

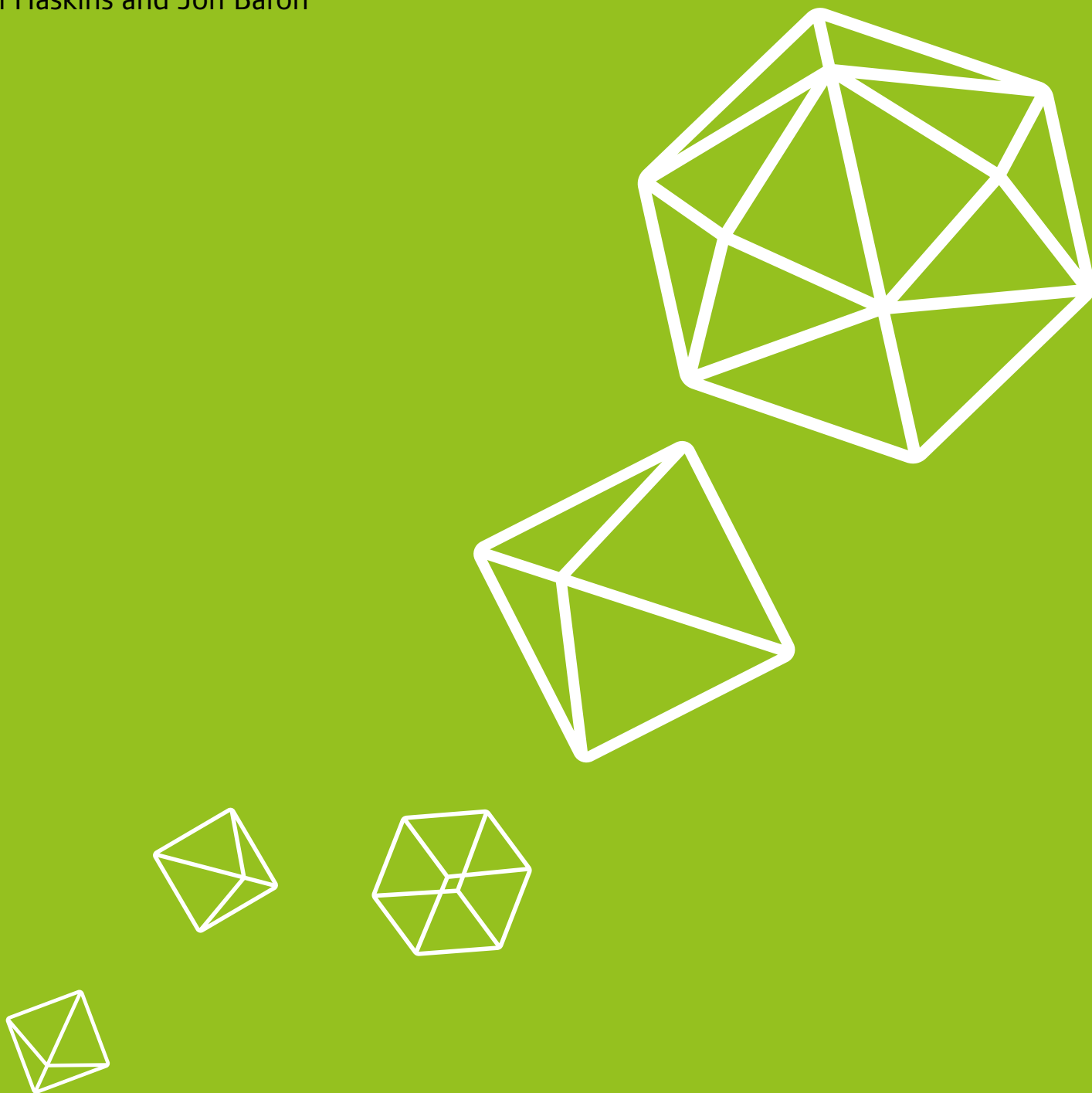
September 2011

NESTA Making
Innovation
Flourish

Building the Connection between Policy and Evidence

The Obama evidence-based initiatives

Ron Haskins and Jon Baron



Building the Connection between Policy and Evidence

The Obama evidence-based initiatives

Foreword

We commissioned this report as part of a series to explore the different approaches being undertaken to improve the effectiveness of social policy and practice. As this report clearly shows, President Obama's evidence-based initiatives are transforming the generation and use of evidence by the US Government, heralding a new chapter in the US evidence agenda. Although the policy context differs, we can learn a lot from their experiences here in the UK.

At NESTA, we are keen to forge connections with a range of organisations to explore how we can strengthen the evidence base for decision making. To co-ordinate these efforts and really drive momentum for change, we are developing a UK Alliance for Evidence. As part of this endeavour, we would like to solicit a range of views and learn from different approaches to developing effective social programmes and policies, helping to improve the effectiveness of our public services and help tackle the challenges facing society today.

As ever, we welcome your views and input.

Ruth Puttick
Policy Advisor, Public and Social Innovation, NESTA

September 2011

NESTA is the UK's foremost independent expert on how innovation can solve some of the country's major economic and social challenges. Its work is enabled by an endowment, funded by the National Lottery, and it operates at no cost to the government or taxpayer.

NESTA is a world leader in its field and carries out its work through a blend of experimental programmes, analytical research and investment in early-stage companies. www.nesta.org.uk

Contents

Building the Connection between Policy and Evidence The Obama evidence-based initiatives

Introduction	4
<hr/>	
President Obama’s evidence-based initiatives	6
<hr/>	
Program areas addressed by the initiatives	11
<hr/>	
The promise of the Obama initiatives	14
<hr/>	
Problems lurk and opportunities abound	15
<hr/>	
Conclusion	17

Building the Connection between Policy and Evidence

The Obama evidence-based initiatives

1. Introduction

There is a growing belief in both the US and the UK that intervention programs addressed to domestic social problems can be greatly improved if policymakers and managers will support programs shown by scientific evidence to produce impacts. Since his inauguration in 2009, President Barack Obama and his administration have developed and are now implementing the most extensive evidence-based initiatives in US history. The purpose of this paper is to trace the evolution of these initiatives and to examine both their promise and problems.

Muddling through vs. rational policymaking

In 1971, Alice Rivlin published a seminal book on decision making entitled *Systematic Thinking for Social Action*.¹ She identified four ‘propositions’ that can be taken as a reasonable summary of the basic elements of what is often referred to as rational decision making. They are:

- Define the problem.
- Figure out who would be helped by a specific program attacking the problem and by how much.
- Systematically compare the benefits and costs of different possible programs.
- Figure out how to produce more effective social programs.²

Rivlin believed that at the time she was writing, economists, statisticians, and other analysts had made good progress on most of the steps in this approach to rational decision making, but that little progress had been made in determining the benefits of particular programs.

A much more skeptical view of the potential for rational, evidence-based policymaking can be seen in the classic 1959 article by Charles Lindblom on making decisions by “*muddling through*.”³ Lindblom argued that no program administrator could actually follow the rational decision making model because the demands on knowledge required to compare all alternative programs are too large, the effects of most programs are not known with any confidence, and not enough time is usually available to perform elaborate analyses before a decision must be made. Thus the choice set faced by managers is limited to incremental adjustments in current policy and practice, and the most important factor in policy choice is usually reaching consensus on a particular alternative. Lindblom argued that this process of what he called “*successive limited comparisons*” among alternatives not radically different from the status quo – or more famously, “*muddling through*” – was both a better description of how policy actually is made and a more practical guide to action than the rational approach.

Our view is that the dichotomy between the rational decision making approach and the muddling through approach is a false one. Policymaking inevitably involves political constraints on choices as well as limitations on evidence and time. But that does not mean there is no evidence available, or

1. Rivlin, A. (1971) ‘Systematic Thinking for Social Action.’ Washington, DC: Brookings.
2. Rivlin, pp. 6-8.
3. Lindblom, C.E. (1959) The Science of ‘Muddling Through’. ‘Public Administration Review.’ 19, (2): 79-88.

that policymakers should ignore the evidence that does exist or fail to devote resources to obtain better evidence. Indeed, Rivlin argued that the case for “systematic analysis” was strong and had been well made, even by 1971, and that “hardly anyone explicitly favors a return to muddling through.”⁴ Rivlin also held that the key challenge is to recognize the limitations of analysis but to nonetheless employ a systematic approach whenever and wherever possible. Rivlin was especially forceful in calling for better evidence of program effects, perhaps the central feature of any systematic approach. Few would disagree that everyone from program managers to senior level policymakers could improve their decisions if they had reliable information about program impacts, or that developing programs with strong positive effects that can be widely replicated should be a fundamental objective in both policymaking and program evaluation.

Rivlin’s propositions today

Updated to 2011, the Rivlin view of rational policymaking is still central to improving policy decisions. Ironically, the Rivlin proposition that now provides the strongest basis for expanding evidence-based policies is the dramatic expansion of high-quality evidence on programs that work (or not), the proposition that Rivlin thought the weakest in 1971. The most important contribution of social science to the public good is the use of scientific designs that allow definitive answers about whether specific intervention programs produce their intended impacts. Given this powerful tool, in a perfect world policymakers could follow a simple decision rule on program funding: if the program works, continue or even expand its funding; if it doesn’t work, reduce or end its funding or find ways to improve it.

Evidence from scientific designs is now available for a large and growing set of interventions in early childhood education, K-12 reading and math, treatment of families that abuse or neglect their children, preparation of high school students to enter the world of work, community-based programs for juvenile delinquents and their families, several program models that reduce teen pregnancy, ‘second chance’ programs for children who have dropped out of school, prison release programs, and many others.

Broadening the evidence-based approach to achieve greater impacts in attacking society’s social problems, government (and the private sector, especially foundations) can employ two approaches. First, as government provides money to establish new social programs, the money should be accompanied by a requirement that the specific programs implemented at the local level be supported by strong evidence from scientific evaluations. Indeed, government might even specify a set of evidence-based programs that can be funded in order to avoid conflicts over what constitutes strong evidence. As we will see, the Obama administration has pioneered methods of identifying evidence-based programs and of ensuring that only evidence-based programs are implemented with government dollars.

Of course, anyone who has watched policymakers in action knows that they will rarely allow evidence on program effectiveness to be the sole or even major factor driving the policy process. Politicians focus on costs, the needs and desires of their constituents, the position of their party leaders, public opinion, their own political philosophy, pressure from lobbyists, the position favored by people and groups that finance their campaigns, and a host of other factors in making decisions about how to vote on program proposals. Allowing an adequate range for all these factors however, does not gainsay the possibility that in some circumstances evidence can have (and has had) a major impact on political decisions.

The second approach to employing evidence to improve social programs is to ensure that programs are implemented in a way that reliable information about program impacts is continuously generated. One of the Achilles heels of social programs is diminishing effectiveness as program models are implemented in more locations. A leading example of this problem is Head Start in the US. Over the past four decades, numerous preschool programs have shown that they can have both immediate and lasting impacts on children learning and other behaviors.⁵ Yet a recent high-quality evaluation of Head Start, a program specifically designed to spread the benefits of preschool to a very large (enrollment in 2010: 900,000 children) group of disadvantaged children, produced only modest impacts that were barely detectable at the end of the first grade.⁶ To combat the problem of diminishing impacts as programs are expanded to new sites, program operators must be vigilant in following the program model, perhaps adapted in some ways to local conditions. The key to replication of effective program models is continuous generation of evidence on program effects on

4. Rivlin, p. 3.

5. Ramey, C., Campbell, F., and Blair, C. (1998) *Enhancing the Life Course for High-Risk Children: Results from the Abecedarian Project*. In ‘Social Programs that Work.’ Ed., Jonathan Crane. New York: Russell Sage Foundation; Schweinhart, L.J. and others (2005) ‘Lifetime Effects: The High/Scope Perry Preschool Study through Age 40.’ Ypsilanti, MI: High/Scope Press; Reynolds, A.J. (2000) ‘Success in Early Intervention: The Chicago Child-Parent Centers.’ Lincoln, NE: University of Nebraska Press; Barnett, W.S. and others (2007) ‘Effects of Five State Pre-Kindergarten Programs on Early Learning.’ Rutgers University: National Institute for Early Education Research; Gormley, W.T. Jr., Phillips, D., and Gayer, T. (2008) *Preschool Programs Can Boost School Readiness*. ‘Science’ 320: 1723-1724.

6. Puma, M. and others (2010) ‘Head Start Impact Study: Final Report.’ Report prepared for the Office of Planning, Research and Evaluation Administration for Children and Families, US Department of Health and Human Services. Rockville, MD: Westat.

participants and adjustments in implementation if the program is not achieving its expected effects. For this reason, enabling legislation should provide a mandate for continuous evaluation and the funding to make it possible.

7. Here is what President Obama said in his inaugural address: "The question we ask today is not whether our government is too big or too small, but whether it works – whether it helps families find jobs at a decent wage, care they can afford, a retirement that is dignified. Where the answer is yes, we intend to move forward. Where the answer is no, programs will end."

8. Funds for an initiative called the Workforce Innovation Fund, which is a mix of systems reforms of the Department of Labor's employment and training programs and funds for high-quality evaluations of employment and training model programs, was in the Obama 2011 budget. However, when Congress was unable to pass the original 2011 appropriations bill, all new discretionary funding (with a few very modest exceptions) was suspended. Then, surprisingly, Republicans in the House allowed the provision to be funded in the final compromise on the 2011 appropriations bill.

9. Carol Weiss has recently called this "imposed" use of research. See Weiss, C.H., Murphy-Graham, E., and Birkeland, S. (2005) *An Alternate Route to Policy Influence: How Evaluations affect D.A.R.E.* 'American Journal of Evaluation.' 26. (4): 12-30.

10. Budget Director Peter Orszag, after a number of speeches in which he talked about how to identify good and bad programs, spelled out his and the President's thinking on using high-quality program evaluations to improve program quality in a blog post and a memo to the heads of all executive branch departments and agencies. See Orszag, P.R. 'Building Rigorous Evidence to Drive Policy.' Office of Management and Budget. Blog, June 8, 2009. Available at: <http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivePolicy> and Orszag, P.R. (2009) 'Memorandum for the Heads of Executive Departments and Agencies.' Executive Office of the President, Office of Management and Budget. Available at: http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-01.pdf

2. President Obama's evidence-based initiatives

Origin of the initiatives

A remarkable aspect of the Obama evidence-based initiatives is that the President intended from the beginning of his administration to fund only programs that had solid evidence of success. In the American system, the president has extraordinary power in agenda-setting. The president can raise the profile of and help to define specific problems, as well as give increased credibility to specific solutions. The president is often the most important single actor in the struggle to enact legislation. He can veto any legislation he doesn't like and because it takes a two-thirds majority in both Houses of Congress to over ride a presidential veto, it is rare that a president's veto, or the threat of a veto, does not spell death for the legislation. The president also has considerable power in enacting legislation. A president whose party controls Congress can almost always force consideration of favored proposals and usually guide them to enactment. Even presidents who face a Congress controlled by the other party can often get his favored issues at least considered, especially if he works to develop bipartisan relations with congressional leaders.

President Obama, encouraged and supported in highly creative fashion by the Office and Management and Budget (OMB) Director, Peter Orszag, intended to place major emphasis on making decisions about social programs based on evidence. OMB, located within the Executive Office of the President, is the lead Executive Branch agency responsible for development and execution of the President's budget and oversight of federal agency performance. As early as his inaugural address, the President made it clear that an important goal of his administration would be to expand programs that work and eliminate programs that don't.⁷ From the earliest days of the administration, senior officials at OMB were planning several initiatives to advance the use of evidence-based program models and to generate high-quality evidence on new program models. When President Obama took office, career officials at OMB, who are often the source of ideas for increasing government efficiency, were already involved in a formal attempt to encourage federal agencies to conduct high-quality evaluations of their programs. Building on this effort which began in the Bush administration and even earlier, by the end of the second year of the Obama administration there were five initiatives well underway and one that was well formulated but stuck in the congressional enactment process.⁸

How the Obama administration is building the initiatives

Based on several interviews with members of the Obama administration and others inside and outside Congress knowledgeable about the Obama initiatives, we think the following outline captures the main steps of the President's evidence-based initiatives:

1. Select an important social problem that would make individual citizens and the nation better off if reduced in magnitude;
2. Identify model programs addressed to the problem that have been shown in randomized controlled trials or other rigorous research to significantly reduce the problem;
3. Obtain funds from Congress to scale-up evidence-based programs of this type that addresses the problem in accord with the proven models;
4. Make the funds available to government or private entities with a track record of good performance to replicate the successful model programs;
5. Continuously evaluate the projects as they are implemented to ensure they are faithfully implementing the model program and producing good results.

The Obama administration has now created a sweeping new opportunity for rigorous evidence to influence policy.⁹ No president or budget director for a president have ever been so intent on using

11. It would be possible for major players in the legislative branch on either committees with jurisdiction over particular social programs (either the authorizing committee or the appropriations committee) to try to enact evidence-based policies. Even the most powerful committee chairman, however, generally has less power and influence than the president. Nonetheless, powerful Congressional players could use an approach like the Obama administration is using to influence the adoption of evidence-based policies.
12. By way of full disclosure, Baron is the President of the Coalition and Haskins is on their Advisory Board.
13. See editorial by Solow, R., and Baron, J. (2010) Long-Term Deficit Reduction: Less Pain, More Gain. 'The Fiscal Times,' August 12, 2010. Available at <http://coalition4evidence.org/wordpress/wp-content/uploads/Solow-Baron-Op-Ed-TheFiscalTimes-8-12-10.pdf>
14. These are 2011 dollars. Examples include: (i) the Riverside Greater Avenues for Independence (GAIN) Program (Freedman, S. and others (1996) The GAIN Evaluation: Five-Year Impacts on Employment, Earnings, and AFDC Receipt. Working Paper 96. MDRC; Riccio, R., Friedlander, D., and Freedman S. (1994) GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program, MDRC); (ii) Los Angeles Jobs-First GAIN (Freedman, S. and others (2000) The Los Angeles Jobs-First GAIN Evaluation: Final Report on a Work First Program in a Major Urban Center, MDRC); and (iii) Portland Job Opportunities and Basic Skills (JOBS) Training Program (Scrivener, S. and others. (1998) National Evaluation of Welfare-to-Work Strategies: Implementation, Participation Patterns, Costs, and Two-Year Impacts of the Portland (Oregon) Welfare-to-Work Program, MDRC; Hamilton, G. and others (2001) National Evaluation of Welfare-to-Work Strategies: How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs, MDRC and Child Trends).
15. Cheh, V. (2001) 'The Final Evaluation Report on the National Home Health Prospective Payment Demonstration: Agencies Reduce Visits While Preserving Quality.' Report submitted by Mathematica Policy Research, Inc. to the Health Care Financing Administration.

evidence to shape decisions about the funding of social programs as President Obama, former Budget Director Orszag, and other senior officials at OMB.¹⁰ The Obama plan turns the normal relationship between policy decision making and use of evidence on its head. Instead of evidence being on the outside of the decision making process trying to get in, Obama brings evidence inside from the beginning. The administration must still convince others – especially those who hold the purse strings in Congress – that the use of evidence will improve policymaking and program outcomes. But those arguments are being made by people inside the administration to retain an evidence-based approach as a fundamental part of the President's legislative agenda rather than fighting from the outside to insert evidence-based policies into the process.¹¹

Although less emphasized, the Obama plan for basing funding decisions on rigorous evidence can be useful for cutting spending as well as funding new programs. In the current age of fiscal austerity in both the US and the UK, when cuts in social programs are inevitable, it will be far better to cut programs that have minimal or no impacts than successful programs or programs that show promise. Yet experience shows that once enacted, federal programs become, as the economists say, 'sticky.' Thus, social programs are hard to get rid of, no matter how bad they are. The nation's social programs are unlikely to be improved until we learn to enact programs supported by rigorous evidence, to improve existing programs based on evidence, and to shut down failing ones, again based on evidence from high-quality program evaluations. Reliable evidence on program effects can be put to good use both in expanding and cutting programs.

Consider a recent example of how ineffective federal agencies in the US are in separating programs that don't work from those that do. The Deficit Reduction Act of 2005 established an Academic Competitiveness Council (ACC) to, among other goals, identify all federal programs with a science, technology, engineering, or math (STEM) education focus and determine whether these programs were effective. The ACC found that the federal government was operating 105 science, technology, engineering, and mathematics programs that spent well over \$3 billion (in 2006). ACC leaders then asked the federal agencies to send their most rigorous STEM program evaluations to the non-profit, non-partisan Coalition for Evidence-Based Policy, a Washington group that works with federal officials to advance evidence-based program reforms.¹² The Coalition independently reviewed these studies for the ACC, with funding from the W.T. Grant Foundation. The Coalition reviewed a total of 115 evaluations, and found only ten scientifically-rigorous impact evaluations (i.e., well-conducted randomized experiments or well-matched comparison-group studies) that had reported findings, as well as 15 others that were underway but had yet to report results. Six of the ten completed evaluations found that the STEM program had weak or no effects on educational outcomes. Thus, only four of the programs had been found by high-quality evaluations to be effective; most of the remainder had never been rigorously evaluated.

Rigorous evidence can also play a key role in deficit reduction in another important way. The solution to the US's long-term deficit problem is sometimes portrayed as a choice among sharp budget cuts, major tax increases, or a combination of the two. Given the magnitude of the problem, some level of sacrifice is unavoidable. But largely overlooked in the discussion are clear examples, from welfare and health care policy in the 1980s and 1990s, in which rigorous randomized experiments identified program reforms that produced important budgetary savings without adverse effects and, in some cases, with improvements in people's lives. Similar cost-saving opportunities already exist in a few areas, and many more could likely be identified through rigorous research.¹³ In the 1980s and 90s, for example, federal officials sponsored many large randomized evaluations of state and local welfare reforms. These studies showed convincingly that certain reform models that emphasized moving participants quickly into the workforce through short-term job search assistance and training – as opposed to providing remedial education – produced large gains in employment and earnings, reductions in welfare, and net entitlement savings (in AFDC and Food Stamps) of \$1,700-\$6,000 per participant.¹⁴ Such findings helped shape the 1996 federal welfare reform act and the major work-focused reforms in state and local welfare programs that followed.

Similarly, in 1995, federal officials launched a rigorous experimental evaluation to test prospective payment of Medicare home health agencies – i.e., paying such agencies an up-front lump sum per patient – against the usual cost-reimbursement approach. The evaluation found that prospective payment reduced costs to Medicare by 20 per cent over three years, compared to cost reimbursement, with no adverse effects on patient health.¹⁵ This finding helped shape Medicare's

nationwide implementation of prospective payment for home health agencies in 2000, producing large cost savings in this \$15 billion program.¹⁶

Such examples illustrate how rigorous experimental studies can build the evidence needed for significant *and* smart spending reductions. But to identify enough of these cost-saving strategies to produce sizable long-term deficit reduction, many more rigorous experiments testing a wide range of strategies are needed.

Issues raised by the Obama initiatives

Evaluating the impacts of federal spending on social programs encounters a special problem with federal funding that goes to states in the form of block grants or funding streams that provide great flexibility in how state and local governments use the funds. Two examples of such funding streams are the Title I education program that provides funds to states to help increase the educational achievement of poor children and the Temporary Assistance for Needy Families (TANF) program that gives states block grant funding to provide financial support to poor families, to decrease nonmarital births, and to promote work. The advantage of providing states with so much flexibility in their use of federal dollars is that they can better match their use of the money to local needs. But the problem is that so much flexibility makes it difficult to figure out how states are actually spending the money, let alone trying to evaluate the effectiveness of such spending. The Obama administration has addressed the problem of flexible spending by making the money in all their evidence-based initiatives contingent on both a clear statement of goals by grant recipients consistent with the goals stated in the federal notice of funding availability and by requiring exact specification of the model programs they will accept for funding.

Another problem with using an evidence-based approach applies to the case of using evidence to improve or eliminate programs shown to be failing or producing modest benefits. Even when ineffective programs have been identified, it does not follow that the administration or Congress will take action. According to Isabel Sawhill and Jon Baron, since 1990 there have been ten instances in which a large-scale federal social program was evaluated by a scientific research design. In nine of these ten cases (including Job Corps, Upward Bound, Even Start, the Job Training Partnership Act (JTPA), and Head Start), highly popular programs were shown to have modest or no impacts on their participants.¹⁷ So far, these disappointing results have resulted in changes in only a few of the programs. The poor results for Head Start caused the Obama administration to use regulations to propose the most radical changes in Head Start's history.¹⁸ The Even Start family literacy program was eliminated this year and funding for the JTPA youth programs was cut. But the other seven programs have continued without major changes.

These examples show that the federal government needs to find a better way to spend money on social intervention programs – and especially a way to overcome the political support that keeps program funds flowing to ineffective projects and interventions. The Obama initiatives for funding social programs are the most important attempts so far to find this better way and could potentially have a major impact on how social programs are funded by greatly elevating the role of program evaluations in program expansion or contraction. Moreover, if the Obama initiatives are effective, the average impact of US social intervention programs on the well-being of children and families will increase and the nation will be better off. Then the taxpayer investment in evaluation research will be fully justified.

Development of the initiatives

As unlikely as it might seem, the story of the Obama initiatives begins with Robert Shea, a senior political appointee at OMB in the Bush administration.¹⁹ Republicans have often been harsh critics of spending on social programs in part because they think that most of the programs don't work. Even so, Shea wanted to develop a system that OMB could use to get federal agencies to improve the effectiveness of their programs. After consulting with people inside and outside the administration, Shea and his colleagues at OMB developed the Program Assessment Rating Tool (PART).²⁰ One important feature of the PART, which OMB ordered agencies to follow in evaluating the effectiveness of their programs, was that it stated that well-conducted random assignment designs were the gold standard for program evaluation, while recognizing that such studies were not always feasible and suggesting second-best alternatives in these cases. This element of PART was based on input from the Coalition for Evidence-Based Policy and other evaluation experts. Not surprisingly, some agencies were not excited about being told to conduct rigorous

16. US General Accounting Office. (2000) Medicare Home Health Care: Prospective Payment System Will Need Refinement as Data Become Available. GAO/HEHS-00-9, April.
17. Sawhill, I.V., and Baron, J. (2010) Federal Programs for Youth: More of the Same Won't Work. 'Youth Today,' May 2010, p. 21.
18. Haskins, R., and Barnett, W.S. (2010) A Test for Head Start. 'The Washington Post,' October 11, 2010; Guernsey, L. (2010) 'Proposed Rules Will Shake Up Head Start.' Early Ed Watch. Blog. New America Foundation, Early Education Initiative, September 23, 2010. Available at: http://earlyed.newamerica.net/blogposts/2010/proposed_rules_will_shake_up_head_start-37234
19. This account is based on the documents that will be cited below and interviews with Grover 'Russ' Whitehurst, former head of the Institute of Education Sciences; Robert Shea, the lead for OMB in developing the Program Assessment Rating Tool (PART) during the Bush administration; David Olds, the developer of the Nurse-Family Partnership who played an important role in the Bush initiative on evidence-based policy; Robert Gordon, the OMB senior staffer for Obama who played a central role in the Obama evidence-based initiative; and Kathy Stack, a senior OMB career official, who has been at the center of emphasizing evidence-based policy during both the Bush and Obama administration.
20. ExpectMore.Gov, 'The Program Assessment Rating Tool (PART).' Available at: <http://www.whitehouse.gov/omb/expectmore/part.html>

impact evaluations of their programs, in part because they did not want to use their funds on expensive evaluations. Even so, the significance of the PART episode is that, even before the Obama administration took over at OMB, senior OMB officials – including career officials who would remain in place after President Obama and the new OMB team assumed office – had taken a firm stand about both the importance of program evaluation and the strongest methodology for conducting the evaluations. Moreover, OMB had put the federal administrative agencies that sponsored intervention programs on notice that it expected sustained attention to using evidence to improve programs.

21. The memo OMB sent to the executive agencies incorporated key concepts suggested by the Coalition for Evidence-Based Policy, regarding which study designs are most likely to yield valid estimates of a program's impact. The Coalition also did a series of workshops with OMB and agency staff on this and related evidence-based policy concepts.
22. Here is the exact language from the House-Senate conference agreement on funds for home visitation: "Within the amount provided for State grants, the conferees include \$10,000,000 for a home visitation initiative to support competitive grants to States to encourage investment of existing funding streams into evidence-based home visitation models. The conferees expect that the Administration for Children and Families will ensure that States use the funds to support models that have been shown, in well-designed randomized controlled trials, to produce sizeable, sustained effects on important child outcomes such as abuse and neglect. The conferees also recommend that the funds support activities to assist a range of home visitation programs to replicate the techniques that have met the high evidentiary standards. In carrying out this new initiative, the conferees instruct the Department to adhere closely to evidence-based models of home visitation and not to incorporate any additional initiatives that have not met these high evidentiary standards or might otherwise dilute the emphasis on home visitation." See Conference Report on H.R.3043, Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act 2008, p. H12486.

At about the same time that Shea and the OMB professional staff were developing the PART program, David Olds, the designer of the Nurse-Family Partnership and a strong advocate of random-assignment program evaluation, came to OMB for a high-level meeting about his Nurse-Family Partnership program and about the Bush administration's use of social science evidence. The White House website had posted information about youth programs that it believed had been found by good evaluations to be successful, when in fact many of the evaluations listed were inferior. Olds wanted to convince the administration that it was inappropriate to classify well-conducted random-assignment evaluations with evaluations that used inferior methods, not least because the criteria followed in such classifications threw Olds' elegant work in with projects with modest or worse evidence. As a result of the Olds meeting, OMB created an interagency group to review evidence standards and compare the criteria for good evidence being used by various agencies on their websites featuring evidence-based practices. Again, the agencies were put on notice by OMB that well-conducted random-assignment designs provided the most reliable evidence, although not the only evidence the administration was willing to accept.²¹ A number of agencies then began changing their best practice websites to reflect OMB's hierarchy of study designs for assessing program effectiveness. But the most important outcome following the Olds meeting was that Shea and his staff decided the administration should use President Bush's 2008 budget to propose funding for expansion of the Olds program on the grounds that it would demonstrate the administration's (and OMB's) commitment to reward programs that had been shown to be effective in convincing random-assignment studies. Thus, the administration proposed and managed to get enacted a small \$10 million pot of funds for states that agreed to use the money, augmented by their own funds, to mount Olds-type home-visiting programs.

During consideration of the Bush home-visiting legislation, the Coalition for Evidence-Based Policy provided input to Congress – as it had previously to the Academic Competitiveness Council and to OMB. Specifically, the Coalition urged the Appropriations Committees in the House and Senate – which are responsible for the annual funding legislation for the federal agencies – to include the administration's \$10 million for home visiting in their respective appropriations bills, along with report language directing the implementing agency (HHS) to adhere to a high evidence standard. Coalition leaders had a long-standing relationship with staffers on the Senate Appropriations Committee and a chief staffer there already believed in the importance of rigorous evidence. The Senate Appropriations Committee included the proposed home-visiting funding and evidence standard in their appropriation legislation for 2009.

But the Appropriations Committee on the House side was a more difficult sell. The Coalition engaged extensively with a senior staffer with the Committee, who had heard that many home-visiting programs, in addition to the Olds program, were supported by high-quality evidence. This position had been put forward by advocates for various home-visiting programs that were being supported by state dollars and wanted their favored programs to be included in the \$10 million federal funding for home-visiting programs. In a series of phone calls and e-mails, the Coalition provided input to the House Appropriations staffer on the specific studies and, in many cases, identified key flaws that likely biased the studies toward finding that a particular home-visiting program 'worked.' The staffer gradually was convinced that evidence from well-conducted random-assignment studies was necessary to know if a program was truly working. In the end, she recommended to the Appropriations subcommittee chairman Ralph Regula, that the \$10 million initiative and the Senate (and OMB) language on giving priority to programs supported by scientific evidence be included in their appropriations bill.²² The most important outcome of this episode is that the small \$10 million appropriation for home visiting was enacted and was ready and waiting to be picked up by the Obama administration when it took over and began to inaugurate its own evidence-based work.

23. Orszag's take on the importance of evidence was succinctly summarized in a now-famous blog post that appeared in the midst of the administration's various evidence-based initiatives; see Orszag, P.R. (2009) 'Building Rigorous Evidence to Drive Policy.' Available at: <http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidenceToDrivePolicy/>

Fortunately, the Obama team at OMB, led by Peter Orszag, Robert Gordon, and Jeffrey Liebman, came into office fully apprised of the value of random-assignment evaluations. Indeed, it is doubtful that any team of senior officials in OMB history was as knowledgeable about and committed to scientific program evaluation as the Obama team. With both a head start from the Bush administration and a new team of powerful OMB officials fully committed to the value of experimental evaluations, the Obama administration lost little time in launching its bold initiative to expand evidence-based social programs.²³

Table 1: The Obama plan for expanding evidence-based programs

Stages in Programs	Teen Pregnancy Prevention	i3 (Investing in Innovation Fund)	Home Visiting	Social Innovation Fund	TAA Community College and Career Training Program	Workforce Innovation Fund
Administering Agency	Health and Human Services	Education	Health and Human Services	Corporation for National and Community Service	Departments of Labor and Education	Departments of Labor and Education
Review of Literature	Completed by Mathematica	Although there was no formal review of literature, the i3 evidence tiers were based on a process for reviewing evidence developed by IES over several years primarily through its work on What Works Clearinghouse and strengthened	Completed by Mathematica	None	None	None
Amount of Awards	\$100 million awarded--\$75 million to replicate existing programs, \$25 million to test new strategies	Up to \$650 million across the three types of grants (development, validation, scale-up) Note: Applicants must obtain 20 per cent of their award amount in matching funds or in-kind donations	\$88 million awarded in year 1, \$1.5 billion over 5 years	\$49.3 million awarded in FY 2010, \$49.9 million appropriated for FY 2011 Note: Each federal dollar must be matched 1:1 at grantee and again at sub-grantee level	\$500 million a year for 4 years	\$125 million appropriated in FY 2011
Review Panel Selection	Panels included both expert peer reviewers and federal staff	Over 330 peer reviewers were selected from 1,400 experts in both subject matter and research/evaluation; reviewers assigned to panels of 3-5 people	Applications were reviewed by grants management officials and program staff	A total of 60 experts were drawn from a pool of experts/professionals and the CNCS reviewer database of 2,300 people; reviewers assigned to panels of 2-4 people	Not Public Yet. Technical review panels will evaluate all applications against evaluation criteria provided in application materials	Undetermined
Selection of Proposals	75 applicants awarded grants to replicate existing programs; 27 grantees awarded grants to test new strategies	49 applications chosen as "highest-rated." All secured matching funds (20 per cent of grant amount)	49 state governments, DC, and 5 territories applied and were awarded funding; second stage of funding to follow	11 grantees chosen for 2010; they have chosen 128 sub-grantees out of over 500 applicants	Not Yet	Undetermined

3. Program areas addressed by the initiatives

The administration has now taken action to implement its evidence-based strategy in six specific areas of social intervention. What follows is an overview of the major characteristics and state of play for the administration's six evidence-based initiatives. Table 1 provides an overview and comparison of the major characteristics of the six initiatives.

Home visiting

Home visiting is a service strategy to help families, usually mothers, in one or more of three domains: maternal and child health, early childhood development, and family functioning.²⁴ Usually conducted by trained social workers or nurses, a number of home-visiting program models have been evaluated in random-assignment studies that estimate impacts on a variety of parenting behaviors and child and maternal outcomes. The home-visiting approach is often advertised as a way to reduce child abuse and neglect, but recent reviews have found that – with a few important exceptions such as the Olds program – many of the leading models produce weak or no lasting effects on these or other important outcomes, such as child cognitive development and family economic well-being.²⁵ As we have seen, home visiting differs from the other evidence-based initiatives in that federal funds had already been made available to evidence-based home-visiting programs during the last year of the Bush administration. The Bush funds were used to create cooperative agreements with 17 grantees that used the money to conduct planning efforts to coordinate existing federal, state, and local funds that could be used for home visiting.

The Obama home-visiting initiative, which has already been awarded \$1.5 billion in guaranteed funding over the 2010–2014 period by Congress, greatly expands the initiative enacted during the Bush administration. In addition, the process of distributing funds differs from the other Obama evidence-based initiatives. The funds are being distributed in three stages. In the first stage, which has been completed, all states were eligible for a share of funding if they submitted proposals that met administration requirements, primarily that they present a plan for conducting an assessment of the need for home-visiting programs in their state. Forty-nine states, the District of Columbia, and five territories were awarded funds to enter the second stage. In the second stage, states are required to complete their needs assessment and submit the results. States that successfully passed the first two stages were then eligible to update their state plans and receive grant funding to actually scale up their home-visiting program. In preparation for the third stage, and parallel in some respects to the Obama initiative on teen pregnancy (see page 12), the administration commissioned a literature review of home-visiting programs by Mathematica Policy Research.²⁶ Among other things, Mathematica found that seven home-visiting model programs met the minimum criteria set out in the program's authorizing legislation to achieve the status of evidence-based programs. Thus, states applying for the home-visiting funds in the third stage must spend at least 75 per cent of their funding on one of the seven evidence-based models, leaving up to 25 per cent of the funds to be used for 'promising' model programs. States are free to spend 100 per cent of their funds on one or more of the seven approved evidence-based models.

At this writing, the Department of Health and Human Services (HHS) is expected to soon issue the final solicitation for grants to initiate the third and final stage of the home-visiting program. But HHS outlined its proposed plan for this third stage in a July 2010 Federal Register notice. As noted above, Mathematica's review identified seven program models that meet the minimum evidence threshold set out in the program's authorizing statute. However, Mathematica's review, as well as the Coalition's review and a 2009 review published in the *Lancet*, found that most of these models – although perhaps meeting the minimum statutory threshold – produced weak or no lasting effects on key outcomes. A few models, such as Olds' program, were found to produce stronger, more durable effects.

Thus, in the third stage of the home-visiting program, HHS plans to allocate program funding that exceeds the first-year (2010) amount through a competition process, in which: *"HHS proposes to give significant weight to the strength of the available evidence of effectiveness of the model or models employed by the State. In this context, the use of program models satisfying the criteria outlined [for the program's initial grants] would be a minimal requirement, but HHS would consider additional criteria that further distinguish models with greater and lesser support in evidence."*

24. Stoltzfus, E. and Lynch, K.E. (2009) 'Home Visitation for Families with Young Children.' (R40705), Washington, DC: Congressional Research Service, October 2009.
25. Howard, K.S. and Brooks-Gunn, J. (2009) The Role of Home-Visiting Programs in Preventing Child Abuse and Neglect. 'The Future of Children.' 19(2): 119–146; see especially Table 2: The Effects of Home-Visiting Programs on Child Abuse, Health, Parenting, and Depression, p. 133; MacMillan, H.I. and others (2009) Interventions to Prevent Child Maltreatment and Associated Impairment. 'Lancet.' 373: 250–266; Paulsell, D. and others (2010) 'Home Visiting Evidence of Effectiveness Review: Executive Summary.' Report prepared for the Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services. Princeton, NJ: Mathematica Policy Research, Inc. Available at: http://homvee.acf.hhs.gov/HomVEE_Executive_Summary.pdf Additional information is available at: <http://homvee.acf.hhs.gov/Default.aspx>
26. Paulsell, D. and others (2010) 'Home Visiting Evidence of Effectiveness Review.' Report prepared for the Office of Planning, Research and Evaluation, Administration for Children and Families, Department of Health and Human Services, Washington DC.

We strongly support this program structure, as it enables the program to evolve toward greater effectiveness over time based on evidence about impact of the various models.

An interesting aspect of the home-visiting initiative is that the administration has already selected MDRC, one of the nation's foremost firms conducting large-scale random-assignment studies, to evaluate the home-visiting programs.²⁷ The details of the evaluation plan have not yet been released, but the involvement of MDRC makes it clear that the administration intends to follow through on its repeated emphasis on basing their initiatives on evidence. In the case of home visiting, evidence played a central role in the selection of model programs and will again play a central role in the evaluation of the impacts of the state programs.

Teen pregnancy prevention

The teen pregnancy prevention initiative has proceeded mostly in accord with the components of the Obama model outlined above. Teen pregnancy is not only a serious national social problem with demonstrated impacts on the mother, the father, and the child, it is also an area of intervention that has a long track record of creative and diverse programs. A comprehensive review of programs by Douglas Kirby published in 2001 found eight program models that had what Kirby called "*strong evidence of success*."²⁸ There is also a comprehensive review of the evidence published by the Campbell Collaboration in 2006 that identified several successful evidence-based programs.²⁹ Thus, the first two components of our outline of the Obama approach to evidence-based initiatives – selecting a serious problem and ensuring that there are evidence-based model programs – have certainly been met in the case of teen pregnancy prevention. As in the home-visiting initiative, the administration commissioned a literature review from Mathematica that was made available to the public. The review identified 28 model programs that were supported by high-quality evidence. However, the review found that only two of these models are backed by well-conducted random-assignment studies showing a sustained effect on the most important measure – the actual reduction of teen pregnancies three to four years after random assignment. The other 26 models are backed by more preliminary evidence – in most cases, random-assignment studies or comparison-group studies showing only short-term effects on intermediate outcomes such as condom use and number of sexual partners, but not the final, most policy-relevant outcomes (pregnancies, births, sexually-transmitted diseases). When programs backed by such preliminary evidence are evaluated in more definitive random-assignment studies with longer-term follow-up, sometimes they are found to produce impacts on the long-term outcomes, but too often they are not. Fortunately, it appears that HHS plans to rigorously evaluate a number of the funded models to determine which are truly effective in preventing teen pregnancies.

Based in part on the Mathematica review, the administration issued its solicitation for grants in April 2010. Because there had been so much previous research in this field, the administration decided to award two tiers of funding. Tier 1, which would receive most of the money, would pay for program models identified in Mathematica's review as having higher-quality evidence of success. Tier 2 would be for programs that had some evidence of success, but did not reach the higher standard reached by Tier 1 programs. The applications for funding were reviewed by a panel of experts based on review criteria published by the administration. Seventy-five projects were selected for Tier 1 funding of \$75 million. In addition, \$25 million was awarded to 27 Tier 2 projects that have some, but not strong, evidence of success. A notable feature of the teen pregnancy initiative is how the administration has made so much of the written material available to the public, including a detailed report from Mathematica on how it conducted the literature review, a list of the projects approved for funding that included extensive information about each project, and more.

The congressional journey of the teen pregnancy initiative demonstrates an unfortunate and perhaps fatal political obstacle faced by some or even all of the Obama evidence-based initiatives. Like the UK and many other rich nations, the US has been running a huge budget deficit that threatens to bankrupt the federal government.³⁰ The UK has recently responded to its deficit crisis with perhaps the most sweeping program of spending cuts and revenue increases in its history.³¹ Congress and the executive branch in the US are now in the process of enacting a series of cuts in spending. The first important deficit-cutting action, although little more than a footnote to the extensive spending cuts and tax increases that will be required to make a serious dent in the nation's deficit, was a package of spending reductions as part of congressional action on the 2011 budget. One of the cuts enacted by House Republicans, who hold the majority, was a complete

27. Haskins is on the Board of MDRC.

28. Kirby, D. (2001) 'Emerging Answers: Research Findings on Programs to Reduce Teen Pregnancy.' Washington, DC: National Campaign to Reduce Teen Pregnancy, p. 179.

29. Scher, L., Maynard, R.A., and Stagner, M. (2006) 'Interventions Intended to Reduce Pregnancy-Related Outcomes among Adolescents.' (Oslo, Norway: Campbell Coalition).

30. National Commission on Fiscal Responsibility and Reform. (2010) *The Moment of Truth* (Washington DC: White House). Available at: http://www.fiscalcommission.gov/sites/fiscalcommission.gov/files/documents/TheMomentofTruth12_1_2010.pdf

elimination of Obama's teen pregnancy initiative. In fact, House Republicans actually voted to zero out the initiative, but was subsequently prevented from doing so by the Senate. Even so, the \$110 million initiative was trimmed to \$105 million and Republicans seemed poised to try again to reduce or eliminate the initiative as part of action on the 2012 budget now being considered by Congress. There is little question that several of Obama's evidence-based initiatives will face further attacks as Congress and the President attempt to cut spending to reduce the federal deficit.

Investing in Innovation Fund (i3)

The i3 Fund and the Social Innovation Fund (SIF; see below) are very different from the home-visiting and teen pregnancy reduction initiatives in that both fund a more diffuse set of programs. In the case of i3, virtually any preschool or K-12 intervention with evidence of success or promise could receive funding. The i3 fund, like the teen pregnancy initiative, recognized multiple levels of evidence-based programs, in this case three levels. The top tier, called scale-up funds, was awarded for programs supported by evidence from rigorous evaluations. Funds from the second tier, called validation grants, were awarded to programs with some but less evidence of success. Finally, development grants were awarded to programs with a reasonable hypothesis. In order to qualify for funding, the programs had to aim to improve outcomes for pre-school children, help students qualify for or succeed in college, help students with disabilities or with limited-English proficiency, or serve schools in rural areas. The initiative is funded at \$650 million. School systems, consortiums of school systems, or nonprofit agencies partnering with school systems were eligible to apply for funding. Awards were announced in August 2010 for all three categories of evidence-based programs. A total of 49 awards were made; all the projects managed to attract the required matching funds in order to receive the federal dollars and are now in various stages of implementation. Again, the administration made public a large number of documents about their grant-making process and about the projects receiving awards, including the comments of reviewers for the highest-rated i3 applicants and summary information and the applicant narratives for the highest-rated applicants for the scale-up, validation, and development awards. The administration has announced that random-assignment evaluations will be conducted for several of the biggest i3 grants.

Social Innovation Fund (SIF)

The President has said that solutions to America's domestic problems *"are being developed every day at the grass roots"*³² and that his administration wants to support those grassroots efforts. SIF is one method by which the administration intends to *"identify and grow high-performing nonprofit organizations"* with experience at the local level.³³ The unique feature of the SIF is the mechanism of awarding funds. SIF funds are awarded in a two-stage process, first to intermediary organizations with a track record of funding successful local, community-based organizations; the intermediary organizations then select local organizations for funding. The intermediary and local organizations must raise matching equal to the value of the SIF grant. The intermediaries were organizations that had *"strong track records of identifying and growing high-performing nonprofit organizations."*³⁴ In July 2010, 11 such intermediaries were awarded \$50 million in funding to go with another \$74 million they had raised in matching funds to be distributed to nonprofit organizations. The nonprofits in turn were to use the money to conduct evidence-based programs addressed to at least one of three broad areas of social policy: economic opportunity, youth development and school support, and promoting healthy lifestyles and avoiding risky behavior.

The 11 intermediaries selected by the administration are reputable organizations with experience funding local programs, but whether they can make judgments about evidence-based programs is something that needs to be examined. In fact, the entire procedure of awarding funds to some organizations to in turn award funds to other organizations is an issue that bears careful study.

Like the teen pregnancy prevention initiative, the SIF initiative and even its federal administering agency (the Corporation for National and Community Service) seem to be in the bulls' eye for spending cuts by House Republicans. Funding for 2011 survived the budget scalpel, but future attacks should be expected.

Community College Challenge Fund

Several of our sources told us that the US Department of Labor had only a modest commitment to rigorous evaluation at the beginning of the Obama administration. Nonetheless, the administration was eventually successful in getting the Department to sign off on an evidence-based initiative

31. The Unlikely Revolutionary. 'The Economist.' August 12, 2010.
32. Obama, B.H. (2010) 'Social Innovation Fund.' Reprinted in Office of Social Innovation and Civic Participation. Available at: <http://www.whitehouse.gov/administration/eop/sicp/initiatives/social-innovation-fund>
33. Corporation for National and Community Service 'Inaugural Social Innovation Fund Grants Awarded to Experienced Innovators.' Press Release, July 22, 2010. Available at: http://www.nationalservice.gov/about/newsroom/releases_detail.asp?tbl_pr_id=1829
34. Corporation for National and Community Service (2010) 'Social Innovation Fund.' Available at: <http://www.nationalservice.gov/about/serveamerica/innovation.asp>

to provide funds for training of displaced and unemployed workers and other young adults by the nation's community colleges. The \$2 billion initiative – \$500 million a year for four years – was enacted in 2010. On January 20, 2011 the Department, in conjunction with the Department of Education which will play a somewhat unspecified role in the grant program, released an announcement of the availability of funds for *“the development and improvement of postsecondary programs of two years or less that use evidence-based or innovative strategies to prepare students for successful careers in growing and emerging industries.”*³⁵ An important characteristic of the grants is that community colleges and other entities receiving the funds are to experiment with existing employment and training materials in order to adapt them for use with young adults who seek employment. It will be interesting to review the basis for awarding the grant funds because there are few education, employment, or training programs for use at the community college level that have been rigorously tested and found to produce impacts on students. Thus, it appears that this initiative will focus on developing new curriculums and testing them with rigorous evaluation designs. The awards will be for between \$2.5 million and \$20 million to support projects employing strategies that have been shown to have *“strong or moderate evidence of positive impacts on education and/or employment outcomes.”* Evaluation is a central feature of the Challenge Fund: 25 per cent of the assessment of proposals is based on their evaluation plan; all evaluations must include treatment and control groups; and the Department of Labor will select some grantees for rigorous evaluation using random assignment designs.

- 35. Department of Labor (2011) 'US Labor Department Encourages Applications for Trade Adjustment Assistance Community College and Career Training Grant Program.' News Release, January 20, 2011.
- 36. US Department of Labor 'FY 2011 Budget in Brief,' p. 3.
- 37. Gray, S.W. and others (1966)

Workforce Innovation Fund

This initiative is also being run by the Department of Labor in conjunction with the Department of Education. Five per cent of the 2011 budgets of the Workforce Investment Act (WIA) Adult program and the WIA Dislocated Worker program were set aside to create this fund of nearly \$108 million. The fund will be used to create competitive grants to states or localities to replicate proven practices in training, employment, and reemployment services, especially for vulnerable groups. Like the other evidence-based initiatives, the fund will also be used to test promising practices. As with community college training programs, there is a paucity of program models for employment and training programs with young adults that have been shown by rigorous evaluation designs to produce impacts on student learning, employment, and earnings. It is anticipated that funds will be focused on 'learn and earn,' apprenticeship, and on-the-job training programs, and that selected programs will be rigorously evaluated to determine their impact on key educational and workforce outcomes.

Federal program evaluation

In addition to these six initiatives, the administration also included money in the 2011 budget for program evaluation. Administration staffers claim that there are enough funds in the 2011 budget to pay for about 20 rigorous evaluations of *“the most promising new programs”* and to build the evaluation capacity of the various administrative departments. Indeed, the budget has well over \$60 million for the Department of Labor alone to *“continue to pursue a robust, Department-wide evaluation agenda,”* including rigorous evaluations of WIA performance measures, effects of job counseling, use of administrative data in workforce programs, incentives for dislocated workers, and effects of Occupational Safety and Health Administration inspections.³⁶ In addition, the White House worked with the Department of Labor to create a Chief Evaluation Office that will manage the new evaluations and work with other components of the Department to assist them in conducting rigorous evaluations of their programs.

4. The promise of the Obama initiatives

These six evidence-based initiatives, plus the new funds for rigorous evaluation across the federal agencies, constitute the most sweeping and potentially groundbreaking emphasis on rigorous program evaluation ever conducted by the federal government. Although normal congressional politics played an important role in the formulation and enactment of the Obama evidence-based initiatives, the role of evidence in all six initiatives was more or less unprecedented. By devising several approaches to bring evidence to the center of policy making in the federal government, both in obtaining funds to implement programs supported by rigorous evidence and in generating new evidence on program effects, the administration was able to achieve two benefits that are not often enjoyed by legislation enacted through a routine legislative process. The first benefit is that

the Obama initiatives focus federal dollars on program models that have at least preliminary – and in some cases moderate or strong – evidence of impacts. It does not necessarily follow that the programs will actually produce impacts because program models backed by preliminary or moderate evidence too often turn out not to work when implemented at scale with a more definitive evaluation. But at least money will be spent on programs that have a good chance of having the desired effects. As compared with money now made available through many federal initiatives, this approach represents a great improvement.

The second advantage of the evidence-based approach is that the initiatives require rigorous evaluation of both program implementation and program impacts. At a minimum, the administration's evidence-based approach requires that a standard set of measures be reported to the federal agency responsible for the initiative. But far beyond reporting a standard set of measures, some of the initiatives require projects, as part of their application, to submit a plan for evaluating their implementation and their impacts. Moreover, the administration is making it clear that the quality of the evaluation plan will be a major criterion for deciding which applications to fund. Again, as compared with the scores of federal programs that have tepid or no evaluation requirements, the emphasis on rigorous evaluation evidence in the Obama initiatives may set a precedent for future legislation. And even beyond these two requirements on evidence, at least two of the initiatives and possibly more are requiring projects applying for funds to submit to random-assignment evaluations performed by third party evaluators. The administration, for example, has already hired a crack evaluation organization to help develop and then carry out random-assignment evaluations of the home-visiting programs. It is not yet known how many programs will be evaluated or which ones, but the use of rigorous third-party evaluations shows the lengths to which the administration is going to send a message to federal and state agency officials and program operators that a strong evaluation plan is the new normal in federal funding of social programs.

'Before First Grade.' New York: Teachers College Press; Gray, S.W. and others (1973) The Early Training Project: A Seventh Year Report. 'Child Development.' 41: 909-924.

38. Pianta, R.C., and others

5. Problems lurk and opportunities abound

Proponents of expanding the role of rigorous evidence in policy choice, not least the two of us, are optimistic about the Obama initiatives. However, it would be a serious mistake to accept the initiatives uncritically and to assume that they will greatly improve the quality and impact of the nation's social programs. For this reason, we turn now to a discussion of several potential problems and issues with the Obama approach, each of which is dealt with in separate sections. In addition, we outline several opportunities to strengthen the evidence-based approach to policymaking.

Politics

Those of us who are hopeful that an enhanced role for evidence will greatly improve the quality and impacts of social programs have sometimes been thought to regard the emphasis on evidence as the repeal of normal politics. Let the experts decide, based on their evidence, what programs should be funded and at what level. But that is not at all the way the two of us view evidence-based policymaking in general and the Obama initiatives in particular. For many years, our major goal has been to increase the role of evidence in political decision making. We are both fully aware, however, that constituent views, the positions taken by party leaders, the political philosophy of elected officials, the positions politicians have taken on similar issues in the past, their campaign promises, and the inevitable political compromises necessary to pass legislation will always play a huge role in political decision making. Against this backdrop of politics as normal, which almost always places political considerations above the usually limited role of evidence, expanding the influence of evidence would be useful and could lead to better decisions. Even so we have no doubt and do not regret the fact that politics will almost always play a determining role in policy choice. We want evidence to be important, not dominant.

Implementation

The field of implementation research is not as advanced as the field of evaluation research. Consider the example of Head Start, outlined earlier. Since at least the late 1950s, rigorous research has shown that high-quality preschool programs can have major impacts on the development of poor children.³⁷ Random-assignment longitudinal studies show that children who have attended quality preschools, as compared with similar peers who stayed at home or attended regular day care

- (2009) *The Effects of Preschool Education: What We Know, How Public Policy Is or Is Not Aligned with the Evidence Base, and What We Need to Know.* 'Psychological Science in the Public Interest.' 10(2): 49-88; Consortium for Longitudinal Studies (1983) *As the Twig Is Bent: Lasting Effects of Preschool Programs.* Hillsdale, NJ: Lawrence Erlbaum.
39. Puma, M. and others (2010) 'Head Start Impact Study: Final Report.' Washington DC: US Department of Health and Human Services.
40. Head Start is not a test of the Abecedarian or Perry Preschool models. Rather, it is a funding stream that supports a variety of preschool programs that have the same general goals but vary along many dimensions and vary greatly in quality.
41. Ioannidis, J.P.A. (2005) *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research.* *Journal of the American Medical Association.* 294 (2): 218-228; Zia, M.I. and others (2005) *Comparison of Outcomes of Phase II Studies and Subsequent Randomized Control Studies Using Identical Chemotherapeutic Regimens.* *Journal of Clinical Oncology.* 23 (28): 6982-6991; Chan, J.K. et al. (2008) *Analysis of Phase II Studies on Targeted Agents and Subsequent Phase III Trials: What Are the Predictors for Success.* *Journal of Clinical Oncology.* 26 (9): 1511-1518.
42. Garet, M.S. and others (2008) *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement.* Washington DC: Institute of Education Sciences, US Department of Education. Available at: <http://ies.ed.gov/ncee/pubs/20084030/>
43. Bloom, H.S., Michalopoulos, C., and Hill, C.J. (2005) *Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects.* In *Learning More From Social Experiments: Evolving Analytic Approaches.* New York: Russell Sage Foundation, pp. 173-235; Cook, T.D., Shadish, W.R., and Wong, V.C. (2008) *Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons.* *Journal of Policy Analysis and Management.* 27(4): 724-50; Glazer, S., Levy, D.M., and Myers, D. (2003) *Nonexperimental versus Experimental Estimates of Earnings Impact.* *The American Annals of Political and Social Science.* 589: 63-93.
44. Auerbach, A. and Gale, W.G. (2010) 'Deja Vu All

facilities, perform better on standardized tests, have fewer placements in special education, are less likely to be arrested as juveniles or young adults, are more likely to graduate from high school, and so forth.³⁸ Yet after nearly half a century, a recent rigorous, multi-site study of Head Start, a broad-scale implementation of preschool intended to boost the development and school performance of poor children, showed that its impacts were barely detectable at the end of first grade (i.e., three to four years after random assignment).³⁹ This seems to be the story of many intervention programs: significant impacts when tested by their designers and on a small scale but modest or no impacts when replicated (often in a much different, more diluted form) on a broader scale.⁴⁰

The lesson here is that implementation is at least as important as program development. If the Obama initiative results in a significant share of federal intervention dollars being spent on programs that have proven track records, that will stand as a great achievement. But experience teaches that poor implementation of good models often fails to produce impacts. Thus, it is to be hoped that the various teams in the agencies implementing the Obama initiatives will focus like a laser on implementation issues. These issues include how to ensure that staff receive adequate training in the intervention approach, how to ensure that staff continue to follow the major features of the intervention program, how to achieve an adequate dosage of the intervention with all program participants, and how to continually monitor and evaluate outcomes. Methods for achieving these and other essential features of quality implementation could be one of the most important outcomes of the Obama initiatives. We are at last focused on evidence of program impacts; now we need to be equally focused on program implementation methods to produce those impacts on a broad scale.

Selecting strong models for scale-up

Some of the Obama initiatives have, by design, selected models for which the supporting evidence is only moderate or preliminary in nature, with the goal of testing them in more rigorous evaluations to determine whether they work. Moderate or preliminary evidence includes nonrandomized comparison-group studies ('quasi-experiments'), or randomized controlled trials with only short-term follow-up, assessment of intermediate rather than final outcomes (e.g., condom use versus reductions in teen pregnancies, abortions, or births), or other key limitations in study design or implementation. These studies can be valuable for decision making in the absence of stronger evidence. Too often, however, findings from such initial studies are overturned in large, definitive randomized controlled trials. Reviews in medicine, for example, have found that 50 per cent to 80 per cent of promising results from quasi-experiments or preliminary 'efficacy' trials are overturned in subsequent randomized controlled trials.⁴¹ Similarly, in education, nine of the ten major randomized controlled trials sponsored by the Institute of Education Sciences since its creation in 2002 have found weak or no positive effects for the interventions being evaluated – interventions which, in many cases, were based on promising quasi-experiments or preliminary trials (e.g., the LETRS teacher professional development program for reading instruction).⁴² Systematic 'design replication' studies comparing large, well-conducted randomized controlled trials with quasi-experiments in welfare, employment, and education policy also have found that many widely-used and accepted quasi-experimental methods produce unreliable estimates of program impact.⁴³

That some of the Obama initiatives – including teen pregnancy reductions and home visiting – may fund program models with only moderate evidence of effectiveness leads us to fear that a number of the models, when rigorously evaluated, will be found to produce weak impacts. But at least a few will likely be found to produce meaningful impacts on important policy outcomes. If the administration can use strong evaluation designs to weed out the program models that produce weak impacts in the initial program years, and reinvest the money saved from these models in models with stronger evidence of success, the Obama initiatives may evolve toward increasing effectiveness over time.

Changing the agency culture

As the episode on the Science, Technology, Engineering, and Math (STEM) programs above demonstrates, federal administrative agencies have often failed to emphasize the importance of evidence in determining whether their programs are producing worthwhile results. Indeed, some agencies appear to be conflicted about evaluation because so many programs that are subjected to rigorous evaluation are shown to produce null results. No agency wants to be administering programs that are known to be ineffective. Moreover, a major rule in the federal government is that power and influence are based on big budgets. If agencies want to expand their programs and their

budget, they need Congress to believe they are conducting effective programs that are providing benefits to the nation. It is a good bet that if every type of federal program received the scrutiny that OMB and the Coalition for Evidence-Based Policy provided for the STEM programs, many federal programs would be exposed as not better than moderately successful or even failures.

A vital part of the Obama evidence-based initiatives is to change the relationship between OMB and the federal agencies so that OMB is a taskmaster for evaluating programs by the use of rigorous evidence. Given the history of the Obama evidence-based initiatives reviewed above, it appears that OMB has already become a strong supporter of rigorous evaluation and is now providing strong leadership – not to mention a source of funding – for federal agencies to evaluate their programs using rigorous methods. This is the role of OMB envisioned by the Bush administration's PART initiative. The modest role of PART in the Bush administration has been expanded and, in a different form, PART's approach to rigorous evaluation now seems to be having an impact on a number of federal agencies that administer intervention programs. In the long run, if federal agencies do not become true believers in rigorous evaluation and transparency about the effects of their programs, the impacts of evidence-based policy making will be minimal.

The exploding deficit

Like the UK, the US is now running huge federal deficits – and as the retirement of the baby boom generation gets into full swing in the years ahead and health care costs continue their relentless rise the deficit will get even worse. According to a realistic baseline of US spending over the next decade,⁴⁴ the deficit will average a trillion dollars a year. One consequence of such profligacy is that by 2020, interest payments on the federal debt will approach \$1 trillion. Although budget hawks have been sounding the alarm for nearly a decade,⁴⁵ Congress and the President are just now beginning to get serious about cutting spending and raising revenues. The first major confrontation between the political parties occurred in the context of the 2011 budget. As part of its package of spending cuts, the House enacted legislation that would have zeroed out the Obama teen pregnancy prevention initiative and the agency responsible for administering the Social Innovation Fund initiative. Both were eventually restored, but more of the same lies ahead. Republicans are now talking about cutting trillions in spending – and in truth, even if revenues are part of a compromise deal, it seems likely that spending would be cut by \$2 trillion or more over ten years. This level of cutting would pose great danger for the Obama initiatives. Even initiatives that are already being implemented, such as the teen pregnancy prevention initiative, are subject to cutting or even termination. In the face of such a major deficit-reduction effort, the argument that evidence can improve social programs tends to lose its force.

There is not much that can be done about the precarious funding of Obama's six evidence-based initiatives. Miraculously, they all survived the 2011 budget fight. The administration may come to the moment when program triage is necessary. As part of a deal that involves big cuts in spending, perhaps the administration can protect three or four of the initiatives while sacrificing the others. In any case, it seems wise to us for the administration to begin figuring out a fallback position on protecting as many of the evidence-based initiatives as possible from the tsunami-like forces moving Congress toward deep spending cuts. Even when spending is under the knife, as we have seen, evidence can prove useful.

6. Conclusion

The federal government and state governments in the US spend tens of billions of dollars each year on social programs that have been shown to produce modest results or worse. In other cases, billions of dollars have been spent on programs and funding streams for many years, and yet little is known from rigorous evidence about whether the programs are producing good outcomes. Meanwhile, social scientists in the US have developed increasingly sophisticated and reliable methods for evaluating program impacts, and the nation's universities and large and widely respected research companies have achieved significant experience in designing and carrying out large-scale, multi-site evaluations using random assignment. Program administrators, especially in the federal Office of Management and Budget – but in other federal agencies as well – have increasingly emphasized the importance of obtaining rigorous evidence about the impacts of

Over Again: On the Dismal Prospects for the Federal Budget.' Washington, DC: Brookings.

45. Rivlin, A.M. and Sawhill, I.V. eds. (2004) 'Restoring Fiscal Sanity: How to Balance the Budget.' Washington, DC: Brookings.

46. As early as 1996, the Welfare Reform Law

their programs. To a limited but probably growing degree, the federal Congress has even required administering agencies to fund rigorous evaluations of their programs.⁴⁶

gave the Secretary of the Department of Health and Human Services the authority and funding to conduct evaluations of welfare reform and of other provisions in the legislation. The law gave the Secretary wide discretion to select important research topics, but the statute directs the Secretary: "to the maximum extent feasible, [to] use random assignment as an evaluation methodology." See Section 413 of Title IVA of the Social Security Act (42 U.S.C. 613).

But the Obama evidence-based initiatives analyzed here are opening a new chapter in the generation and use of evidence by the federal government. Subsequent administrations may not have the same commitment to evidence that senior officials in the Obama administration have, but to a considerable degree the Obama emphasis on evidence is now being institutionalized in the federal agencies. At least three cabinet-level agencies (HHS, Education, and Labor) as well as the Corporation for National and Community Service are now administering multi-year evidence-based initiatives. Of greatest importance, OMB has become a strong advocate for generation and use of rigorous evidence and, beginning at least with the PART initiative during the Bush administration, has been leading all the federal departments that administer social intervention programs to generate and use rigorous evidence.

The existence of the six Obama initiatives, the actions of OMB over at least a four or five year period, the increasing use of evidence by federal agencies, and the approval and funding of evidence-based initiatives by Congress lead us to believe that the role of rigorous evidence in federal policymaking and program implementation is here to stay.

NESTA

1 Plough Place
London EC4A 1DE

research@nesta.org.uk
www.twitter.com/nesta_uk
www.facebook.com/nesta.uk

www.nesta.org.uk

Published: September 2011