
that focus on the goals of safety, permanency, and child well-being, they take a
wide variety of approaches when designing the data dashboards that they review
in their regular senior staff meetings. In particular, it is quite common for these
meetings to focus on volume of service metrics, rather than measures of safety,
permanency, and child well-being. Common metrics are the size of the caseload,
the number of case closings, and the number of children referred to different
types of services. Agencies often focus on volume metrics because these metrics
determine the rate at which the agency exhausts its annual budget allocation, and
how large the caseloads are for social workers. While monitoring volume metrics
is important and can lead to decisions that affect the quality of services delivered,
when leadership attention is focused exclusively on volume metrics, opportuni-
ties are missed to improve service delivery and outcomes.

Interpreting implications of data in a way that can drive operational improvements

We regularly observe meetings at which agency leadership is looking at the
right data, but not asking the right questions about them and at which the staff
are presenting data, but do not offer any useful interpretation of what the data
mean.[7] For example, a histogram might be displayed of the distribution of social
worker caseloads, showing a mode of approximately twenty and that 90 percent
of caseworkers have caseloads between fifteen and twenty-five. Everyone will
nod when the data are presented each month and say "the range of caseloads is
interesting, maybe we should do something to reduce it," but no one asks any
operationally relevant questions. Are the large caseloads all from a single regional
office that is understaffed? Is the heterogeneity in caseloads across caseworkers

appropriate with some caseworkers given a larger number of low-intensity cases and others given a smaller number of higher-intensity cases? Is it resulting from some caseworkers failing to do the paperwork to close inactive cases promptly? It would be straightforward to check in with field office managers about outliers and reasons for this heterogeneity and, if necessary, to pull a sample of cases to sort out what is going on. But this does not happen, and so data are presented month after month without any useful interpretation or follow-up.

Collaborating effectively with service providers outside of government to make sure services are delivered effectively

As I mentioned briefly, government social service agencies rarely do a good job of managing their contracts with private sector social service providers. This is the problem with which the GPL has the most experience helping agencies. To date, we have helped governments to improve their contracting processes for homelessness services, child welfare services, prisoner reentry services, adult basic education services, workforce development services, early childhood home visiting services, pre-K services, and juvenile justice services. Typically, governments contract for services without identifying the strategic purpose they are trying to achieve, and simply pay for slots in programs. They usually fail to measure the outcomes achieved by contractors or to build effective outcome reporting or performance incentives into contracts. Most importantly, they typically fail to actively manage provider performance once contracts are executed. In our work, we help government agencies to track in real time whether individuals referred to services actually receive services, whether the individuals are progressing successfully through the service model, and what their subsequent outcomes are. We help agencies to set up regular (monthly or quarterly) meetings with service providers to review performance metrics and to discuss how the government and the providers can collaborate to achieve better results. For example, in Rhode Island we have helped the state child welfare agency there establish monthly "active contract management" meetings with the four large providers of front-end family preservation services to review performance data and conduct deep dives on issues where the group thinks systems reengineering can produce better results for children and families. The key elements of the active contract management approach are high frequency data-driven purposeful efforts to improve outcomes implemented in such a way that a culture of collaboration develops between government agencies and service providers.

Comparatively evaluating different types of services to rethink the service mix

We find that agencies fall into patterns in which they contract every year for the same set of services with the same set of providers and miss opportunities to alter the service mix to achieve better results. For example, we worked in one state in which rates of placement in long-term congregate care had been steadily

rising for more than a decade in large part because the state had reduced spending on in-home services that could enable families to stay intact. In another state, referrals to a long-established in-home service offered by four providers had declined over five years to substantially less than contracted capacity, leaving the service providers (who were compensated largely based on families served) scrambling to cover their fixed costs. In yet another state, there was a bureaucratic backlog that prevented new families from getting approved as foster care providers, resulting in long stays of children in congregate care. Regular review of rates of referrals to different types of placements and services and regular tracking of cases in which a child or family is referred to a less optimal service because the first choice service did not have available slots can help to direct attention to resource misallocation and to bottlenecks that need to be eliminated. And comparing outcomes for similar individuals referred to different services can help to inform decisions about which services are most effective (and most cost-effective).

These examples are not exhaustive of all the ways social service agencies could improve results through better use of data, but my hope is that they are sufficiently concrete to provide an understanding of the potential for data and data analysis to be used to improve results for priority populations.

There are two main obstacles that are preventing this work from occurring: First, many agencies lack leadership with a time horizon that is sufficiently long to prompt performance improvement projects, that is philosophically oriented toward using data to drive change, and that is willing to bear the stresses associated with driving change. Second, many agencies lack staff with the combination of spare capacity, expertise, and desire necessary to lead data-driven reform projects. In some cases, there are several people in the agency who could lead this kind of work, but they are already occupied more than full time, making sure the trains run on time and all the required quarterly and annual reports are produced. In other cases, there is no one around with the ability to look at data and ask questions about them in the way that is necessary to drive reform. In still other cases, agency leadership fails to empower the capable analysts to roam beyond their narrow silos to address cross-cutting issues; analysts and program specialists need to collaborate effectively. And in still other cases, the capable personnel decline to take on challenging reform projects because the initiatives require extra work and involve risk taking.

There are other more minor obstacles that sometimes arise.[8] For example, some agencies have farmed out data management to private contractors and can only regain access to the agency's own data by paying a large fee to the contractor. In other cases, data warehouses have been set up at local universities, and the gatekeeper to that data becomes a bottleneck. Sometimes agencies have seven-year-old computers meant for word processing and cannot easily procure even a $1,000 work station and a copy of Stata to use for data analysis. Other times outcomes data need to be obtained from a different state agency or a federal agency, and the lawyers for the various agencies operate in a risk-adverse, "it is not permitted" mode rather than in a problem-solving mode. Such "legal"

bottlenecks invariably get solved not by further negotiations between lawyers, but by intervention from an agency head or state budget officer who directs the lawyers to start operating in problem-solving mode. Finally, while most of the necessary data matching and analysis are simple, occasionally there is the need for technical assistance that can allow an agency to perform a more advanced statistical analysis than an agency can perform on its own. But the key point is that in most projects the data matching and data analysis per se are not very hard. What is hard is overcoming inertia and resistance to change and making sure there are people in the agency who are capable of looking at numbers and asking the right operational questions about them.

In thinking about how to assist agencies in overcoming these obstacles it is important to focus both on sustainability and on replication. Technical assistance needs to be provided in a way that allows agencies to continue doing whatever the assistance enabled, even after the assistance ends. Ideally, a technical assistance project that helps a government to do a data-driven management reform not only allows an agency to continue to implement that particular reform, but also to undertake additional projects on its own, without further technical assistance, based on the learning that occurred from during the initial reform. On the replication front, the human service agencies in all fifty states (and many counties and cities) are all essentially in the same business. In theory, a successful data-driven management reform can be transported to other jurisdictions quite easily. In practice, the GPL has found that even when all the steps are known from doing a project in one jurisdiction, the resources necessary for replicating a project in a second jurisdiction can be 70 to 80 percent of what they were in the initial jurisdiction.[9] The hardest parts of reform work are getting leadership and other agency staff to buy into the project, training staff in taking an analytical approach, and building relationships with service providers and other community actors. To a first approximation, all this work needs to be done from scratch in each new jurisdiction. What does improve in replication is one's confidence that the process will produce the desired results.

How can we rapidly increase the number of government social service agencies making effective use of data? There are things that the technical assistance community already knows how to do and simply needs the resources and focus to achieve and things that we still need to find better models for. Organizations such as the Annie E. Casey Foundation, Pew Center on the States, the National Governor's Association, and the Harvard Kennedy School GPL have shown that various mixes of on-site and remote technical assistance can help a willing government to successfully implement a reform project. While there are important lessons that have been learned along the way—how one verifies that a government really is committed to a project, what one can do to increase the chance that the project gets completed in six months rather than in two years, and how to maximize the probability that a new way of doing business sticks after the technical assistance is complete—there are now known techniques for providing technical assistance that result in successful execution of reform projects in government human service agencies. This is not to say that the work is easy; overcoming inertia and deliberate resistance to change is hard. But if we had a

willing agency head, it would be relatively straightforward to provide them with four outside full-time equivalents (FTEs) who could work with four internal FTEs to implement a project that aimed to purposefully use data to reengineer a large number of their systems over a two-year period to achieve better performance (one might want more than four FTEs for a complicated agency such as Medicaid and fewer for a relatively straightforward one such as TANF). One could imagine doing this in two jurisdictions for each of the twelve or so main state human service agencies (they would not need to be the same jurisdictions for all twelve agencies). At the end, we would have templates that could be brought to the other forty-eight states.[10] There are also good existing models for how to share best practices across jurisdictions once the best practices have been developed—models of peer learning, executive education, cohort-based technical assistance and the like.

What currently seems harder is to provide state human service agencies with the human capital necessary to continue to do data-driven reforms on their own once the technical assistance ends. Strong human capital would also greatly facilitate the spreading of best practices once they are developed, because there would be able and willing people in each jurisdiction to implement the new models—we would not have to provide dedicated FTEs to each via outside technical assistance.

Thankfully, the scale of this challenge is not that large. Suppose our goal was to have four data-focused leaders in each of the twelve key social service agencies in each of the fifty states. That is only 2,400 people who need to be recruited and trained: 480 a year if we want to achieve this in 10 years and assume 50 percent turnover. Including the human service agencies in the fifty largest cities might increase the target number by 50 percent. The best models for how to do this lie in cities such as Boston, Denver, and Louisville, which have started by hiring a few pioneer data-focused individuals. The pioneers have helped to create a culture that attracts other young workers with analytic talent.[11] Denver and Louisville have also made systematic efforts to train up a large fraction of their existing staff alongside attracting new talent.

One could imagine a state making a dedicated effort to bring in an initial cadre of a dozen analytically oriented new staff members and spreading them across the human service agencies. While they would initially be somewhat isolated within their agencies, they would have both an interagency peer network and a cross-jurisdiction peer network. And over time as the group within their jurisdiction grew from twelve to forty-eight, they would have new peers within their agencies and have found allies among the existing employees, and the culture around the use of data would shift.

Put simply, it seems to me that we are in a place in which with a purposeful effort to inject data-driven management approaches, we could greatly improve the administration of our most important social programs nationwide in a relatively short period of time. I am optimistic that by improving the administration of these programs we would make major progress in addressing difficult social problems, but that is an unproven hypothesis at this point based largely on

observing how far from the frontier we are right now, and therefore how much low-hanging fruit there appears to be.

Breaking Down Silos with Data-Driven, Outcome-Focused, Community-Level Collaboration

Many of our most difficult social problems cross agency boundaries, and families often receive services from multiple agencies. A problem with how social services are often delivered today is that each service provider is focused on providing units of a specialized service, but no one is accountable for thinking holistically about what it would take to get the individual or the family to an overall success-ful outcome. This tunnel vision exists within government human service agencies as well. Most agencies are organized around program managers who are account-able for making sure their program dollars get spent and for counting how many individuals receive services from their program, but no one is responsible for tracking and managing the overall well-being of the target population.

I had a conversation recently with an expert in a child welfare agency about the challenge of obtaining substance abuse treatment for parents whose addic-tions were leading to neglect of their children. The child welfare agency would refer such parents to the state's substance abuse program, but when a parent failed to show up for treatment, the substance abuse program would simply cross the parent off their list of individuals who were interested in treatment without thinking of the case as one that still needed attention. Too often, social service agencies are structured in these myopic ways—focused on the administration of their own programs not on making progress on population outcomes.

We need to make progress in four different areas to address the problem of silos.

The first and simplest is that agencies, as discussed here, need to properly define the population that is their target, measure outcomes for the full target population, and manage their operations to improve those outcomes. They need to break down the silos and unit-of-services focus within their particular agencies.

Second, we need explicit cross-agency collaborations that allow multiple pro-grams spread across multiple agencies to jointly define the target populations and outcomes they are trying to affect and whose teams meet regularly to review data and spot opportunities to troubleshoot, collaborate more effectively, and improve performance.

Third, we need to experiment with efforts to identify the highest-need families in a jurisdiction and to provide appropriate case management that connects the families to the right mix of services. My GPL recently published a case study about the UK Troubled Families Program that has been applying this approach nationwide in the UK (see Economy and Gong 2017).

There is also a large research project that needs to be done to inform this approach. Some jurisdiction should look at all the individuals who had really bad

outcomes—children involved in the juvenile justice system, children who were maltreated, children who were expelled from school, and so on—and map out how many different families these cases represent, what the overlap is between those showing up as problem cases in different systems, the first time a human service agency became aware of a problem with the individual or family, and where the first opportunity to intervene occurred. Then we need to develop predictive models to inform early intervention efforts that appropriately balance the risks of overproviding and underproviding services.

Fourth, we need to find ways to channel the momentum being produced through collective impact efforts such as the Strive Partnership so that they actually change how services are being delivered on the ground. The collective impact movement has correctly observed that better collaboration and a results-focused orientation are needed not only within government but with all the partners in a community—including philanthropic foundations, employers, and school districts—who impact outcomes for children and other vulnerable populations. But even when there is considerable high-level buy-in for these efforts, it has been hard to translate this buy-in into improvements in service delivery. In some cases, the challenge appears to be that the collective impact effort is not being driven by the entities that control either the funding or the data, and GPL's best current hypothesis is that these initiatives will be more successful when the entities that control the data and the funding—typically government agencies—are at the center of these initiatives.

I have written previously about what I call the "10-year challenge," which could be a framework for bringing about the data-driven, outcomes-focused, community-level collaborations that I believe are necessary to move the dial on complex social programs (Liebman 2013). A funder, either the federal government or a large philanthropic foundation, would choose one or more social problems on which it wanted to make significant progress; examples could be reducing recidivism among ex-offenders, raising third-grade test scores among low-income children living in high-poverty neighborhoods, preventing high-risk youth from dropping out of high school, retraining individuals who have been unemployed for more than nine months, increasing the rate of community college completion, reducing obesity-triggered diabetes, eliminating chronic and/or family homelessness, or helping developmentally disabled youth make successful transitions into the adult workforce, among many others. All the problems would be ones where the specific individuals in the population to be served can be identified and baseline outcomes can be established; these two factors will provide an observable baseline against which improvement can be measured.

Through a grant competition, ten communities would be selected for each problem in an effort to transform outcomes for the specific population within 5 to 10 years. In a first step, the funder would issue planning grants of approximately $250,000 each to several dozen communities that were interested in putting together proposals. Then ten communities would be selected for more substantial funding based on how likely the proposed project is to make significant progress in addressing the social problem, the potential for the project to yield rigorous evidence about what works, and the extent to which the approach

demonstrated by the project could be spread nationwide. Although it would be terrific if all ten communities were successful, the real goal would be to discover two or three transformative approaches for each policy problem—solutions that could then be developed and implemented nationwide. In the original proposal, I suggested that average-sized projects would spend $10 million a year on services and serve approximately one thousand to two thousand individuals (with flexibility depending on the nature of the intervention and the size of the community and of the target population). I suggested that the primary funder (the federal government or a philanthropic foundation) would cover one-third of the costs of the intervention, state and local governments would need to agree to providing another one-third, and private community partners would be required to cover the remaining third. In addition, each chosen jurisdiction would receive $1 million per year for technical assistance on data analysis and evaluation. In total, this initiative would cost the primary funder approximately $40 million per year for 5 to 10 years per social problem. Clearly all these numbers can be adjusted by varying the number of people served and the number of communities working on solutions to each problem.

Final Words

The modern era of social-policy development and evaluation dates back to the late 1960s when large datasets and randomized experiments began to be used regularly to evaluate federal policy initiatives. Decades later, we have a better sense of what works to address certain social challenges, but we are still very far from where we need to be. We still lack proven, cost-effective, scalable solutions to most social problems, and, despite significant government investment, we are failing to make sufficiently rapid progress in addressing our most serious challenges. Today, if a governor were to ask his or her policy advisors for a state-wide program that could cut recidivism among individuals recently released from state prison by one-third, or a program that could raise the employment of welfare recipients by 10 percentage points, there is no intervention currently available for those advisors to offer the governor that has more than a 50 percent chance of working. Even in early childhood education, where the evidence of successful interventions is strong, if the governor were to ask for an initiative to eliminate half of the gap in third-grade test scores between more- and less-affluent students, it is far from certain that an initiative could be designed and implemented to achieve that target with what we know today.

Part of the reason we lack solutions to many social problems is that the problems are hard and human beings and their social environments are complex. But it is also the case that our current mechanisms for funding and evaluating social programs do not produce a culture of continuous learning and improvement; nor do they generate opportunities for fundamental reengineering of systems to produce better results. Sustained purposeful efforts to actually move the dial on a particular social problem in a particular community are rare. My conjecture is

that if we take advantage of the great expansion in the availability of data and analysis tools to actually try to move the dial on social problems in a data-driven, outcomes-focused way, we might find that we succeed.

Notes

1. Bridgeland and Orszag (2013, 63) estimate that "less than $1 out of every $100 of government spending is backed by even the most basic evidence that the money is being spent wisely."

2. For example, Baron and Sawhill (2010, 21) report that nine of ten evaluations of entire federal social programs found "weak or no positive effects."

3. The social entrepreneur George Overholser has observed that "evidence melts like ice cream." See Overholser (2014).

4. The GPL provided technical assistance to the state of South Carolina on a project that is expanding funding for NFP and providing a platform for a new randomized evaluation of the impact of NFP. We are also assisting two other states that are developing similar projects.

5. There is a parallel between my argument and recent developments in federal management policy. The Bush administration's Program Assessment Rating Tool (PART) focused on systematically assessing whether each government program was achieving its goals and was often perceived by agencies and congressional committees as a mechanism to apportion programs into those that were worth keeping and those that should be eliminated (Moynihan 2008). The Obama administration's performance.gov approach focused on tracking and improving performance trends. See Metzenbaum (2009) for an influential presentation of these ideas, and Metzenbaum and Shea (2017) for a recent assessment.

6. Further details on the Seattle project are available in Azemati and Grover-Roybal (2016).

7. See Behn (2014) on the difference between creating a white elephant performance dashboard and actually undertaking a data-informed performance leadership strategy. Behn notes that in contrast to the popular expression, data do not in fact speak for themselves.

8. See Goerge (this volume) for an insightful discussion of barriers researchers face in trying to access state government administrative data.

9. Perhaps the clearest demonstration of this is in our pay for success/social impact bond work. There are now sixteen pay-for-success projects in the United States. The GPL has provided government-side technical assistance on ten of them. But even though we understand all the steps are necessary to implement a project and have done it many times, it is still taking at least two years from conception to service delivery in these projects. Certain steps in the process—getting decisions from government decision-makers, obtaining legislative authority, helping local service providers become comfortable with the pay-for-success model and able to understand the financial implications of increasing the scale of their operations, setting up the data systems necessary to track outcomes, and waiting for investors to be recruited—need to be done anew in each project.

10. Bloomberg Philanthropies' What Works Cities initiative provides a good example of how to do multijurisdictional technical assistance at scale. See Blauer (this volume).

11. On the Boston example, see Steve Poftak (2016). On Denver, see Brian Elms (2016). See Blauer (this volume) for a broad vision of how cities might make better use of data in the future.

References

Azemati, Hanna, and Christina Grover-Roybal. 2016. *Shaking up the routine: How Seattle is implementing results-driven contracting practices to improve outcomes for people experiencing homelessness.* Cambridge, MA: Harvard Kennedy School Government Performance Lab.

Baron, Jon, and Isabel Sawhill. May 2010. Federal programs for youth: More of the same won't work. *Youth Today.*

Behn, Robert D. 2014. *The PerformanceStat potential*. Washington, DC: Brookings Institution Press.

Blauer, Beth. 2017. Building the data city of the future. *The ANNALS of the American Academy of Political and Social Science* (this volume).

Bridgeland, John, and Peter Orszag. July/August 2013. Can government play moneyball? *The Atlantic*.

Economy, Christina, and Gloria Gong. 2017. *UK Troubled Families Programme: Lessons from local authorities*. Cambridge, MA: Harvard Kennedy School Government Performance Lab.

Elms, Brian. 2016. *Peak performance: How Denver's Peak Academy is saving money, boosting morale, and just maybe changing the world*. With J.B. Wogan. Washington, DC: Governing.

Goerge, Robert M. 2017. Barriers to accessing state data and approaches to addressing them. *The ANNALS of the American Academy of Political and Social Science* (this volume).

Liebman, Jeffrey. 2013. *Building on recent advances in evidence-based policymaking*. Hamilton Project. Washington, DC: Brookings Institution.

Metzenbaum, Shelley H. 2009. *Performance management recommendations for the new administration*. Washington, DC: IBM Center for the Business of Government.

Metzenbaum, Shelley, and Robert Shea. 11 January 2017. Memo to the president: Performance accountability, evidence and improvement. Government Executive.

Moynihan, Donald P. 2008. *The dynamics of performance management: Construction information and reform*. Washington, DC: Georgetown University Press.

Overholser, George. 9 September 2014. Statement of George Overholser. Hearing on Social Impact Bonds and Families in Need. House Committee on Ways and Means Subcommittee on Human Resources.

Poftak, Steve. 2016. *City Hall's data and technology journey: Using data to improve the lives of citizens*. Boston, MA: Rappaport Institute for Greater Boston.