AMERICA
ACHIEVES
RESULTS FOR AMERICA

THE
HAMILTON
PROJECT

Building on Recent Advances in Evidence-Based Policymaking

Jeffrey B. Liebman, Harvard University

April 2013

BROOKINGS

Building on Recent Advances in Evidence-Based Policymaking

Jeffrey B. Liebman, Harvard University

April 2013

A paper jointly released by Results for America and The Hamilton Project

Abstract

The current fiscal environment makes it imperative that we produce more value with each dollar that government spends. Doing so will require better use of evidence in policymaking. The good news is that over the past decade new government strategies have begun to emerge—at the federal, state, and local levels—that simultaneously offer the potential to make better use of taxpayer dollars and speed up progress in addressing serious social problems. These strategies: subsidize learning and experimentation so that new solutions are developed, increase the amount of evidence on the effectiveness of existing and potential new programs, make greater use of evidence in budget and management decisions, make purposeful efforts to target improved outcomes for particular populations, and spur innovation and align incentives through cross-sector and community-based collaborations. This paper describes the new strategies. It also proposes several steps to advance the use of evidence-based policy in the federal government, including giving agencies the authority to reserve a percentage of program spending to fund program evaluations and expanding the use of tiered evidence standards in grant competitions. Finally, it recommends two initiatives that would supplement the diffusion of these evidence-based practices with a more-focused approach that aims to supply solutions for specific high-priority social problems. The Ten-Year Challenge would tackle ten social problems by establishing data-driven, outcome-focused initiatives in one hundred communities. A federal Pay for Success initiative would help state and local governments establish Pay for Success projects in areas like early-childhood education where state and local activity has the potential to achieve important federal policy objectives or produce significant federal budget savings.

Contents

Introduction

At every level of government, policymakers are facing the same challenge: How do we continue to innovate and make additional progress in addressing our nation's problems when budget cuts are making it difficult, if not impossible, to hold onto the gains we have already made?

The fiscal pressures that governments are experiencing today will persist even after the economy fully recovers from the recent recession. With spending on Medicare, Medicaid, and Social Security projected to rise by over 3 percent of GDP in this decade as the baby boomers retire and health-care expenditures continue to grow, the resources available for everything else government does will be squeezed.

The only way to keep making progress in this fiscal environment is to produce more value with each dollar that government spends. Doing so will require better use of evidence in policymaking. We need to reallocate funds from less-effective programs to more-effective programs. We need to leverage government resources to encourage the development and evaluation of new, more-effective solutions. We need to improve performance by setting outcome-focused goals and then using leadership strategies such as data-driven performance reviews and community collaboratives to make the changes to systems necessary to achieve those goals. And to the extent that additional revenue may be part of the way we close the fiscal gap, we need to produce the evidence on program effectiveness necessary to demonstrate to taxpayers that their additional tax dollars are being put to good use.

Although fiscal pressures make the need for more-effective government particularly acute, the imperative for government to achieve more per dollar spent would exist even if budgets were flush.

Despite spending billions and billions of dollars each year, we are simply not making rapid enough progress in addressing social problems. From recidivism to school readiness, and obesity to workforce development, we have either failed to develop effective solutions, failed to prove that the solutions work, or failed to scale the solutions that do work widely.

These failures clearly have adverse consequences for the affected individuals and their communities. But they also have serious consequences for the U.S. economy. When there are seven million youth who are neither in school nor working (Belfield, Levin, and Rosen 2012), when half of low-income fourth graders are not even reading at the "basic" level (Fiester 2010), when one in fifteen African-American men is incarcerated, and when the United States ranks fourteenth among Organisation for Economic Co-operation and Development (OECD) countries in college graduation rates, clearly we are not maximizing the potential of our labor force. As a result, our standard of living is lower, and economic growth is reduced.

Spreading the new strategies at all levels of government

The good news is that over the past decade, and particularly over the past five years, new government strategies have begun to emerge—at the federal, state, and local levels—that simultaneously offer the potential to make better use of taxpayer dollars and speed up progress in addressing serious social problems. These strategies:

- Subsidize learning and experimentation so that new solutions are developed

- Increase the amount of evidence on the effectiveness of existing and potential new programs

- Make greater use of evidence in budget and management decisions

- Make purposeful efforts to target improved outcomes for particular populations

- Spur innovation and align incentives through cross-sector and community-based collaborations

Despite the encouraging progress, it remains the case that most government spending is not allocated based on evidence or with a focus on innovation or performance. Even in federal agencies that have been the most creative in developing new ways to allocate grant funds, such as the U.S. Departments of Education and Health and Human Services (HHS), the bulk of spending is distributed through conventional grants and other traditional mechanisms.

Therefore, it is an important priority to spread these new strategies widely—to additional programs within the federal agencies that are already using them, to additional federal agencies, and to more state and local jurisdictions. To this end, this paper describes nearly two dozen of the new strategies and provides a framework for identifying which strategy is the best fit for achieving a given policy objective.

Advancing the use of evidence-based policy in the federal government

To advance the use of evidence-based policy in the federal government, the paper also proposes the following specific steps:

- All agencies that administer social programs should have the authority, similar to that currently provided to the U.S. Department of Labor (DOL), to reserve a percentage of program spending to fund program evaluations.[1] In addition, a small evaluation fund should be provided to the Office of Management and Budget (OMB) for cross-agency evaluation initiatives.

- In high-priority policy areas where sufficient evaluation evidence is available, agencies should be required to use tiered evidence standards (described in detail below) in grant competitions.

- As the evidence base becomes strong enough, federal formula funding provided to state and local governments should begin to require that a portion of such funds (e.g., 5 percent) be spent on programs that have been proven effective.

- Congress should task either the Congressional Budget Office (CBO) or the Government Accountability Office (GAO) with producing cost-effectiveness reports in several priority policy areas. Sufficient funding should be provided to prevent this from becoming an unfunded mandate on the agency that is given this assignment.

If these evidence-based approaches to policymaking spread widely, we will achieve better outcomes with government expenditures by replacing less-effective government programs with programs that work better, and we will develop new, more-effective approaches.

Developing solutions to our most serious social problems

But if our goal is to make significant progress in addressing our most serious social problems, simply expanding the use of these strategies is unlikely to be enough to produce the results we require. We need to supplement the wide diffusion of these practices with a more-focused approach that aims to supply solutions for specific high-priority populations.

Consider the challenges of reducing recidivism among ex-offenders, raising third-grade reading and math skills among low-income children living in high-poverty neighborhoods, preventing youth from dropping out of high school, helping chronically unemployed individuals obtain and keep jobs, raising community college completion rates, reducing obesity-triggered diabetes, eliminating chronic and/or family homelessness, and helping developmentally-disabled youth make successful transitions into the adult workforce, among many others.

Although dozens of programs spend billions of dollars serving these populations, services are delivered in a highly fragmented manner. One program provides mental health services, another provides housing services, and still another provides job-readiness training. But in most cases, no one is responsible for ensuring that a specific cohort of individuals in a particular community achieves successful outcomes. Moreover, it is not just government programs that determine whether success is achieved. Results depend on the joint actions of multiple levels of government, diverse not-for-profit and for-profit service providers, private businesses, and philanthropic and other community partners.

This paper proposes two initiatives designed to overcome this fragmentation and produce the community-level collaboration necessary to make real progress on our most persistent social problems.

The Ten-Year Challenge: Solve ten problems in ten places in ten years

Congress and the President should work together to identify ten social-policy problems where it is a national priority to find solutions. For these ten problems, policymakers and communities must have the ability to identify specific individuals in the population to be served and the ability to measure the population outcomes; these two factors will provide an observable baseline against which improvement can be measured. Then, through a grant competition, ten communities would be selected for each challenge—one hundred communities overall—in an effort to transform outcomes for the specific population within five to ten years.

The goal would be to discover two or three transformative approaches for each policy problem—solutions that could then be applied nationwide. The theory behind this initiative is that solving most of these problems will require a creative reengineering of systems and practices by multiple partners in each community, both governmental and non-governmental, and that for this to happen there needs to be a data-driven collaborative focus on achieving measurable improvements in outcomes for specific cohorts of individuals.

Strategic support for state and local Pay for Success initiatives

The federal government should become a strategic partner with state and local governments in establishing Pay for Success/Social Impact Bond projects by helping to establish Pay for Success projects in areas where state and local activity has the potential to achieve important federal policy objectives or produce significant federal budget savings. The highest priority would be to create an initiative around early childhood interventions including home visiting, early learning, and preschool, or initiatives spanning birth to second grade. Evidence of success in early childhood programs is strong, but state and local governments are having trouble establishing Pay for Success projects in this area because much of the financial benefit of these interventions accrues to the federal government.

The Need to Strengthen Evidence-Based Policymaking

We are flying blind

The simple truth about most government social spending (and many other kinds of spending) is that we have no idea how effective it is. Performance is rarely assessed, and measurement tends to focus on tracking the number of people served and the amount of service provided, rather than the outcomes that are achieved. Although some local governments and federal agencies have improved their measurement of outcomes in recent years, most programs have never been rigorously evaluated. Just as problematic, most agencies lack sophistication in coding and analyzing the data they do collect so they fail to spot patterns and variations in performance that might point to promising or problematic practices. When a government contracts with multiple service providers, it almost never puts systems in place to compare their relative performance so that future funding can be allocated to the best performers. There is no systematic ranking of the cost-effectiveness of different programs.

The recent experience of the federal Head Start program provides a telling example. The program provides early childhood education to low-income children between the ages of three and five. Established in 1965, it is a $7 billion-a-year program that serves over 800,000 preschoolers annually. Yet until recently we had no idea how effective it was. When it was finally submitted to an impact evaluation beginning in 2002, the results were disappointing. The evaluation found "modest to moderate positive impacts" that "largely disappeared by the end of first grade" (Advisory Committee on Head Start

Research and Evaluation 2012, 27–28). In addition, it had become apparent that the quality of different Head Start centers varied widely and that there was a need to systematically assess center quality on an ongoing basis.[2]

The point of this example is not to question the effectiveness of Head Start. The balance of the evidence suggests that its benefits significantly outweigh its costs.[3] Moreover, the 2007 reauthorization legislation, cosponsored by Senators Edward Kennedy and Mike Enzi and signed by President George W. Bush, and the 2011 HHS regulations establishing the Designation Renewal System, now require providers that fail to meet performance benchmarks to recompete for funds. The point is, if we can go forty years without knowing whether a program as important as Head Start is working and without spotting performance problems that could have been fixed with management reforms, we have a long way to go in introducing evidence-based practices more broadly.

The economic consequences

The lack of evidence-based practices has economic, fiscal, and social consequences. At the most basic level, this nation's prosperity depends on the productivity of its workforce and the share of Americans who are employed. When more Americans are working and when each is producing more per hour, our per capita output, and therefore our standard of living, increases. International evidence consistently shows that countries with more human capital are richer and have more rapid economic growth (Barro and Lee forthcoming;

Hanushek and Woessmann 2012; Mankiw, Romer, and Weil 1992; Cohen and Soto 2007).

Although the United States ranks high on human capital measures, it also has considerable untapped potential. In particular, we are not doing an adequate job of educating low-income children.[4] Analysis of data from the National Assessment of Educational Progress finds large gaps in educational achievement between children in middle-income and low-income households, gaps that translate later in life for the latter into lower employment rates and wages, and higher rates of incarceration and disability.

All levels of government devote significant resources to this challenge, from up-front investments such as prenatal services, early childhood education, and special education to later investments such as grants for college tuition and job training for high school dropouts. It is therefore essential to improve the effectiveness of the interventions we are already making, reallocate funding to the approaches that have the greatest impact, and develop better approaches. In fact, demographic trends are making this an increasingly urgent challenge, with population groups that are failing to achieve their full educational potential becoming a larger share of the workforce (Ellwood 2001). As a result, our success in educating low-income children will be one of the most important determinants of how fast the U.S. economy grows over the next couple of decades.

Auguste, Hancock, and Laboissiere (2009) conclude that closing income-based academic-achievement gaps in the United States would raise the level of U.S. GDP by between 3 percent and 5 percent one decade after the achievement gaps are closed. It takes a long time for educational improvements to translate into a more productive workforce, but if half of the income gap could be closed over the next decade, these estimates suggest that doing so would add about 0.1 percent to average annual growth rates over the next twenty years—equivalent to more than an extra year of productivity growth. Government also plays an important role in other productivity- and living standard–enhancing investments including

infrastructure, public health and safety, research and innovation, and environmental protection. Thus, the potential economic gains from more-effective policymaking could be significantly larger than the human capital calculation suggests.

Now more than ever

There is also a clear fiscal policy imperative for evidenced-based policymaking. With health-care expenditures rising, and costs associated with the baby boomers' retirement crowding out spending on everything else, the need to "do more with less" has never been greater. The discretionary spending caps set by the Budget Control Act of 2011 are scheduled to reduce non-defense discretionary spending to 2.7 percent of GDP, compared to a forty-year average of 4.0 percent of GDP (CBO 2013). In this context, smart decisions need to be made about which government activities to continue, which to expand, and which to eliminate.[5] For those that are continued, funding constraints mean that we will need to achieve much better results per dollar of spending than we do today. Investing in what is most effective is a basic stewardship requirement for those entrusted with public dollars. It is irresponsible not to make the investments in evaluation and performance measurement necessary to make good decisions.

The economic and fiscal need for more evidence-based policymaking is strong. But more directly, we are simply not doing well enough in developing solutions to social problems. Since 1990, ten major federal social programs have been evaluated using randomized experiments. Nine of those evaluations "found weak or no positive effects" (Baron and Sawhill 2010). Meta-analyses of interventions in broad policy areas like job training and prevention of recidivism consistently find disappointing results.[6] For most social-policy objectives, we have no proven solutions.[7] To rectify this, we need to reform the way government does social spending so that we produce more experimentation, perform more rigorous evaluation of innovative ideas, and pay more attention to performance in funding decisions.

Five Strategies for Evidence-Based Policymaking

Over the past decade, and particularly over the past five years, state, local, and federal governments have implemented a large number of new evidence-based practices. Evaluation experts Jon Baron of the Coalition for Evidence-Based Policy and Ron Haskins of the Brookings Institution have written that the Obama Administration's efforts in this area "constitute the most sweeping and potentially groundbreaking emphasis on rigorous program evaluation ever conducted by the federal government," while also describing how the Obama initia-

tives evolved from work done at OMB during the George W. Bush Administration (Haskins and Baron 2011, 25–51). The amount of new activity at the state and local levels has been even greater.

To help set the stage for a more systematic discussion of principles to guide greater use of evidence-based policymaking, boxes 1–3 provide examples of three particularly innovative evidence-based initiatives.

Box 1. Tiered Evidence Standards in the Department of Education's Investing in Innovation Fund (i3) Competitive Grant Program

The Department of Education's i3 program provides competitive grants to local education agencies to expand innovative practices that have been demonstrated to improve student achievement, increase high school graduation rates, or increase college enrollment and completion. The program established three tiers: scale-up grants to fund expansion of practices for which there is already strong evidence, validation grants to provide funding to support promising strategies for which there is currently only moderate evidence, and development grants to provide funding to support "high-potential and relatively untested" practices.

The Department of Education also established standards for evidence. "Strong evidence" requires a prior randomized trial or a rigorous quasi-experimental design. "Moderate evidence" is defined as promising research that had a flaw such as insufficient sample sizes or a potential for selection bias that limited the amount of confidence that could be placed in the research.

Over three rounds of competitions, the Department of Education has awarded five scale-up grants, twenty-eight validation grants, and fifty-nine development grants with total grants of $940 million. Most of the scale-up grants provided approximately $50 million each; most of the validation grants were for approximately $15 million; and most of the development grants were for approximately $3 million. Thus, the program reserved the largest blocks of funding for proven practices, while also investing in promising but not fully-proven approaches. It also required rigorous evaluation plans from grantees, so that unproven programs can, over time, become proven programs if they are shown to work.

Box 2. New York City's Center for Economic Opportunity

The Center for Economic Opportunity (CEO) was established in 2006 by New York City Mayor Michael Bloomberg to develop and evaluate innovative approaches to reducing poverty. It is supported by a $100 million annual fund that is used to implement and evaluate new solutions. Initiatives that are successful then become eligible for permanent funding through the city budget. The center has produced more than fifty initiatives, many of which have been evaluated using randomized controlled trials. Its annual report provides a rigorous accounting of what has worked and what has not. It describes the successful initiatives that are receiving ongoing funding, the pilot programs that are still in the development and evaluation stage, and the discontinued programs that were unable to demonstrate sufficient outcomes to earn continued funding.

Initiatives that produced demonstrated results and transitioned to receiving ongoing funding include:

- The Accelerated Study program at the City University of New York that provides a range of academic and support services to assist students in earning associate's degrees within three years

- The WorkAdvance program that helps low-wage workers advance in the labor market by providing services such as individual advancement coaching, access to training and education programs, and job placement assistance

- $aveNYC, which offers eligible individuals a 50-percent match if they deposit a portion of their tax refund into a savings account and maintain the deposit for a year

- Sector-Focused Career Centers that focus workforce development services on a single economic sector, providing low-income workers with access to jobs in transportation, manufacturing, and health care

The Center for Economic Opportunity has received funding from the federal Social Innovation Fund to expand five of its initiatives both within New York City and in additional cities.

Box 3. The Greater Cincinnati Strive Partnership

The Strive Partnership is an initiative in Cincinnati, Ohio, and two neighboring cities in Kentucky—Covington and Newport—that aims to improve student achievement "from cradle to career." The partnership has brought together a wide range of community members to focus in a data-driven way on achieving eight outcomes: kindergarten readiness, fourth-grade reading proficiency, eighth-grade math proficiency, high school graduation rates and ACT scores, as well as postsecondary enrollment, retention, and completion.

The partnership takes an "every child" approach and regularly tracks the number of students who are not achieving target outcomes to form strategies to deliver the services necessary to raise the number of successful outcomes. Different community organizations are taking the lead on various components of the initiative. For example, the United Way of Greater Cincinnati's Success By 6 initiative has been focused on improving the quality of early childhood education; several organizations have established collaborative tutoring and mentoring programs; and the partnership has recently launched a Talent Pipeline Initiative to better connect employers and high schools in order to establish more robust and accessible career pathways for students.

Kindergarten readiness has risen in all three communities since the partnership began, and the Cincinnati Public Schools became the first urban school district in Ohio to be rated "effective." What is remarkable about this initiative is that it is not simply a partnership to track outcomes. It represents a commitment among community members to doing business differently to create pathways to success for children in the community.

Five challenges in evidence-based policymaking

Conceptually, there are several challenges in using evidence-based practices to improve the performance of government and make better use of taxpayers' dollars. Most of the new evidence-based practices can be seen as a response to one or more of these five challenges.

First, innovative solutions need to be developed. The United States has many successful social programs. Our schools have given us one of the world's most productive workforces.[8] Social Security has nearly eliminated poverty among the elderly (Meyer and Sullivan 2010; Engelhardt and Gruber 2006). Medicare and Medicaid have reduced the risk associated with out-of-pocket health-care expenditures (Finkelstein and McKnight 2008). After forty years of applying rigorous evaluation methods to social policy, we have learned a lot both about what works and what does not. But for many of our most important social problems, we still do not have any proven, cost-effective, scalable strategies.

The challenge is how to finance the needed program development and experimentation. If we fund ten promising early childhood interventions and only one succeeds, and that one can be scaled nationwide, then the social benefits of the overall initiative will be immense. But nine out of ten of the individual projects will have been failures. Therefore, we need funding mechanisms—whether from private philanthropy, public R&D-style grants, or prize competitions—that pay for the learning (and failure) necessary to ultimately come up with successful solutions.

Second, the innovative solutions need to be evaluated rigorously, and we need to implement faster and less-expensive ways of figuring out what works. Today, the outcomes of most programs are never measured, much less evaluated by comparison to a rigorous counterfactual scenario. Occasionally, a program will be subjected to a large program evaluation that, by the time the results are available, tells us whether or not the program worked in the past but not necessarily how the program is working today.

The information-technology revolution offers the means to do much better. Government administrative data systems can be used to provide important outcome measures on a monthly or quarterly basis, and to compare the performance of different providers. These systems can also be used to implement low-cost evaluation strategies, randomizing clients to different providers or using quasi-random approaches based on pro-

gram wait lists or eligibility cutoffs, among other methods, to establish reasonable counterfactuals. It is time to bring the private-sector practices of rapid-cycle innovation and use of data for continuous improvement to the public sector.

Third, governments need to actually use the evidence that is generated. Today we tend to fund exactly the same things we funded last year, and since we generally do not measure performance, we have no basis for doing anything differently. This means we continue to fund things that don't work, we miss opportunities to improve performance, and we fail to adopt effective new solutions. Most of the obstacles here are due to lack of information. We don't have basic facts or when we do, we lack the cost-effectiveness analysis necessary to compare approaches. But there are political obstacles to using evidence as well. Legislatures sometimes earmark funding for particular providers, thwarting efforts to award grants through merit-based competitions. An evidence-based initiative that threatens to reduce the market share of an incumbent provider who has friends in the legislature may not endure, while an ineffective program that is politically connected may be immortal. Haskins and Baron (2011) note that programs with weak evaluation results tend to continue to be funded at essentially the same level as before the evaluations were conducted.

Fourth, strategic efforts need to be undertaken to target improved outcomes for particular populations. Most government programs fund specific providers who serve a set number of people who meet particular eligibility requirements. Typically, the providers recruit clients, and clients often receive multiple types of services from different programs managed by different agencies. In this fragmented system, it is no one's job to identify all of the people in need of service, track their outcomes, and make sure they receive the services they need to be successful. This needs to change. We need to reorganize both government and service provision to focus on achieving better outcomes for entire cohorts of high-risk individuals in specific communities.

For example, cities and states should work with community partners in a data-driven way to:

- Identify all of the low-income preschoolers in high-poverty neighborhoods and allocate available resources to maximize the number that reach kindergarten ready to learn

- Identify all of the at-risk youth who are involved in the juvenile-justice and child-welfare systems and come up with a strategy for all of them to make successful transitions to adulthood

- Identify all of the individuals living in public housing, determine which ones have the potential for economic self-sufficiency, and target services to make that possible

- Identify all of the adults being released from prison each quarter and make sure processes are in place to connect them with the right post-release services

Governors and mayors should be demanding to see quarterly reports showing the outcomes being achieved in these populations and holding their managers accountable for performance trends. The PerformanceStat approach to city, state, or agency management provides evidence that data-focused leadership can improve outcomes.

Fifth, structures need to be developed that provide proper environments and incentives for cross-sector collaboration. Outside of education, most social-service provision is done by private providers while most of the financing comes from government. This means that government's task is to create an environment with its grant competitions and procurements so that providers and their philanthropic partners can be successful in innovating, producing evidence on what works, and scaling up effective services.

The new government strategies

As described above, new strategies have begun to emerge—at the federal, state, and local government levels—that offer the potential of simultaneously making better use of taxpayer dollars and speeding up progress in addressing serious social problems. These initiatives can be grouped by which of the five challenges they primarily target, though many of them target more than one challenge.

1. Subsidizing learning and experimentation to develop new solutions

The development of a successful social innovation is a public good. Once a solution is developed, it can be adopted widely. As with other areas of innovation, the full benefits of social innovation are not captured by the innovator. Thus, without substantial philanthropic or government funding we are likely to see too little social innovation. Three main tools are being used to subsidize the basic research stage:

Investing directly in innovation. The simplest way to increase the amount of social innovation is to fund it directly through either government or philanthropic grant competitions and innovation funds. The New York City CEO approach of dedicating $100 million a year to testing innovative approaches is an example of this. So is the funding of development grants in the Department of Education's i3 program. A challenge in using this direct approach is that funders may not have the information necessary to make good decisions about which innovative approaches are worth testing. Requiring grantees to raise matching funds, as in the Social Innovation Fund (discussed further below), can decentralize the information gathering by using the test of whether additional funders can be convinced of the merits of the intervention as a screen for the project's potential. Requiring and funding a rigorous evaluation is essential in order to have these direct investments pay off by generating learning about which approaches work.

Guaranteeing payments if solutions are developed. In producing private consumer goods, an innovator knows that if he or she produces a terrific product there is a potential market waiting for it. In contrast, when successful social innovations are developed there is no guarantee that governments will decide to "buy" and scale them up. If governments can guarantee that there will be buyers for successful social innovations, then there will be more incentive for philanthropies and social entrepreneurs to invest in developing them.

President Bill Clinton's Millennium Vaccine Initiative, which promised to subsidize sales of new vaccines developed for malaria, tuberculosis, and HIV/AIDS, is the most famous example of this kind of structure. Social Impact Bonds (described further below) implicitly function this way because they build a relationship with the government in which a successful pilot is highly likely to be taken to scale. Grants that use tiered-evidence standards (such as the i3 program) also implicitly create a market for proven social innovations by awarding the largest grants for solutions that are backed by the strongest evidence of effectiveness.

Awarding prizes. In private markets, a successful entrepreneur has the potential to earn profits that are many times his or her initial investment. Undertaking a project that has a one-in-three chance of success can make sense if there is the potential to make millions or even billions of dollars if the project should succeed. With innovation that is targeted at solving public problems there is typically no such opportunity for a huge pay-

off for the entity undertaking the investment, making it difficult to raise funds for risky initiatives even when those initiatives have a chance of producing substantial social value.

One way to approximate the private-sector reward structure is to establish large prizes for the development of solutions to important public problems. Through guidance issued by OMB in 2010, and further authority obtained in the 2011 American Competes Act reauthorization, the Obama Administration is working with federal agencies to promote prizes to "establish an ambitious goal without having to predict which team or approach is most likely to succeed; benefit from novel approaches without bearing high levels of risk; reach beyond the 'usual suspects' to increase the number of minds tackling a problem; bring out-of-discipline perspectives to bear; increase cost-effectiveness to maximize the return on taxpayer dollars; and pay only for success" (Office of Science and Technology Policy 2012). Forty agencies have posted more than 150 prizes on the Challenge.gov website. NASA, the Department of Energy, and the Department of Defense have all succeeded in generating solutions to difficult problems using the prize approach.

2. Increasing the amount of evidence on what works

Although developing and evaluating new approaches is an important objective, we also need to figure out whether existing programs work and how to improve them.

Prioritizing evaluation funding. Spending a few hundred million dollars more a year on evaluations could save tens of billions of dollars by teaching us which programs work and generating lessons to improve programs that don't. In preparation for President Obama's FY 2011 budget, OMB established an evaluation initiative.

Agencies were invited to submit proposals for program evaluations and provided with technical assistance in developing the proposals. OMB set aside a pot of money for the evaluations so that any funding provided to agencies for the evaluation initiative would be above and beyond the spending level negotiated with the agencies for their regular operations. This ensured that there would be no opportunity cost to agencies in terms of other budget priorities from agreeing to have their programs evaluated. The proposal also included funding to allow agencies to beef up their evaluation capacity by hiring evaluation experts. In total, the President's budget proposed $100 million to per-

form thirty-five evaluations of programs spanning seventeen agencies. However, gridlock between the White House and Congress resulted in the federal government operating under a continuing resolution for the entire year, and funds for the evaluation initiative were never appropriated.

This is an example of a broader problem with attempts to evaluate federal programs. By the time a specific evaluation proposal is developed and authority is received from Congress, several years are likely to have passed. Then it can take a couple of years to set up the evaluation. Often the staff member who initiated the evaluation will have moved on and without this sponsor's enthusiasm, the project never moves forward. The four other approaches to generating evaluation evidence described in this section are, in part, responses to the difficulty involved in directly funding evaluations through the appropriations process.

Building evaluation into program DNA. When he was OMB director, Peter Orszag famously blogged that the Administration "was designing new initiatives with evaluation standards built into their DNA." His point was that if evaluation can be incorporated into the core operations of a program, then the lengthy, and likely unsuccessful, attempt to get a specific appropriation from Congress to do an evaluation will not be necessary. Thus, many of the new grant competitions initiated during the Obama Administration have made rigorous evaluations a condition for making awards, and the quality of the proposed evaluation has been used as a selection criterion. The New York City CEO initiative similarly builds evaluation directly into the initiative itself.

Setting aside a portion of program funding for evaluation. FY 2012 appropriations for the DOL provide transfer authority allowing the secretary of labor to reserve up to 0.5 percent of program spending for program evaluations overseen by the Office of the Chief Evaluation Officer. This provision provides a good model because it gives the chief evaluation officer the flexibility to focus evaluation dollars on the agency's top priorities.

Incentives for producing evaluation evidence. Another way to generate more evaluation evidence is to require evidence as a condition of funding. The more programs there are like the i3 program, which reserve the largest grants for applicants whose effectiveness is backed by rigorous evaluation evidence, the more likely it is that providers and philanthropies will work to make more such evidence exist. The department has recently issued a proposed rule that would revise its General Administrative Regulations (EDGAR) to allow any of its programs that make dis-

cretionary grant awards to give "special consideration to applications supported by strong or moderate evidence of effectiveness." This would enable any department competitive grant program to use the i3 tiered approach.

Low-cost evaluations using administrative data. In the past, a major obstacle to building up evaluation evidence has been the cost associated with data collection (an in-person survey can easily cost $1,000 or more per sample member) and the delay (often five to ten years) between when an evaluation is proposed and when evidence from the evaluation becomes available. Although some delay is necessary to measure the long-run impacts of a program, with advances in information technology and the availability of large administrative data sets measuring key outcomes such as earnings histories, college enrollment, and criminal trajectories, there is the opportunity to do evaluations more cheaply and more rapidly and to provide ongoing assessments of current program effectiveness rather than one-time snapshots that simply tell us whether the program worked at some point in the relatively distant past (Overholser 2011).

3. Making greater use of evidence in budget and management decisions

There are several challenges to making budget decisions based upon evidence of program effectiveness. For many programs conclusive evidence does not exist. Even when evidence of program impacts does exist, it can be analytically challenging to extrapolate from that evidence to broader policy impacts or to systematically compare the relative effectiveness of different policy options. When funding is distributed to lower levels of government or is used to finance provision by numerous private-sector service providers, effectiveness can vary widely in different locations. In some cases, underfunding is what causes programs to be ineffective. Thus, the best policy response to poor performance can be to increase funding. On the management side, the main challenge is to produce evidence on a high enough frequency to allow a feedback loop between management practices and outcomes.

In recent years we have seen several interesting models emerge for taking evidence into account in both budget and management decisions:

Tiered evidence standards in grant competitions. As the description of the i3 program above illustrates, tiered evi-

dence standards offer a way for an agency to allocate its largest grants to proven interventions while simultaneously investing in the development and evaluation of promising new approaches. In addition to the Department of Education, several other agencies have adopted this tiered approach. HHS has used it for its home visiting program, requiring grantees to allocate at least 75 percent of funds to proven approaches but allowing the remainder to be spent on promising approaches as long as they are rigorously evaluated. HHS also used a tiered evidence approach with its grants for teen-pregnancy prevention. The DOL's Workforce Innovation Fund and Trade Adjustment Assistance Community College and Career Training Program have used a tiered evidence approach, as have the Corporation for National and Community Service's Social Innovation Fund and the U.S. Agency for International Development's Development Innovation Ventures Fund. A federal interagency working group is currently developing a common evidence framework that would make it easier for additional agencies to adopt a tiered funding approach. Even without tiered evidence standards, grant competitions can make evidence a selection criterion to allocate funding to initiatives with the best evidence of success. Formula grants can also make adoption of best practices a requirement for some or all of their funding.

Comparing the performance of different providers. Historically, the federal Head Start program has automatically renewed existing providers. But under new regulations designed to respond to concerns about uneven quality at Head Start sites, providers are awarded five years of funding and then evaluated, with the bottom 10 percent of performers required to compete with other potential providers for funding. This approach of recompeting the worst performers is a good model for types of social services where it is not easy to scale the size of operations up or down. In other cases, a more direct comparison of outcomes for different providers can be used to allocate more program slots to the providers with the greatest performance.

Alternatively, the government can publicize the performance of each provider and then give the individuals who are eligible to be served a voucher to choose among providers. This latter approach is used in the Australian employment program, Jobs Services Australia, an extensive attempt to systematically compare results across a large number of social-service providers. In that program, individuals eligible for assistance are classified into four categories based on characteristics that are predictive of how difficult it will be to find them jobs. A larger payment

is made to service providers for serving harder-to-employ individuals. The Australian government compiles results on re-employment rates and other quality metrics for each provider and publishes a rating (one to four stars) for each.

Another tool is simply to assess provider performance at the end of each contract. Early in the Obama Administration, the Office of Federal Procurement Policy finally achieved the goal of a unified procurement data system, to which all agencies centrally report their assessments of contractor performance at the end of each contract in a way that can be searched by any agency when decisions are being made on new contracts. However, it is unclear what fraction of the assessments are of sufficient quality to be useful.

Comparative cost-effectiveness analysis. In some policy areas where dozens of different interventions have been evaluated in separate research studies, often using different outcome measures and different evaluation methodologies, it can be difficult for policymakers to make sense of all of the evidence, especially when the cost of the interventions varies widely. A review of the evidence by a policy-focused team of experts can help determine the relative cost effectiveness of different approaches. The Washington State legislature created the Washington State Institute for Public Policy (WSIPP) to produce studies of this sort. WSIPP's reviews of the evidence in criminal-justice policy are used by state policymakers nationwide who are seeking to determine which policies are the most cost-effective for reducing recidivism, and the Pew Center on the States is helping to adapt the WSIPP criminal justice benefit-cost model for use in other states.

Spreading successful practices. Another approach to allocating resources based on evidence is to target efforts around spreading proven practices. The New York City CEO Social Innovation Fund project is explicitly aiming at this objective. The project is helping other cities implement programs that have been successful in New York City. The federal Charter Schools Program, which funds successful, high-quality charter schools to expand enrollment or set up additional schools with the same model, is another example. Since 2010, the Department of Education has awarded competitive grants to support the creation of 251 new high-quality charter schools and the expansion of thirty-four existing charter schools. In both of these examples, the same teams that set up the original models are involved in the replications. Although narrow-purpose programs can sometimes be excessively rigid

and stifle innovation, when they support proven innovations they can be another way to spread successful practices.[9]

Simply cataloging successful practices can help them to spread. The Department of Education's What Works Clearinghouse, the Department of Justice's CrimeSolutions.gov, Substance Abuse and Mental Health Services Administration's National Registry of Evidenced-Based Programs and Practices, and DOL's Clearinghouse of Labor Evaluation and Research (CLEAR) are helpful steps towards making evidence on what works more available. Private-sector efforts such as the Coalition for Evidence-Based Policy's Social Programs that Work list are similarly useful.

Eliminating funding for ineffective programs. One of the most direct ways to use evidence in allocating resources is to stop funding things that are found not to work. An example of this occurred early in the Obama Administration when federal funding for the Even Start program, which provides literacy training to parents and children in homes where English is not the first language, was eliminated after a randomized control trial found that outcomes were no better for those served by the program than for those in the control group. This decision was controversial. Supporters of the program argued that it had improved subsequent to the period covered by the evaluation. More broadly, the Bush Administration's Performance Assessment Rating Tool (PART) categorized 1,015 federal programs (98 percent of the total) as either "effective," "moderately effective," "adequate," "ineffective," or "results not demonstrated." PART scores were often used by the Bush Administration to justify budget decisions, but Congress questioned the credibility of the ratings. The Obama Administration discontinued PART assessments to refocus performance assessment as a management tool to track performance trends and improve performance in real time, rather than as a binary assessment of whether programs are effective.

PerformanceStat efforts to improve performance trends. PerformanceStat is a leadership strategy in which the chief executive, agency head, or a top deputy holds regular meetings with key staff to review up-to-date data on progress toward achieving performance goals in an ongoing effort to improve performance.[10] The PerformanceStat movement began at the local level with New York City's Compstat and NYCStat, Baltimore's CitiStat, and Maryland's StateStat. Use of these techniques corresponded with falling crime in New York City, demolition of blighted properties in New Orleans, improved

Medicaid claims processing in Los Angeles County, and reduced homelessness in Washington, DC (Behn 2013). In the last two years, this approach has blossomed within the federal government as well, with nearly two-dozen agencies using the strategy.[11]

4. Making purposeful efforts to target improved outcomes for particular populations

Most government social service funding today is dedicated to purchasing slots in programs. Programs are managed to deliver a defined set of services to a fixed number of people rather than to achieve any particular outcome. Multiple government programs often provide "stove-piped" assistance or services to a given individual, with none of them accountable for getting the individual to achieve success. Rather than managing programs based on the quantity of services provided, government agencies need to track outcomes for specific target populations and manage their programs to achieve outcome goals. Successful population-focused efforts will generally require extensive collaboration with many non-governmental community partners, including businesses, non-profit service providers, and philanthropies.

Taking a whole-population approach to providing services. The Strive Partnership described above is perhaps the clearest example of an effort to define a target population and coordinate services in a strategic way to make sure everyone in the population receives the services they need to succeed. Community collaboratives are said to have reduced teen pregnancy in Milwaukee, gun violence in Boston, high school dropout rates in Philadelphia, and violent crime in Atlanta.[12]

Client matching. An important component of population-focused efforts is matching the right services to the right individuals. Some services work better for particular population subgroups than for others, and some are cost-effective only for certain population subsets. As part of its Work for Success employment initiative for the formerly incarcerated, New York State is developing a client matching system to connect prisoners with the appropriate set of services upon release from prison.

Setting outcome-focused goals and managing to them. The U.S. Department of the Interior's initiative to reduce violent crimes in Indian communities illustrates the power of goal-setting, frequent and population-focused performance measurement, replication demonstrations, and scaling. Over

three years, violent crime has fallen by 55 percent in four reservations. The agency is now working to expand the initiative to two additional reservations.[13]

5. Spurring innovation and aligning incentives through cross-sector and community-based collaboration

Much of the work necessary to enable evidence-based policymaking by the federal government needs to occur either outside of government, through efforts of service providers, social entrepreneurs, and philanthropists, or within government at state and local levels. Creating structures that promote productive collaboration and align incentives so that the right initiatives are developed, evaluated, and ultimately funded is therefore essential. The federal government has recently created six such structures:

The Social Innovation Fund. The Corporation for National and Community Service's (CNCS) Social Innovation Fund (SIF) makes grants to grant-making intermediaries that match the federal awards dollar for dollar with funds raised from other sources. Each intermediary then runs a competitive process to make grants to expand community-based non-profits with evidence of strong results. These grantees also are required to match the grants they receive. To date, the SIF has made grants to twenty intermediaries, which have selected 197 non-profit sub-grantees. CNCS reports that the $138 million in federal grants has yielded $350 million in non-federal cash match commitments. Each grant is expected to be rigorously evaluated, though given the relatively small size of the grants and the early stage of development of some of the organizations receiving grants, it is unclear how many impact evaluations will actually emerge from the initiative.

Waivers and performance partnerships. Disadvantaged communities and individuals are often the recipients of services from multiple federal programs spanning several federal agencies. In many cases this funding is not well coordinated, and no one is responsible for whether or not the combination of funding produces successful outcomes. In his FY 2013 budget, President Obama proposed to allocate $200 million in existing funding to Performance Partnerships in which states and localities would be given the flexibility to propose better ways to combine federal resources in exchange for greater accountability for results. The initial partnership proposals were targeted at the areas of disconnected youth and neighborhood revitalization. Several states gave examples of

the projects they would like the flexibility to undertake. For example, Iowa was interested in developing a coordinated approach to providing services to high-risk youth—those involved in the child welfare, juvenile justice, mental health, and vocational rehabilitation systems.

More broadly, federal waiver authority is an important tool for developing innovative approaches. By letting state and local governments test and evaluate new strategies, they can demonstrate solutions that can then be spread more broadly through legislative changes. Dozens of welfare-reform demonstration projects preceded the 1996 legislative changes. A more recent example is the HHS Title IV-E waivers for child-welfare funding. Nine states are currently testing new strategies for serving children and families involved in the child-welfare system. To be successful, a waiver program must not only give states the flexibility to try new solutions, it must also incorporate rigorous evaluation so that informed decisions can be made about whether to spread the new strategies to additional jurisdictions.

The SSI PROMISE program, a joint project of four federal agencies (Department of Education, Social Security Administration, HHS, and DOL), is another good example. It will make awards to states that present promising strategies to produce better outcomes for children who receive Supplemental Security Income (SSI). In order to align incentives between the federal government and its state partners, the program will make incentive payments to states that are able to reduce federal SSI costs by helping SSI youth make successful transitions to postsecondary education and adult employment.

Place-based strategies. An important category of initiatives in which state and local governments are being given the flexibility to combine federal funding streams in creative ways are place-based initiatives. These include the Department of Housing and Urban Development's (HUD) Choice Neighborhood Program, which invests in improvements in housing, schools, transportation, access to employment, and other public needs in a coordinated effort to transform distressed neighborhoods; the Department of Education's Promise Neighborhoods program, which funds communities engaging in "cradle-to-career" educational initiatives; and the Sustainable Communities Initiative, a joint effort of HUD, Environmental Protection Agency, and the Department of Transportation, which supports regional planning and development efforts.

Using federal grants to facilitate local partnerships. The availability of a federal grant competition can serve as an action-forcing event that enables government, philanthropic, and private partners at local levels to come together and produce an innovative plan to improve outcomes. Perhaps the most successful example of this has been the Department of Education's Race to the Top Fund, which offered grants to states for educational reform efforts and required all of the necessary stakeholders to be partners in the state proposals. What was striking about the Race to the Top effort was that many of the states that were runner-ups for the federal funding nonetheless decided to go ahead and implement the reform plans that they had developed in the response to the Race to the Top competition. As with waiver authority, the key to producing learning from initiatives like this is not simply to promote innovation but to also rigorously assess the results of the different strategies so that these state initiatives can truly function as "laboratories of democracy."

A federal–state partnership fund. Many federally funded social programs are administered by state and local governments, and efforts to improve the administration of these programs require cooperative efforts across levels of government. In 2010, Congress authorized the Partnership Fund for Program Integrity Innovation, managed by OMB, which received $32.5 million to be spent over a multi-year period. OMB set up a Collaborative Forum with representatives from state and local governments, non-profits, federal agencies, and participation by the public to come up with ideas for pilot projects that could demonstrate best practices and inform policy decisions. The OMB director approves pilot concepts and then transfers funds for the pilot to a lead federal agency, which implements the pilots in collaboration with state and local partners. As one example, the Department of Justice is working with three state and local juvenile justice agencies to develop a cost-effectiveness scorecard. By showing the cost effectiveness of different evidence-based juvenile justice interventions, the scorecard will help program leaders make better service contracting decisions and also help frontline service providers make better decisions about particular interventions for youth.[14]

Social Impact Bonds. Pay for Success contracts using Social Impact Bonds (SIBs) are a particularly novel approach to financing social programs. Under the most common SIB model, the government contracts with a private sector intermediary to obtain social services. The government pays the intermediary entirely or almost entirely based upon achievement of performance targets. Performance is rigorously mea-

sured by comparing the outcomes of individuals referred to the service provider relative to the outcomes of a comparison or control group that is not offered the services. If the intermediary fails to achieve the minimum target, the government does not pay. Payments typically rise for performance that exceeds the minimum target, up to an agreed-upon maximum payment level. The intermediary obtains operating funds by raising capital from independent commercial or philanthropic investors who provide upfront capital in exchange for a share of the government payments that become available if the performance targets are met. The intermediary uses these operating funds to contract with one or more service providers to deliver the interventions necessary to meet the performance targets. Figure 1 illustrates SIB cash flows.

The first U.S. SIB is a New York City initiative that is providing services to sixteen- to eighteen-year-olds who are jailed at Rikers Island with the aim of reducing recidivism and related budgetary and social costs. Services are being delivered to approximately 3,000 adolescent males per year, from September 2012 to August 2015. MDRC is serving as the intermediary, overseeing day-to-day implementation of the project and managing the two non-profit service providers that are delivering the intervention. Goldman Sachs is funding the project's delivery and operations through a $9.6 million loan to MDRC. The city will make payments that range from $4.8 million if recidivism is reduced by 8.5 percent to $11.7 million if recidivism is reduced by 20 percent. Bloomberg Philanthropies is guaranteeing the first $7.2 million of loan repayment. Massachusetts and New York State are also developing SIBs.

Although SIBs are still a highly experimental approach to financing social programs, they are also highly promising because they directly address all five of the challenges described at the top of this section. Because they shift the risk of failure away from taxpayers and on to the private sector, they enable governments to try new and innovative solutions. Because rigorous real-time impact assessment is an essential component of a SIB, SIBs generate additional evidence on program impacts. Because an intervention that results in a successful SIB will likely be scaled up while an unsuccessful one will not receive further funding, the evidence from a SIB project is highly likely to be influential in budget decisions. Because SIBs assign specific populations to service providers and hold them accountable for outcomes, they avoid the traditional fragmented and slot-based approach to service provision. Because SIBs are multi-party contracts that combine government, service providers, and private investors in a multi-year effort to achieve performance goals, they promote cross-sector collaboration.

Figure 1

Cash Flows in a Social Impact Bond
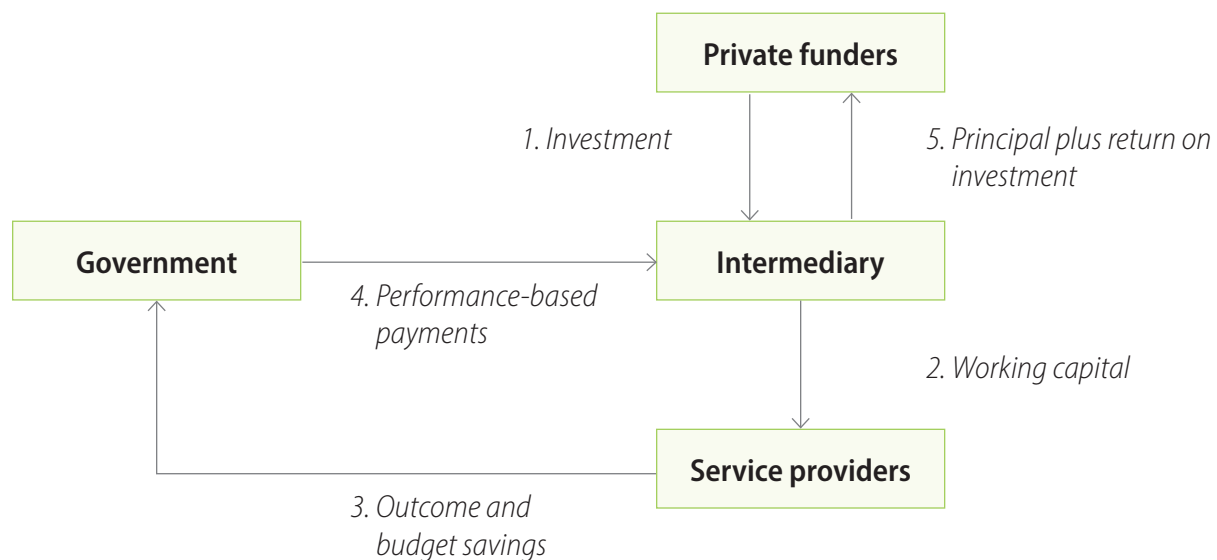
Federal Actions to Support Evidence-Based Policymaking

These evidence-based practices described above are likely to spread on their own as more federal agencies and state and local governments become aware of them. At the federal level, they will spread faster and be more likely to survive changes in administrations if Congress takes action to support their adoption. There are a variety of both legislative and administrative actions that would help spread these practices.

Provide funding authority for evaluations

All agencies that administer social programs should receive authority, similar to that currently provided to the DOL, to reserve a portion of program spending to fund program evaluations. The DOL's authority is better crafted than similar authority that has been provided to other agencies because it gives the agency the flexibility to use the evaluation funds to evaluate the highest-priority initiatives rather than tying the funds to evaluation of specific programs.[15] The agency authority should be provided only to agencies that establish a chief evaluation officer position or have a similar office that is dedicated to producing independent and rigorous evaluation evidence about the agency's programs.

In addition, a small evaluation fund should be provided to OMB for cross-agency evaluation initiatives. The administrative structure could be similar to that of the Partnership Fund described above and would enable coordinated, cross-agency evaluation planning for related programs. OMB would be encouraged to consult with Appropriations Committee leadership in deciding which programs to evaluate and to demonstrate the potential benefits of planned studies.

Expand the use of tiered evidence standards in grant competitions

In high-priority policy areas where sufficient evaluation evidence is available, agencies should be encouraged to use tiered evidence standards in grant competitions. Structuring a grant competition in this way ensures that the largest share of federal funding goes to practices that have been shown to be effective, while also investing smaller amounts of funds in developing evidence about promising but not yet proven approaches. Tiered evidence grant competitions are expensive to administer because they generally require agencies to recruit outside evaluation consultants to review the evidence base of each proposal the agency receives. Thus, they are only worth using for large, high-priority grant competitions. In addition, tiered standards are useful only in policy areas where there is a sufficiently wide evidence base for it to be possible to distinguish among different levels of evidence quality.

Reserve a portion of formula funding for proven practices

The majority of federal social spending is administered through formula funding to state and local governments rather than through grant competitions. Recent advances in evidence-based practices have had very limited impact on spending that is allocated via formula funding. Today we do not have a sufficient evidence base for most activities funded via formula funding for it to make sense to restrict the use of these funds to evidence-based practices. So, a portion of these funds should be used to build the evidence base. Then, over time, as the

evidence base becomes strong enough, formula funding provided to state and local governments should begin to require that a portion of such funds be spent on programs that have been proven effective. Depending on the policy area and the amount of evidence that exists, this requirement could start by stipulating that 1 percent of such funds be allocated to proven practices, then have the requirement rise over time to 5 percent.

Direct OMB to submit an annual report on evidence-based techniques and practices

The 2010 Government Performance and Results Modernization Act was important because it codified the new agency performance reporting framework in a way that makes the framework more likely to endure. Similarly, Congress could take steps to show that it supports the recent executive branch evidenced-based initiatives. At a minimum, a congressional committee could have an annual hearing at which it invites the OMB director and some agency representatives to describe their progress in using evidence-based techniques. Congress could also require OMB to produce an annual report listing evidence-based practices by agency and reporting the percentage of each agency's grants that were made using evidence-based techniques. Many such reporting requirements quickly devolve into time-wasting compliance exercises. Therefore, if such a report is requested, the requirement for producing it should be sunsetted after five years and apply only if Congress also rescinds several existing OMB reporting requirements. Another option would be for members of Congress to ask OMB leadership at a hearing to commit to devoting a few pages each year in the Analytic Perspectives volume of the budget to an update on progress toward implementing evidence-based funding practices.

Task CBO or GAO with cost-effectiveness reports

The Budget and Appropriation Committees should jointly identify five priority policy areas each year and ask either CBO or GAO to produce a report in each area identifying (1) what evidence currently exists about the outcomes being produced by federal spending in that area, (2) what is known about the relative cost effectiveness of different strategies in the policy area, (3) what promising strategies exist that have not been rigorously evaluated, and (4) where promising strategies are lacking, and for these, what funding strategies and incentives

could be used to spur development of new solutions. Illustrative areas for such reports include job training, early childhood education, dropout prevention, and asthma and diabetes prevention, among many others. It would likely require an increase in the agency budget of at least $3 million per year to carry out this work, but this modest investment would enable Congress to target resources more effectively. Once the agency has gone through the effort of producing an initial report in each area, it should be updated every two or three years.

Compile federal program evaluations into a comprehensive website

In 2010, OMB issued guidance requiring agencies to make information readily available online about all federal program impact evaluations that are planned or already underway. This effort was intended to be analogous to the HHS clinical trial and results data bank (ClinicalTrials.gov) that aims to prevent drug companies from hiding negative trial results. Although some agencies complied by posting their evaluations on their agency websites, OMB never moved forward to compile all of the agency information into a comprehensive federal evaluation website. In addition to preventing the suppression of results, a comprehensive federal evaluation website would help Congress and the public understand which programs have been evaluated, which ones haven't, and what the results showed. A challenge with government transparency efforts is that they often provide a large quantity of information without the interpretive context necessary for consumers of the information to use it. Thus, any federal effort of this sort will likely need to be accompanied by a parallel effort by organizations like the Coalition for Evidence-Based Policy to interpret the information.

Make administrative data more accessible for measuring outcomes

Today evaluation studies that wish to use government administrative data such as earnings histories or records of benefit receipt to measure program outcomes typically need to go through extensive one-off negotiations to arrange access to data. Establishing a standardized way for evaluators to access common federal data sources would increase the number of evaluations that can occur.

A More-Focused Approach

If the evidence-based approaches to policymaking described above spread widely, we will achieve better outcomes with the government spending we do, we will be able to stop spending money on programs that don't work, and, most importantly, we will develop new, more effective, strategies.

But if our goal is to make significant progress in addressing our most serious social problems, simply expanding the use of these approaches is unlikely to be enough to produce the solutions we require. We need to supplement the wide diffusion of these practices with a more-focused approach that aims to find solutions for specific populations.

The Ten-Year Challenge: Solve ten problems in ten places in ten years

Congress and the President should work together to identify ten social problems where it is a national priority to find solutions. Examples could include: reducing recidivism among ex-offenders, raising third-grade 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, and helping developmentally-disabled youth make successful transitions into the adult workforce, among many others. All of 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—one hundred communities overall—in an effort to transform outcomes for the specific population within five to ten years. A single agency would be the granting agency for each initiative, though many of the initiatives will require extensive cross-agency collaboration. The granting agency would first issue planning grants of roughly $250,000 each to several dozen communities that were interested in putting together proposals. Then ten communities would be selected for 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 nationally if it is successful. As with the Department of Education's Race to the Top program, we would expect many of the communities that apply for the grants and fail to receive them to nonetheless go ahead and implement their proposals.

Funding for the initiatives could come from repurposing existing grant and formula funding, or from a congressional appropriation. In addition, the federal government would waive program rules, so that these communities could take existing funding streams and use them in more flexible ways to fundamentally redesign systems. Communities would be expected to supplement the federal assistance with resources raised locally.

Although it would be terrific if all one hundred projects were successful, the real goal would be to discover two or three transformative approaches for each policy problem— solutions that could then be taken nationwide. The appendix gives further details on this proposal.

Establish specific Pay for Success projects in early childhood and other strategic areas

The federal government should establish specific Pay for Success projects in areas where state and local activity has the potential to achieve important federal policy objectives or produce significant federal budget savings. The highest priority would be for early childhood interventions including home visiting, early learning, and preschool. Evidence of success in early childhood programs is strong, but state and local governments are finding it challenging to establish Pay for Success initiatives in this area because a significant portion of the financial benefit of these projects is likely to accrue to the federal government in the form of lower Medicaid and transfer program spending and higher federal tax revenue. Federal involvement could enable the economics of these projects to

work. The appendix contains a more detailed early childhood Pay for Success proposal. As another example, if a state government can set up a Pay for Success project that helps disabled individuals return to work rather than apply for federal disability insurance benefits, the federal government should offer to reimburse the state government for any success payments it makes in the project.

The federal government should also consider a broader role as a strategic partner with state and local governments in establishing SIB projects. The rationale for a broader federal role is that if one state or local government discovers a solution to a social problem, it will have tremendous value to the nation as it gets scaled nationwide. As mentioned above, a process through which the federal government subsidizes learning at the state level can help overcome the underinvestment in learning that is likely to occur when individual jurisdictions cannot capture the full value of their discoveries. The federal subsidy could come in the form of an extra prize-like payment to investors in successful SIB projects. Alternatively, the federal government could "backstop" a portion of the losses in unsuccessful projects.

Conclusion

During the past decade, governments have shown tremendous creativity in coming up with new approaches to foster experimentation and learning and to allocate spending to programs with the best evidence base. As the results of the programs funded through the new mechanisms emerge, we will know a lot more than we know now about what works and what doesn't. If these new evidence-based practices continue to spread, the amount of learning will accelerate and the country will reap both fiscal and economic benefits.

But without a strategic effort to develop approaches that target entire populations of at-risk individuals in specific communities, it is unlikely that we will move the dial on our most pressing social problems. What is needed is a decade in which we make enough serious attempts at developing scalable solutions that, even if the majority of them fail, we still emerge with a set of proven solutions that work.

Appendix

Appendix A: The Ten-Year Challenge: Solve ten problems in ten places in ten years

Motivation

The modern era of social-policy development and evaluation dates back to the late 1960s when large data sets 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 many 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. Too often, the focus is on individual programs or service providers, and not enough focus is on systems or outcomes across an entire population in a community. We rarely measure outcomes and, even when we do, we generally measure them too infrequently and fail to disaggregate them by service provider. We contract for services in a fragmented way, with multiple providers delivering services to the same individual, but no one held accountable for whether the individual achieves a successful overall outcome. We rarely track the entire population of individuals in need of service to determine who is failing to receive services and whether we are doing a good job of connecting the right people to the right services. We fail to coordinate well among the various parties in the community—service providers, employers, governments, philanthropic organizations—whose actions jointly determine outcomes for target populations.

The Ten-Year Challenge is motivated by the hypothesis that data-driven, outcome-focused initiatives that aim to produce community-level collaboration to achieve better results for entire populations of high-priority individuals offer the best chance of producing transformative solutions to our most persistent social problems. The key ingredient that makes this the right time to undertake such an initiative is the increasing availability of outcome data from government data systems that can be used to monitor results and evaluate program im-

pacts at a high enough frequency to allow feedback loops between changes in practices on the ground and results.

The proposal

Congress and the President should work together to identify ten social problems where it is a national priority to find solutions. Examples could include: 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, reconnecting disconnected youth ages sixteen to twenty-four to school or work, retraining individuals who have been unemployed for more than six months, increasing the rate of community college completion, reducing obesity-triggered diabetes, eliminating chronic and/or family homelessness, and helping developmentally-disabled youth make successful transitions into the adult workforce, among many others. All of 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—one hundred communities overall—in an effort to transform outcomes for the specific population within five to ten years. A single agency would be the granting agency for each initiative, though many of them will require cross-agency collaboration. The granting agency would first issue planning grants of $250,000 each to several dozen communities that demonstrate a commitment to cross-sector partnerships and organizational capacity to develop competitive proposals. Final awards would be made to those communities that successfully propose a data-driven, collaborative approach to transform delivery of services to achieve measurable improvements in outcomes for specific cohorts of individuals.

Average-sized projects would spend $10 million a year on services and serve approximately 1,000 to 2,000 individuals (with flexibility depending on the nature of the intervention and the size of the community and of the target population). The federal grants would cover one-third of the cost of service provision, with state and local governments providing one-third, and private community partners covering the remaining third. The federal government would also cover up to $1 million per year

per project of evaluation expenses. In total, the federal share of the initiative would cost approximately $400 million per year. Funding for the initiative could come from repurposing existing grant and formula funding, or from a congressional appropriation. In addition, the federal government would waive program rules as appropriate, so that communities could take existing funding streams and use them in more flexible ways to fundamentally redesign systems.[16]

As with the Department of Education's Race to the Top program, we would expect many of the communities that apply for the grants and fail to receive them to nonetheless go ahead and implement their proposals. Although it would be terrific if all one hundred initiatives 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 taken to scale nationwide.

In selecting among final proposals, criteria should include:

- Likelihood, based on existing evidence, the project's design, and the level of coordinated commitment of community partners, that the project will make significant progress in addressing the target social problem

- Potential for the project to produce rigorous evidence that would add to what we currently know about the effectiveness of particular strategies

- Extent to which the proposed strategies represent significant advances over current practices

- Potential for scaling the project, if successful, both within the given state and to other states, and to other similar populations

Additional background

Why ten problems? While it would be easy to come up with a list of several dozen social problems for which solutions are needed, there is a limit to how many projects of this complexity federal agencies can administer effectively. With ten grant competitions, the major agencies that administer social programs—Department of Education, Department of Justice, DOL, HHS, the Social Security Administration—could each administer one or two projects (HHS could potentially ad-

minister two health-related projects and two human service related projects.) Tackling fewer than ten would be a significant lost opportunity to make progress in addressing our most important social problems.

Why ten communities? To find transformative solutions, we will need to test many different approaches. A reasonable conjecture is that one-third of communities will develop highly successful strategies, another third will exhibit results that are decent but not strong enough to merit replication, and a final third will fail to make any significant progress at all. By focusing on each problem across ten communities, we will have a good chance of finding two to three solutions to each problem that can be scaled nationwide. Also, by involving one hundred communities in this initiative, the federal government will help spread the adoption of results-focused, community-level collaborations in a way that is likely to produce benefits beyond the specific targeted policy areas.

Why ten years? The process of coming up with transformative solutions will often be iterative, requiring a combination of continuous improvement in processes and more fundamental midcourse corrections. Successful social entrepreneurs often talk of failing several times before finding a successful model. It would be a shame to pay only for the failure phase and never get to the success phase. Also, some interventions (early childhood, for example) involve a several-year delay between the time the interventions are delivered and the time when results can be observed.

Why target entire cohorts of individuals? The goal is to get both government and service providers to focus on achieving better outcomes for entire cohorts of high-risk individuals in specific communities, such as all low-income preschoolers in high-poverty neighborhoods or all of the individuals living in public housing. The Ten-Year Challenge seeks to replace the current fragmented approach to service delivery—where service providers serve individuals who are referred to them on an ad hoc basis, where no one is accountable for the overall well-being of individuals who are often served by multiple service providers, and where no one tracks how many individuals fail to receive needed services—with a new approach that identifies all of the individuals in a target population and comes up with a coordinated way to target community resources to produce successful, comprehensive outcomes for as many people as possible.

What is the size of the geographic region or community where services can be delivered? This will depend on the policy area. A target population could be statewide—for example, all youth in a state who are aging out of the juvenile justice or foster-care system in a given year. Or, it could be much more geographically targeted—for example, all preschool-age youth in a particular neighborhood. In most cases, interventions will involve 1,000 to 2,000 individuals per year. With fewer individuals, it will be hard to have enough statistical precision to know what the results are. And budgetary resources are unlikely to permit significantly larger groups to be served except in cases in which the cost per person served of the intervention is very low.

How should the ten problems be chosen? The ten problems would be chosen through bipartisan discussions between Congress and the Administration. The aim would be to achieve consensus among members of Congress from the relevant committees and the Administration around a set of social problems to be targeted that could be proposed next February in the President's budget. The problems should be those where there are solutions that show promise and that are priority challenges that will impact the long-term strength of our economy and health of our communities.

Appendix B: An early-childhood Pay for Success initiative

Motivation

There is considerable evidence that some early childhood investments for disadvantaged populations have benefits that significantly exceed their costs (Karoly, Kilburn, and Cannon 2005; Ludwig and Phillips 2007). However, fiscal constraints are making it difficult for state and local governments to adopt or expand these proven early childhood investments even when they know such investments will more than pay for themselves in the long run. Pay for Success financing mechanisms, in which governments agree to pay service providers for results when the results are actually achieved, offer a way to expand investments in early childhood and produce additional evidence on the effectiveness of the interventions.

Several state and local governments are actively considering early childhood Pay for Success projects, but are discovering

that a significant portion of the benefits of these projects is likely to accrue to the federal government in the form of lower Medicaid and transfer-program spending and higher federal tax revenue. Federal involvement could enable the economics of these projects to work.

The proposal

Congress should give the Department of Education and HHS the authority to repurpose a total of up to $125 million of existing funds to support five state or local early-childhood Pay for Success initiatives. These funds would be used to match, on a dollar-for-dollar basis, state or local success-based payments. A typical-sized project would provide $10 million of services per year for four years and serve approximately 1,000 families per year. With a state/local commitment of $25 million and a federal commitment of $25 million, success-based payments of up to $50 million would be possible. Funds repurposed in this way would remain available until expended rather than expiring at the end of the fiscal year.

In selecting projects, the Department of Education and HHS should aim to test the Pay for Success financing approach for a range of early-childhood investments, including home visiting, early learning, and preschool. In selecting among applications, criteria should include:

- Likelihood, based on existing evidence, that the project will achieve its performance goals

- Potential for the project to produce social benefits that significantly exceed its resource costs

- Potential for scaling the project, if successful, both within the given state and to other states

- Potential for the project to produce rigorous evidence that would add to what we currently know about the effectiveness of particular strategies

- Potential for the project to be used as a study platform, even after the project is concluded, to generate evidence of the long-run (e.g., through early adulthood) impacts of early-childhood interventions

The recent DOL "Solicitation for Grant Application for Pay for Success Pilot Projects" provides a useful model for a federal Pay for Success grant competition and contains additional useful selection criteria (U.S. Department of Labor 2012).

Additional background

Early-childhood investments are a good fit for the Pay for Success model because of the strong evidence that preventive spending is this area can lead to both significant social benefits and government savings in the future. However, there are two challenges facing a state or local government that is considering the Pay for Success model in this policy area. First, half or more of government savings produced by early-childhood investments are likely to accrue to the federal government. Second, much of the benefits and government savings come many years after the initial investments are made. For example, Miller (2013) studies the Nurse–Family Partnership home visitation program, a program that costs about $8,580 per family served. Extrapolating from the many randomized controlled trials that have been done on the program, he estimates that total discounted government savings per participating family are $29,605, and that more than half of the benefits ($16,205) accrue to the federal government. He also found that only about a quarter ($7,510) of the total government savings and $2,900 of the state government savings occur in the first four years after services begin.

Rationale for a federal role. The 50-percent match suggested for this initiative is based upon the observation that at least half of the government savings produced by an early-childhood initiative are likely to accrue to the federal government. A further federal subsidy could be justified because evidence about the effectiveness of interventions gathered in one state can be used to improve decision-making in forty-nine other states.

Accounting for long lags between service delivery and results. Early-childhood interventions produce a combination of short-term, medium-term, and long-term benefits. For example, home visiting programs have been shown to produce short-term improvements in maternal and child health that also reduce Medicaid costs. Additional, medium-term benefits such as reduced child maltreatment, increased maternal employment, and reduced need for special-education services in early elementary school years may also occur. Finally,

long-term benefits such as reduced criminal activity during late teenage years and higher adult earnings could also result.

There are limits to the length of Pay for Success contracts. Investors are reluctant to have their money tied up for more than a few years, and long lags between service delivery and the measurement of results can eliminate the opportunity for feedback loops that improve performance (Godeke and Resner 2012). It is likely that the outer bounds of what might be possible would be a philanthropically financed ten-year Pay for Success contract for a prenatal home visiting initiative in which short-term payments were made based upon Medicaid savings and medium-term payments were made based upon kindergarten readiness, third-grade reading scores, and/or special-education utilization. But most early-childhood Pay for Success initiatives will be of three- to six-year durations and will need to make payments based on a combination of actual short-term benefits and projections of future benefits based upon the shorter-term results. In order to maximize what we learn from these pilot initiatives and establish the evidence base necessary to decide whether to expand and replicate the pilots, these shorter-duration initiatives should include measurement components that extend well beyond the period during which the project is operating—ideally until the children reach early adulthood.

Scale. The proposal recommends interventions that spend approximately $10 million per year on services for four years. This reflects lessons learned in working with states to design some of the initial U.S. Pay for Success initiatives (Liebman and Sellman 2013; Azemati et al. forthcoming). First, sample sizes of at least 500 new individuals served per year for several years are necessary to achieve a sufficient degree of statistical precision for most research designs. Second, the projects are complicated to set up and are therefore worth doing only if at least several hundred people per year are going to be served. Third, projects should last for long enough so that mid-course corrections can be made if necessary, and so that early results become available in time to make decisions about whether to renew or expand the program. In most cases, a total of $20 million of services (i.e., $5 million per year for four years) will be the smallest Pay for Success project worth embarking on, and $40 million projects (i.e., $10 million per year for four years or $6.6 million per year for six years) will be more likely to demonstrate the full potential of the Pay for Success model.

Author

Acknowledgments

Jeffrey B. Liebman

Malcolm Wiener Professor of Public Policy, John F. Kennedy School of Government, Harvard University

Jeffrey B. Liebman is the Malcolm Wiener Professor of Public Policy at the Harvard Kennedy School where he teaches courses in social policy, public sector economics, and American economic policy. In his research, he studies tax and budget policy, social insurance, poverty, and income inequality. During the first two years of the Obama Administration, Liebman served at OMB, first as Executive Associate Director and Chief Economist and then as Acting Deputy Director. From 1998 to 1999, Liebman served as Special Assistant to the President for economic policy and coordinated the Clinton Administration's Social Security reform technical working group.

Endnotes

1. The DOL provision allows transfers of up to 0.5 percent of program spending (see section 108 of the general appropriation provisions for DOL in P.L. 112-74). America Achieves has proposed that up to 1 percent of program spending be reserved for evaluations.

2. A companion program, Early Head Start, serves even younger children and has stronger evaluation results.

3. The Head Start National Impact Study compared children attending Head Start to children who in many cases attended other forms of preschool. Thus, the evaluation does not allow an assessment of the impact of Head Start relative to no services. Non-experimental studies have confirmed the finding that short-term test-score impacts fade, and yet have also found significant long-term benefits of Head Start (Garces, Thomas, and Currie 2002; Deming 2009). Ludwig and Phillips (2007) review the evidence and conclude that "the program is likely to generate benefits to participants and society as a whole that are large enough to justify the program's costs."

4. International comparisons of student performance suggest a broader potential for improvement. See America Achieves (2013).

5. Indeed, simply making *any* decisions would be progress as the most common approach to dealing with tight budgets is to avoid making choices and simply enact across-the-board cuts, often leading to underfunding that undermines effectiveness.

6. Greenberg, Michalopoulos, and Robins (2006) study thirty-one evaluations of government training programs for the disadvantaged and conclude that "government-funded training programs rarely produce large effects on earnings, but that these effects are typically larger for adult women than for adult men and are negligible for out-of-school youth." Visher, Winterfield, and Coggeshall (2005) review eight randomized evaluations of employment programs for ex-offenders and conclude that "this group of community employment programs for ex-offenders did not reduce recidivism."

7. The Coalition for Evidence-Based Policy searched across nine broad policy areas and could identify only ten programs that met top-tier evidence standards.

8. See United Nations International Labour Organization (2011) and OECD (2013). The United States leads the world in output per worker. On an output per hour basis, Norway and Luxembourg are the only countries that consistently exhibit higher productivity than the United States.

9. The National Guard Youth ChalleNGe is a good example of an evidence-based narrowly purposed program. A randomized evaluation by MDRC found the program to be highly effective.

10. See Behn (2006).

11. See GAO (2013). The HUDStat website (http://goals.performance.gov/hudstat) provides a good example of how this approach is being used by a federal agency.

12. See Jolin, Schmitz, and Seldon (2012). It is unclear how many of these collaboratives really moved the needle on outcomes because most lacked a rigorous counterfactual.

13. See U.S. Department of the Interior (2013).

14. Metzenbaum (2008) provides a thoughtful analysis of the challenges involved in learning what works in federal programs that are administered with state and local government partners.

15. The Obama Administration's FY 2013 budget and the Senate FY 2013 Appropriations Bill for the DOL, HHS, and the Department of Education both contain language that would give the Department of Education additional flexibility to pool evaluation funds across elementary and secondary education programs so that evaluation dollars could be targeted at answering the highest-priority questions.

16. The Obama Administration's proposed Performance Partnership Pilots for Disconnected Youth provide a useful model for this approach.

References

Advisory Committee on Head Start Research and Evaluation. 2012. *Advisory Committee on Head Start Research and Evaluation: Final Report.* Submitted to the Secretary of the U.S. Department of Health and Human Services. August.

America Achieves. 2013. *Middle Class or Middle of the Pack? What can we learn when benchmarking U.S. schools against the world's best?* April.

Auguste, Byron, Bryan Hancock, and Martha Laboissiere. 2009. "The Economic Cost of the U.S. Education Gap." *McKinsey Quarterly*. June.

Azemati, Hanna, Michael Belinsky, Ryan Gillete, Jeffrey Liebman, Alina Sellman, and Angela Wyse. Forthcoming. "Social Impact Bonds: Lessons Learned So Far." Federal Reserve Bank of San Francisco *Community Development Investment Review*.

Baron, Jon, and Isabel V. Sawhill. 2010. "Federal Programs for Youth: More of the Same Won't Work," *Youth Today* (May).

Barro, Robert J., and Jong-Wha Lee. Forthcoming. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*.

Behn, Robert D. 2006. "The Varieties of CitiStat." *Public Administration Review* 66 (3): 332–340.

———. 2013. *The PerformanceStat Potential: A Leadership Strategy for Producing Results*. Unpublished manuscript.

Belfield, Clive R., Henry M. Levin, and Rachel Rosen. 2012. "The Economics Value of Opportunity Youth." January. Paper in association with Civic Enterprises for the Corporation for National and Community Service.

Cohen, Daniel, and Marcelo Soto. 2007. "Growth and human capital: good data, good results." *Journal of Economic Growth* 12 (1): 51–76.

Congressional Budget Office [CBO]. 2013. *Budget and Economic Outlook.*

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111–34.

Ellwood, David T. 2001. "The Sputtering Labor Force of the 21st Century. Can Social Policy Help?" In *The Roaring Nineties: Can Full Employment Be Sustained?*, edited by Alan Krueger and Robert Solow. New York, NY: Russell Sage Foundation.

Engelhardt, Gary V., and Jonathan Gruber. 2006. "Social Security and the Evolution of Elderly Poverty." In *Public Policy and the Income Distribution*, edited by Alan Auerbach, David Card and John Quigley. New York, NY: Russell Sage Foundation.

Fiester, Leila. 2010. "Early Warning! Why Reading by the End of Third Grade Matters." Annie Casey Foundation, Baltimore, MD.

Finkelstein, Amy, and Robin McKnight. 2008. "What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending." *Journal of Public Economics* 92: 1644–1669.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92 (4): 999–1012.

Godeke, Steven, and Lyel Resner. 2012. "Building a Healthy & Sustainable Social Impact Bond Market: The Investor Landscape." Godeke Consulting, New York, NY.

Government Accountability Office [GAO]. 2013. *Managing for Results: Data-Driven Performance Reviews Show Promise But Agencies Should Explore How to Involve Other Relevant Agencies.* GAO-13-228.

Greenberg, David H., Charles Michalopoulos, and Philip Robins. 2006. "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.

Hanushek, Eric A., and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth* 17 (4): 267–321.

Haskins, Ron, and Jon Baron. 2011. "Building the Connection Between Policy and Evidence: The Obama Evidence-Based Initiatives." In *Using Evidence to Improve Social Policy and Practice*, edited by Ruth Puttick. London: Nesta Trust.

Jolin, Michele, Paul Schmitz, and Willa Seldon. 2012. "Needle Moving Community Collaboratives: A Promising Approach to Addressing America's Biggest Challenges." May. The Bridgespan Group.

Karoly, Lynn A., M. Rebecca Kilburn, and Jill Cannon. 2005. *Early Childhood Interventions: Proven Results, Future Promise.* Santa Monica, CA: RAND Corporation.

Liebman, Jeffrey, and Alina Sellman. 2013. "Social Impact Bonds: A Guide for State and Local Governments." Harvard Kennedy School Social Impact Bond Technical Assistance Lab.

Ludwig, Jens, and Deborah Phillips. 2007. "The Benefits and Costs of Head Start." *Social Policy Report* 21 (3): 3–19.

Mankiw, N. Gregory, David Romer, David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407–437.

Metzenbaum, Shelley H. 2008. "Federal Agencies as Learning Leaders in the Information Age." In *Intergovernmental Management for the 21st Century*, edited by Timothy J. Conlan and Paul L. Posner, 207–242. Washington, DC: Brookings Institution Press.

Meyer, Bruce D., and James X. Sullivan. 2010. "Consumption and Income of the Poor Elderly Since 1960." NBER Retirement Research Center Working Paper NB10-08. National Bureau of Economic Research, Cambridge, MA.

Miller, Ted R. 2013. "Nurse–Family Partnership Home Visitation: Costs, Outcomes, and Return on Investment." Pacific Institute for Research and Evaluation.

Office of Science and Technology Policy. 2012. *Implementation of Federal Prize Authority: Progress Report*. March.

Organisation of Economic Co-operation and Development [OECD]. 2013. OCED Stat Extracts. Available at www.stats.oecd.org (accessed April 5, 2013).

Overholser, George. 2011. "George Overholser on Pay For Success." *Tactical Philanthropy*. May 18. Available at http://www.tacticalphilanthropy.com/2011/05/george-overholser-on-pay-for-success/.

United Nations International Labour Organization. 2011. *Key Indicators of the Labour Market, Seventh Edition.*

U.S. Department of the Interior. 2013. "Department of the Interior Agency Priority Goals: Reduce violent crime in Indian communities." Available at http://goals.performance.gov/goal_detail/DOI/381.

U.S. Department of Labor. 2012. "Notice of Availability of Funds and Solicitation for Grant Applications for Pay for Success Pilot Projects." Available at http://www.doleta.gov/grants/pdf/pfs_sga_dfa_py_11_13.pdf.

Visher, Christy A., Laura Winterfield, and Mark B. Coggeshall. 2005. "Ex-offender employment programs and recidivism: A meta-analysis." *Journal of Experimental Criminology* 1 (3): 295–316.

Building on Recent Advances in Evidence-Based Policymaking

Summary of findings

Federal, state, and local governments spend billions of dollars each year addressing serious social problems from recidivism to school readiness and obesity to workforce development. For most of these problems, however, we have either failed to develop effective policy solutions, failed to demonstrate which solutions work, or failed to scale widely the solutions that do work. As a result, we are not making enough progress toward combating many of America's most pressing social challenges, resulting in negative outcomes for the affected individuals and for the U.S. economy.

Given mounting fiscal pressures, it is imperative to produce more value with each dollar that the government spends. This means reallocating funds from less-effective programs to more-effective programs and developing new, more-productive solutions.

Over the past several years, governments at all levels have developed a set of new approaches that simultaneously make better use of taxpayer dollars and speed up the pace of innovation. Jeffrey Liebman describes these new strategies and shows how they overcome the barriers that have in the past prevented governments from implementing an evidence-based approach to policymaking.

In addition to outlining these new strategies, Liebman describes several specific steps that can be taken to make federal spending more effective. Some of these steps can be implemented immediately. For instance, policymakers can make greater use of tiered evidence standards that restrict the largest grants to those programs with strong evidence of effectiveness, while also making smaller investments in promising, but not yet proven, approaches. He also argues that all agencies that administer social programs should be able to reserve a portion of their budgets to fund evaluation of current programs and new approaches.

Liebman also proposes a Ten-Year Challenge, a program in the mold of Race to the Top, where the federal government would fund communities to develop innovative, evidence-based solutions to a variety of social problems. And he suggests expanded use of Pay for Success contracts, where the government pays for social services based on the results that are achieved rather than on the quantity of services that are delivered.

Fast facts

- There is surprisingly little evidence available on the effectiveness of government spending on important social programs. Performance of these programs is rarely assessed, and the measurements that do exist seldom focus on how these policies improve outcomes for those affected.

- The lack of evidence-based practices has profound economic, fiscal, and social consequences. The workforce is less productive than it could be, and individuals are adversely affected by continuing social problems. In the current fiscal climate, policymakers are facing pressure to find solutions to these problems while simultaneously reducing the amount government spends on them.

- Making greater use of evidence in policymaking is critical. Drawing on recent government innovations, Liebman describes five strategies that lawmakers and government agencies can use to improve the effectiveness of government spending:

  1. Subsidize learning and experimentation so that new solutions are developed

  2. Increase the amount of evidence on the effectiveness of existing and potential new programs

  3. Make greater use of evidence in budget and management decisions

  4. Make purposeful efforts to target improved outcomes for particular populations

  5. Spur innovation and align incentives through cross-sector and community-based collaborations

- Liebman proposes a Ten-Year Challenge that would help put these strategies into practice. Under this program, the federal government would identify ten pressing social-policy challenges. Then, through a grant competition, ten communities would be selected for each challenge, where each community would be expected to make meaningful and demonstrable progress on addressing its challenge over the next decade. The goal of the program would be to discover two or three transformative approaches for each policy problem—solutions that could then be taken nationwide.

- Liebman also recommends that the federal government make greater use of Pay for Success contracts, and he identifies early-childhood investments as the top priority for a Pay for Success initiative.