
3 With a scope so wide: using evidence to innovate, improve, manage, budget

Ron Haskins

Brookings Institution

Abstract

Evidence from social science research and evaluation is used for at least two broad purposes in improving program effectiveness. One purpose is to influence the decisions of policy makers; a second purpose is to contribute to continuous program improvement by influencing program management and implementation. This paper examines the role of evidence in three recent episodes of policy making in the United States: the welfare reform legislation of 1988 and 1996; funding for after-school programs in 2003; and the ongoing debate over establishing a new federal home-visiting program. The examples demonstrate that, although evidence is often not a primary factor in policy debates, in some cases its role can be important if not decisive. In reviewing these examples, characteristics of evidence that make it most likely to be useful to policy makers are examined. The paper concludes with a brief review of several guidelines for using evidence to improve program management and implementation.

3.1 Two uses of evidence in policy formulation

This paper has a split personality. The dominant personality wants to examine the influence of high-quality evidence from program evaluations on the formation and enactment of social policy. Three interesting examples from policy making in the United States are examined, although the paper focuses on welfare reform legislation enacted in 1996 after a long and bitter debate in Congress that ended the entitlement to cash welfare and required work by all welfare recipients. The examples serve two purposes. First, they illustrate how evidence from social science research and evaluation can have an impact on policy — or not. Second, the examples suggest several generalisations about how good evidence can be developed, communicated, and used to influence policy.

But another, less confident, personality wants to examine whether evidence can be built into a broad management system that a government can use to bring continuing attention to what it is trying to do, how it is trying to do it, how to improve it, and how to pay for it. Confining attention to the United States, I am well aware of the many thoughtful schemes that have been previously implemented to improve government effectiveness and efficiency. These include ‘program planning and budgeting’, ‘zero-based budgeting’, the Government Performance and Results Act, and the Program Assessment and Rating Tool. In every case, there seems to be general agreement that results have fallen short of expectations (Radin 2000).

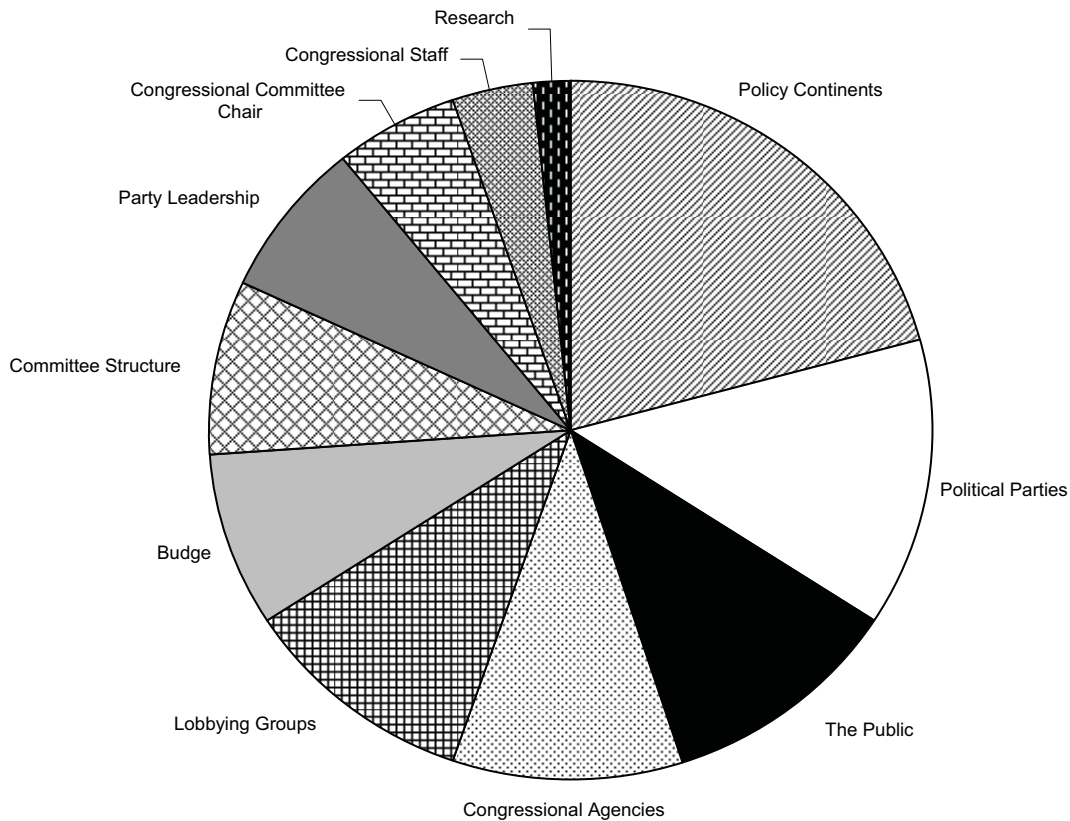
Throughout, I pursue a strategic question that reformers must face: Is it better to try to launch a government-wide initiative that attempts to bring order across a huge range of agencies and programs using a top-down initiative or to focus most attention on individual programs and how to improve them in an initiative that starts from the bottom and leaves open the question of when and whether the initiative should try to build towards something that involves many agencies and programs?

In the end, I argue that we should devote most of our attention to program evaluation and using evaluation results to enact, implement or improve individual programs. Although broader schemes seem reasonable, experience shows that putting them into practice has certain costs and uncertain benefits. Even so, experience and research also suggest that improvements may be possible and that evidence could become an important part of a broad movement to manage and budget wisely.

Applying evidence to policy choice

Consider the somewhat whimsical portrayal of the factors that influence policy formulation portrayed in figure 3.1. Although this paper focuses on research, the policy process often does not. Having participated in policy formulation and enactment directly from inside the US House of Representatives and the White House for a decade and a half, and having observed — and occasionally tried to influence — from outside for more decades than I would care to discuss, experience leads me to observe that research rarely drives policy. Of the factors that compete with research as prime influences on policy debates, at least three are typically of much greater importance than research: policy continents and inertia; the philosophy of political parties; and powerful politicians.

Figure 3.1 **Factors that influence policy choice**



Data source: Illustrative proportions only; Ron Haskins, The Brookings Institution.

Policy continents demand a word of explanation. All the industrial democracies have highly developed social policies for the elderly, children, the disabled, and the poor and destitute. Those policies are typically embodied in thousands of pages of legislation and regulation. In the United States, many of our most important social programs — including cash welfare, adoption and foster care, medical care, child support enforcement, welfare-to-work programs and programs for the disabled — are located in the Social Security Act (CWM 2005). Each line on every page of this wonderful statute is watched by hawk-like individuals and interest groups. Many of them are looking for an opportunity to change the statute; many are guarding the statute so no-one else can change it. Thus, the statutes, usually developed over many decades, own a kind of energy to ensure that they are not changed without a fight. If someone wants to change them, they had better be armed. Like continents, important statutes move slowly.

Perhaps an even more powerful influence on policy formulation is political philosophy. I am not well versed in the politics of other industrial democracies, but

in the United States the two political parties offer a great contrast in philosophies, which translates to a striking difference in agendas — and a continuing feast of arguments in the federal capital and state capitals all over the nation. There are exceptions, but in general Democrats are the party of big government and high taxes, especially on the rich — the protestation of President Bill Clinton notwithstanding (CNN 1996). At this moment, Democrats are trying to reform health care to achieve universal coverage and to feature a government insurance plan that would be available to all. Government already pays for half of the medical care in the United States, but if reform passes the share of health care paid for by government will expand even faster. Conservatives — and not just conservatives — are fearful that, if the reforms include a government insurance program open to all, Democrats will subsidise the plan with public funds and drive private insurance companies out of the market. Government would then be bigger, taxes would have to be raised, and government could have a health care monopoly after a decade or so. In effect, the long hoped-for introduction of competition in the American health market would be dead. All this is taking place in an atmosphere in which the federal government has already taken control of two of the nation's three largest car companies and major parts of the financial system, and is attempting to exert much greater control over the nation's energy infrastructure.

Democrats are now putting another one of their philosophical tendencies on display. Despite the fact that the upper 5 per cent of earners already pay 60 per cent of federal income taxes and 75 per cent of corporate taxes (CBO 2007), Democrats plan nonetheless to increase tax rates on people earning over \$250 000. In addition, despite the fact that the bottom 40 per cent of households already pay negative federal income taxes (they get a cheque from government in the mail each year), Democrats hope to lower the taxes of those households further (or send them a bigger cheque) by creating a new tax credit for workers (Tax Policy Centre 2009).

In contrast to Democrats, Republicans want smaller government and lower taxes for everyone (some would add 'especially the rich'). The reputation of Republicans for smaller government took a major hit under President Bush when they enacted a huge and unfunded expansion of the Medicare program by providing the elderly with a drug benefit, but their reputation for tax cuts was dramatically reinforced. Bush cut taxes more than any other president, even when it became obvious that the federal government would once again run very large deficits. Republicans even went so far as to produce one-sided budget rules that would apply strict limits to spending but not to tax cuts. One could conclude that Republicans were willing to sacrifice their reputation for fiscal responsibility in order to cut taxes.

The point is that, in the face of these fundamental goals of the two major parties, evidence from social science research has little to no chance of shifting decisions.

Budget projections, for example, are a kind of social science, although the artistic side of projections must be admitted — especially, as Mark Twain observed, when projections involve the future. Yet the long-term projections of fiscal doom for the federal budget have been largely ignored by elected officials from both parties (Antos et al. 2008). Even the Obama administration, headed by a president who has pledged to cut the deficit in half and a budget director who has been a fiscal hawk his entire career, has increased the federal deficit to the previously unheard of level of over \$1.17 trillion for 2010 (OMB 2009, table S-2, p. 114) (partly justified by Keynesian spending to rescue the nation from the worst recession since the Great Depression) and at least \$1 trillion per year thereafter (with no justification known to man or God) (Auerbach and Gale 2009). Evidence that these deficits are out of all proportion to any reasonable definition of what government should be doing have no impact. Evidence that past deficits have been associated with inflation and high interest rates is ignored. Democrats want to expand programs and add new ones; Republicans want to cut taxes. Who cares that our children and grandchildren must pay the bill? Nothing could provide stronger evidence of the extent to which our political parties are driven by their philosophies than their joint willingness to bankrupt the nation to achieve their philosophical ends.

In addition to the power of the statutory continents and the political philosophies and historical goals of our major parties, another force that is more powerful than evidence is the personal views and political strength of individual politicians. Most people familiar with the American system of government have an appreciation for the power of the president, but congressional party leaders and committee chairs, many of whom are barely known to the public, are also immensely powerful — sometimes so powerful they can modify or even defy the agenda of the president. Committee chairs have two powers that are especially useful for exercising their legislative muscle. First, they can call a public hearing of the committee and then invite witnesses that will present the views and policies favoured by the chair. In this way, a savvy chair can shape the political debate to tilt toward her favoured outcome. Second, when chairs decide to write a bill, they can author the first draft of the bill themselves. If they craft their bill properly so that it reflects (or at least does not flagrantly violate) the public will and the views of members of their own party and if they are clever in building support for their legislative goals, the bill can often survive its long legislative journey with many or even most of its major provisions more or less intact. Thus, if the chair wants a particular provision because his constituents favour it or because his financial backers want it or because his political philosophy demands it, the state of evidence for or against the provision is of modest, if any, concern.

I note in passing that a president, chairman, or other powerful political figure who thinks evidence should be a major consideration in program enactment and funding

could greatly increase the role of evidence in policy formulation. Indeed, as we will see, President Obama and his budget director, Peter Orszag, appear to illustrate this point in spectacular fashion.

My goal in this opening section is to lower expectations about the potential role of evidence in the policy process because I think it should be clearly understood that evidence will only rarely be a dominant force in a debate over policy formulation. As suggested by figure 3.1, there are simply too many powerful forces that operate above and beyond evidence. When I left the scholarly world to work in the US Congress, I thought it was important for the social science community to try to convince elected officials that their decisions should be shaped by evidence whenever possible. But it did not take long to realise that even the members of Congress most disposed to pay attention to social science evidence simply regarded it as one small room in the mansion that is political debate and decision making. When occasions arose on which I had the opportunity to advise members about votes on social programs, my approach was to make them aware of whether there was evidence for or against a particular program and some idea of the quality of the evidence. Over the years, I dropped the idea that, if only those rascal politicians had their heads screwed on straight, they would listen to evidence and even seek it out. I regarded my role as being one of slightly expanding the purview of members to include at least some attention to evidence.

Arguably, the best way to bring evidence to the attention of members of legislative bodies is to have experts in organisations that elected officials trust report to them on a regular basis (table 3.1). In the US capital and in nearly all the state capitals, there is at least one organisation that helps legislative bodies keep track of their budgets and the cost and budget impacts of specific pieces of legislation. Either that budget organisation or other agencies also provide advice to members and committees about whether programs are working and whether new legislation is consistent with evidence about previous programs.¹ The US Congress has three agencies of this type that provide reliable and nonpartisan advice: the Congressional Budget Office, the Congressional Research Service and the Government Accountability Office (formerly the General Accounting Office; see table 3.1). These organisations specialise in bringing evidence to bear on the policy process. Many of their senior staffers have advanced degrees and know a great deal about the programs in their purview, as well as about legislative procedure. All three

¹ A good example of a state organisation of this type is the Washington State Institute for Public Policy. Created by the Washington state legislature in 1983, the institute ‘carries out practical, non-partisan research — at legislative direction — on issues of importance to Washington State’. Issues studied by the institute include education, criminal justice, welfare, children and adult services, health, and general government. See www.wsipp.wa.gov/default.asp.

organisations have established a tradition of being nonpartisan and of presenting neutral analyses of budget and program issues. Importantly, most members of Congress trust them. These organisations, in short, are in position to bring neutral and evidence-based positions to the attention of Congress. But, of course, it is well beyond their power to make members pay much attention to the evidence. Even so, democracies need these neutral and respected agencies in order to know the financial implications of pending legislation and to maximise the chance that evidence will find an important role in policy debates and decisions.

Table 3.1 Congressional agencies that provide non-partisan analyses

<i>Agency</i>	<i>Year founded</i>	<i>Mission</i>
Congressional Research Service (CRS)	1914	CRS supports an informed national legislature by developing creative approaches to policy analysis, anticipating legislative needs, and responding to specific requests from legislators in a timely manner. CRS provides analysis that is authoritative, confidential, objective and nonpartisan.
Government Accountability Office (GAO)	1921	GAO helps Congress improve the performance and ensure the accountability of the federal government by providing Congress with timely information that is objective, fact-based, nonpartisan, nonideological, fair and balanced.
Congressional Budget Office (CBO)	1974	CBO provides Congress with objective, nonpartisan and timely analyses to aid in economic and budgetary decisions on the wide array of programs covered by the federal budget and with any information and estimates required for the Congressional budget process.

Sources: See <http://www.loc.gov/crsinfo/>; <http://www.gao.gov/about/index.html>; <http://www.cbo.gov/aboutcbo/factsheet.shtml>.

As it happens, the United States is now in the midst of an episode that nicely illustrates the vital role played by one of these nonpartisan analysis agencies. President Obama and the Democrats are in the midst of a serious attempt to bring universal health care to the nation. Because of the nation’s yawning deficit, the President and his budget director, Peter Orszag, have promised to pay for any health care expansions that are enacted. However, as almost always happens, when the bills began to work their way through Congress in June and July 2009, it became increasingly obvious that the Democratic majority — precisely like the Republican majority when it enacted the elderly drug benefit in 2002 — found it too difficult to include the tough measures that would provide the financing for their new health benefits. When the first bill emerged from committee, amidst great claims of success by Democrats and the President, Douglas Elmendorf, the Director of the

Congressional Budget Office (and himself a Democrat who had served in the Clinton administration), testified that the bill was not fully financed and would increase the nation's already spectacular deficit by nearly \$240 billion over 10 years. For this piece of honest analysis, he was berated by leading Democrats. Harry Reid, the Democrats' Majority Leader in the Senate, said that Elmendorf should run for office if he wanted to play such a decisive role in legislative battles — never mind that making cost estimates is the single most important job of the CBO director. Reid was probably upset because he knew that Elmendorf and the CBO have much more credibility on budget issues than he and other elected officials do. Not surprisingly, the media accepted Elmendorf's estimate and told the nation that the Democrats were trying to pull a fast one (Montgomery and Murray 2009; Pear 2009). It is precisely for situations like this that 'Anonymous' invented the old aphorism about the king's new clothing.

Having sufficiently lowered expectations about the political power of evidence, I now argue that evidence can nonetheless play an important role in policy choice under some circumstances. More specifically, in what follows I describe three legislative struggles in which evidence from social science research played an important role in policy debate. The first example is from the seminal welfare reform legislation of 1996, in which several ways that evidence influences policy were on display. The second is from a much more concise episode, in which evidence formed the basis for a policy choice by the Bush administration and in which the evidence was quickly swept aside and cast into a deep and ignominious grave. The third is still ongoing and, like welfare reform, shows the power of evidence to play an important role in a major congressional debate.

The Welfare Reform Law of 1996

Much of the discussion in this section is based on *Work over Welfare* (Haskins 2006). I had the good fortune of being a Republican staffer on the Ways and Means Committee² in the US House of Representatives when Republicans won control of both houses of Congress in the elections of 1994. Thus, I had a ringside seat for the festivities that followed. A small group of Ways and Means Republicans had been working on welfare reform for three or four years before 1994, had formed a coalition that involved Republican leaders in the House and on other committees, and had introduced several bills. Of course, House Republicans had been in the minority in the House for four decades and, as usual, their bills were ignored. But

² When referring to the Ways and Means Committee, scholarly and press reports often use the phrase 'the powerful Ways and Means Committee'. I was on the committee staff for four months before I figured out that the committee's official title was not 'Powerful Ways and Means Committee'.

once Republicans took over control of both houses of Congress in 1994, the importance of their bills and the ideas they represented took a quantum leap forward.

The opportunity for Republicans to fundamentally reform welfare was created by none other than the great moderate Democrat, Bill Clinton. In the presidential election of 1992, Clinton had led the nation to believe that he was going to ‘end welfare as we know it’ (Weaver 2000), but then failed to follow up on his promise during his first two years in office. Instead, he squandered much of his prestige and power on a health care reform bill that failed completely. He did send a thoughtful and sweeping welfare reform bill to Congress late in 1994, but the senior leadership of the Ways and Means Committee, joined by the House Democratic leadership, did not like the bill and simply ignored it — a clear example of the power of senior congressional leaders to shape the fate of legislation, even when it is sponsored by the President.

Unlike Clinton, after the 1994 congressional elections that led to such a surprising Republican victory, Ways and Means Republicans had a welfare reform bill ready to introduce. Even more to the point, Republicans in the House were united on almost all the major welfare reform issues, leaving no doubt that they would be able to pass a bill in the House. This they did in short order, although the bill received hardly any support from Democrats (194 House Democrats voted against the bill; 9 voted in favour). In the record time of less than three months from the opening of the congressional session, House Republicans sent their bill to the Senate. After an exciting and drawn-out drama typical of the Senate, a bill somewhat more moderate than the House bill was passed on a surprisingly strong bipartisan vote in the autumn of 1995 (only 11 Democrats voted against the bill). A House–Senate conference committee then worked to resolve the differences between the two bills, and a version of the bill was sent to President Clinton in December as part of a huge budget bill designed to reduce the deficit. This compromise welfare reform bill, closer to the House than the Senate bill, lost much of its bipartisan support in the Senate. In addition, the budget bill of which the welfare bill was only one part was strongly opposed by Democrats. Thus, the lack of Democratic support for the huge bill in Congress gave cover to President Clinton to veto it. After the President’s veto in December 1995, Republicans extracted the welfare reform bill from the budget bill and sent welfare reform separately to the president. Clinton vetoed the second bill as well. After the president’s second veto, congressional Republicans worked with governors on a bipartisan basis, made some important changes in their bill, and then passed the revised bill on strong bipartisan votes in both the House and the Senate (about half the Democrats in both the House and the Senate supported the bill). President Clinton then signed the bill into law on 22 August 1996 (Haskins 2006). The third time was the charm.

I know of no example in which evidence played such a decisive role in a major legislative battle. By far the most important role played by evidence was already obvious when Democrats, joined by a large majority of Republicans and President Reagan, enacted a welfare reform bill called the Family Support Act in 1988. The 1988 Act, like the more radical welfare reform legislation passed eight years later, was designed primarily to boost work by welfare applicants and participants. Even in the highly contentious debate of 1996, as well as in the calmer debate of 1988, almost everyone believed that programs designed to increase work would actually lead more welfare mothers into the labour force and save money. Why did members of Congress believe work programs could be so effective?

The answer is that a body of high-quality research had accumulated, both before and after the 1988 reforms, showing that programs that emphasise helping mothers on welfare find a job and communicate the clear expectation that mothers should work would increase employment and reduce welfare costs. These studies are nicely reviewed by Judy Gueron (2003), the former president of MDRC, a widely-respected research firm that conducted many of the pre-1988 studies.³ I draw attention to four characteristics of this line of research because the studies are vital to understanding how research could have such influence and perhaps suggest ideas for creating future occasions on which research is as pivotal as it was during the welfare reform era in the United States.

The first characteristic, and perhaps the most remarkable, is that the research on welfare reform was of such high quality. The remarkable thing about the quality is that the field of random-assignment evaluations under real-world conditions was in its infancy at the time the studies began in the late 1970s. Before that time, there were no major random-assignment studies on welfare demonstrations to test whether work programs changed the behaviour of recipients.⁴ But, starting with large-scale demonstrations to test various approaches to increasing work levels, a style of research that had immense influence on subsequent welfare evaluations began to take shape (Gueron and Pauly 1991). Over the next decade and a half or so, numerous studies were initiated that were large-scale; featured random-assignment; were conducted in cooperation with states and with welfare offices within states; were conducted by professional companies that developed great skill in both the design of research of this magnitude and in the human-relations and organisational capacity needed to conduct complex studies of this type; were often

³ For a thorough review of the welfare reform studies, see Besharov (2009b, in press). In the spirit of full disclosure, I am now a member of the MDRC board of directors.

⁴ The income maintenance experiments were an exception, but they did not test job search and similar welfare-to-work programs; see Munnell (1986).

funded by a coalition of government and foundations; and featured tests of reforms that could be generalised to other settings.

These studies often involved changes in programs that were not specifically allowed by federal welfare statutes. What made such studies possible without violating the law was that in 1962 Congress had enacted an obscure statutory provision⁵ that allowed the Secretary of the Department of Health and Human Services to waive provisions of welfare law in order to test innovative programs. Eventually, about 40 states conducted demonstration programs under this waiver provision, and most of those demonstrations experimented with ways to increase work by parents on welfare. The partnership between researchers and state officials who applied for the waivers and who often contributed to financing the demonstrations offered a durable model for additional studies of research on this scale. One especially desirable outcome of this approach is that the research results have a ready-made audience of senior government officials who not only support the research but have a great interest in learning from the demonstrations because they think the results can have wider application and help them achieve the political goals of elected officials.

Another important characteristic of this research is that the demonstrations gave states — and in some cases local government — the opportunity to develop the practical experience and skills needed to implement these big and unwieldy reforms. Imagine two scenarios. In the first, states are required by the federal government to do something they had not done before and with which many people disagree. The states often use the term ‘unfunded mandate’ to refer to federal legislation that requires them to do something that imposes costs on their budgets. Given the level of invective states direct at such federal mandates, one way to ruin a policy that states must implement is to impose a federal mandate on them. In the second scenario, states not only agree with the approach but actually have some experience with developing similar approaches and implementing them by their own bureaucracies. Welfare reform clearly fell into the second scenario. The demonstrations produced a bottom-up movement of people who believed in helping mothers work, had the skills necessary to communicate the work requirement to mothers, and were learning how to build programs that would successfully help mothers find work. Thus, not only did almost every state support the radical reforms under discussion in Washington in 1995 and 1996, but they were already in the process of developing their own programs that, perhaps with some adjustments, would meet the new federal requirements.

⁵ *Social Security Act 1962*, section 1115.

A third characteristic of the demonstration movement was that governors themselves began to take notice of their own state programs. Here is some shocking news: governors don't know everything that is going on in their states, even when the government they head is sponsoring the activity. Most of the demonstrations produced positive impacts on work and saved money — outcomes that governors were bound to like. That got their attention. Many governors, both Republicans and Democrats, began to laud their own programs and to think about and discuss how much more they could do if they had more flexibility from federal requirements. Thus, the governors themselves became an important constituency for welfare reform — and for implementing aggressive programs once the federal legislation passed in 1996.

Finally, it should not go unremarked that MDRC and the other research companies conducting the welfare-to-work experiments developed great skill in bringing attention to their findings. Not only did MDRC give snappy testimony and briefings to important congressional committees, members, and staffers in Washington, DC, and state capitals, but they also developed expertise at working with the media to bring attention to their results. MDRC and a few of the other research companies thereby fulfilled two important goals that are difficult for individual researchers to achieve: they kept up a steady stream of information about successful demonstrations to key players in Washington and state capitals, and they stimulated media stories about the growing successes of welfare reform.

By the time President Reagan and the federal Congress decided to reform welfare in 1988, a background assumption of both Democrats and Republicans was that work programs could help mothers get jobs and leave welfare. There was even a widespread belief that these programs could save money (Long 1988). So popular were these findings that both Democrats and Republicans framed their welfare reform arguments around the claim that their fundamental goal was to advance the work agenda. Never mind that the 1988 bill actually did little to require anyone to work (Haskins 1991). This same understanding of the effectiveness of job search programs also served as a background condition for the debate leading to the 1996 reforms that really did require mothers on welfare to work or suffer rather serious consequences. The 1996 debate produced lots of arguing, some of it bitter, but not about whether programs could successfully help mothers get jobs and thereby fulfil the major goal taxpayers had for welfare reform — requiring work and not permitting long stays on welfare that reinforce idleness. The single greatest achievement of research in the 1988 and 1996 welfare reform debates was to create this background belief, backed by solid data everyone accepted and trusted, that it was possible to increase the employment rates of poor mothers and save public dollars in the process.

I think that almost any reasonable person studying the events I have summarised or who participated in them would agree that research and demonstrations on welfare reform created an important predicate for both the 1988 and 1996 reforms. In a real sense, research was one of the most important factors that created the opportunity for the Aid to Families with Dependent Children (AFDC) program, the nation's major cash assistance program for destitute families since the New Deal of the mid-1930s, to be repealed and replaced by the Temporary Aid for Needy Families (TANF) program. Every elected president from Nixon to Clinton, except the first Bush, had proposed major welfare reform initiatives, but all the initiatives had foundered because they cost too much or failed to find a compromise between encouraging work and allowing continued welfare dependency (Haveman 1995). But the experimental evidence that welfare mothers could and did find work when helped or cajoled into doing so allowed Republicans to argue that they and their children would be better off leaving welfare and, eventually, allowed Democrats to believe that Republicans might be right.

Moreover, the work requirements in the Republican bill were based on the assumption, well supported by research, that if ways could be found to get more mothers to find work, the goal of helping mothers become self-sufficient would be advanced. As shown in table 3.2, the major characteristics of the TANF program designed by Republicans in 1996 were intended to encourage or, when necessary, force mothers into the labour force. Ending the entitlement (the legal right to a cash benefit) sent a strong signal that mothers had to work and also made it easier for states to establish tough work requirements without violating the mothers' legal rights.

Table 3.2 Temporary Aid for Needy Families program requirements

1	End Cash Entitlement
2	Block Grant Funding
3	Work Requirements
4	Sanctions
5	5-Year Time Limit

Source: Haskins (2006).

To make sure that states mounted programs that would require work and that mothers would conform to the work expectation, the law included numerical work standards that states had to meet. To ensure that states would meet the work standards and that individual mothers would comply with the work requirements, the law included financial sanctions on states that did not meet the percentage requirement (50 per cent of their caseload had to be in work programs when the

requirement was fully phased in) and mandatory financial sanctions on mothers who did not cooperate with the state work requirements. None of these specific provisions by which the work requirements were to be implemented had been carefully tested by research, but all were justified indirectly by research because they were designed to make sure states established programs that would implement the work requirement and that welfare recipients would have a legal duty to participate in the work programs.

While research established a common base of knowledge that programs could help mothers on welfare find work, another goal of research is to influence political debate so that data and research become a vital part of the discussion on important public issues. Most people who have reflected on this goal of research would not expect research to be dispositive in settling any particular dispute, but it should insert itself into the debate in such a way that all sides must contend with facts about the policy problem and potential solutions. Let us now turn to a consideration of whether research fulfilled this goal in the case of the 1996 reforms.

A useful approach is to begin by identifying the major arguments Democrats made against the Republican bill and then examine those claims in the light of evidence that was available in 1995 and 1996, when the debate was in full bloom. Democrats used four major arguments against the approach to mandatory work taken in the Republican bill. Perhaps the Democratic argument with the most potency was that welfare reform requiring work would harm children. The original purpose of the AFDC program was precisely to protect children from the ravages of destitution. If Democrats could convince the public that requiring single mothers to work rather than entitling them to benefits, as the AFDC program had done, would harm children, they would strike a major blow against the bill.

During the entire two years of congressional debate, however, I never heard a Democrat cite any study or other kind of evidence that welfare reform would in fact hurt children. There was a rather substantial child development literature on whether mothers' work harmed children, but there were no random-assignment studies and the correlational evidence was not consistent in showing harm to children, even preschool children (Ainslie 1984; Gruber et al. 1994). Further, little of this research had been done on poor mothers. The most potent 'evidence' among developmentalists was essentially a theoretical argument about disrupting the attachment between mother and child if they were separated too often when children were young (Bowlby 1969). But even developmental scientists were not in agreement that mothers' work actually disrupted the attachment bond, and the empirical evidence was modest to weak. In any case, it was rare to see this body of evidence marshalled against the Republican bill on editorial pages, and I do not think any member of Congress cited it during committee meetings or during floor

debates. Democrats simply asserted that children would be hurt by the Republican bill.⁶

This issue shows yet again that evidence is not necessarily a trump card in debates about important social issues. Under any reasonable set of rules for rational debate, assertions such as harm to children should be backed up by evidence. There was some evidence (albeit equivocal) available, but it was not used very often, if at all. Without evidence, Democrats had a hard time shaming Republicans into changing their strong work requirement — especially because millions of American mothers, as well as mothers in virtually all the industrial democracies, had been voting with their feet on this issue for three decades. If millions of mothers, including college-educated mothers and single mothers, were flocking to the workforce, how could Democrats successfully argue that welfare mothers were hurting their children by working when they would also be arguing by implication that millions of other mothers were hurting their children (and when many of these middle class working women were themselves members of Congress or the spouses of members of Congress)? The mass movement by women into the labour force — and even into elected office — had been underway for decades, making it all but impossible to use the harm to children claim as a reason to kill welfare reform.

A second major argument Democrats made against the bill was that there were not enough jobs available for all the mothers Republicans expected to leave welfare for work. Prominent labour economists, such as Rebecca Blank (1995) and Gary Burtless (1995), both Democrats, had written about this issue in some depth. Blank and Burtless agreed that the availability of jobs was not likely to be a problem, but that mothers would be forced by their low level of education to take low-wage jobs, and that their advancement to higher wages would be painfully slow. Blank analysed the growth in jobs generated by the American economy and compared the growth in numbers of people who had jobs during the 1960s, 1970s, and 1980s with the percentage increase in the population (table 3.3). In each decade, the growth in jobs had exceeded population growth. In other words, the fraction of adult Americans with jobs had been increasing for at least three decades. Thus, the best evidence indicated that the Democrats' claim that jobs would not be available was exaggerated. But again, evidence on job availability was not often cited and the evidence did not prevent many Democrats from simply asserting that there were not

⁶ A host of good studies, many based on random assignment, published after the 1996 welfare reform law had passed showed that preschool children were probably helped when their mothers went to work (because they went to decent child care centres where they learned more than they would have learned at home); children of elementary school age were neither helped nor hurt when their mothers went to work; and adolescents may have been modestly harmed, as indicated by slight increases in trouble at school when their mothers went to work. See Morris et al. (2005).

enough jobs, while perhaps adding an anecdote about how hard it was to find jobs in their state or district.

Table 3.3 Changes in US population and employment, 1960 to 1990

<i>Year</i>	<i>Percentage population increase</i>	<i>Percentage employment increase</i>	<i>Increase in employed workers (millions)</i>
1960–70	16.9	19.6	12.9
1970–80	22.4	26.2	20.6
1980–90	12.1	18.7	18.6

Source: Economic Report of the President (1994, table B-33).

The third argument Democrats made against the bill was that mothers on welfare should get more education before being thrown into the job market where they would earn only low wages. It was certainly obvious, as Blank and Burtless showed with persuasive evidence, that the overwhelming majority of welfare mothers would receive low wages. Even so, despite low wages, most welfare mothers who worked even half-time would be financially better off than they had been on welfare because of cash benefits from the Earned Income Tax Credit, health coverage, food stamps and child care subsidies from the nation’s growing work support system (see below), which provided cash and in-kind subsidies to low-income working families, especially those with children (Coe et al. 1998; Haskins and Sawhill 2009, ch. 9).

But there was an even more fundamental problem with the argument that mothers on welfare should get education and training before being required to work. Despite around \$240 billion spent on employment and training programs between 1968 and 1995 (Burke 2003), reviews of this literature showed that most of the programs produced modest impacts, if any. As Burtless (1995, p. 100) concluded, based on a careful review of these studies, ‘Even though training can improve the earnings prospects of women who are dependent on AFDC, it will not cause enough of an improvement to remove many low wage single mothers from poverty.’ Other reviewers were even less generous in their conclusion about the impacts of employment and training programs than Burtless (LaLonde 1995).

There were other problems with the plea by Democrats for more education and training. Demanding that welfare mothers be trained before they could be required to work overlooked the fact that the public had already paid many thousands of dollars providing a public education to welfare recipients. True, many mothers had dropped out of high school, but dropping out was a personal decision. Moreover, there were substantial public subsidies available for both education and training that mothers on welfare could get on their own. If mothers on welfare wanted education and training, there were many federal and state programs that could help them

(Crawford 1994). But most welfare mothers did not want training. Perhaps they would select education or training if the alternative was work at \$7 or \$8 an hour, but mothers on welfare were not beating down a path to get training on their own. Thus, the evidence that training resulted in better jobs was weak, lots of other Americans were working for low wages, and mothers on welfare were not clamouring for more education and training. What taxpayers and many policy makers wanted was for them to fulfil their duty to avoid welfare by finding work.

Even if the federal Congress had decided to spend additional billions on employment and training, it is doubtful that most welfare mothers would have gotten jobs that paid much higher wages. Nonetheless, even if they knew the research literature on the tepid results from employment and training programs, most Republicans were reluctant to cite this literature because it seemed to place them in the position of arguing against education. The typical Republican response when Democrats urged more education and training was simply to assert that welfare reform was about employment. Mothers should be required to work and gain some work experience.

The debate over work vs. education was fascinating because Republicans could accurately have made a research-based argument that Democrats were simply wrong to think that employment and training programs would provide much of a boost to the economic value of welfare mothers in the labour market. Although a few Republicans did make this argument, mostly Republicans preferred to simply assert that the goal of welfare reform was to increase work, not education. Indeed, Republicans often argued, education got in the way of work.⁷

A fourth argument made by Democrats was that, even if mothers found work, their wages would be so low that they could not survive. The problem with this argument was that it ignored the mathematics of the value of benefits available to low-income working mothers outside the welfare system compared with the very modest benefits available to mothers while they were on welfare. Consider the figures in table 3.4. The typical mother on welfare who did not work in 1995 on the eve of the welfare reform debate would receive \$7774 in cash and food stamps and she and her children would be covered by Medicaid. By contrast, if the mother took a job for \$7.25 an hour, she would have earnings of \$15 000, a cash payment of \$2842 from the Earned Income Tax Credit, and \$1538 in food stamps, for a total income of \$19,280 (\$17 812 after deducting taxes). A little known fact before the welfare debate, often discussed publicly by Republicans during debate on the legislation,

⁷ Even so, the final bill did allow some education to count as work. Ignoring lots of complications, states were allowed to fill roughly 30 per cent of their work requirement by placing mothers in education and training programs. Despite this provision of the bill, few states chose to put very many mothers in education and training programs.

was that since the mid-1980s or so the federal government had created or modified a series of programs that provided support to low-wage working parents. The Earned Income Tax Credit had been dramatically expanded, Medicaid had been modified to cover children who were not on welfare, food stamps were available to low-wage workers, and the 1996 welfare reform legislation approximately doubled the amount of money available for child care. Thus, mothers could greatly increase their income by taking low-wage jobs, even if they had \$3000 or \$4000 in work expenses.

Table 3.4 Is low-wage work a good deal?

Work vs. welfare, 1996

<i>Income category</i>	<i>Mother on welfare</i>	<i>Mother working (\$7.25 an hour)</i>
	\$	\$
Earnings	0	15 000
Earned Income Tax Credit	0	2 842
Social Security Tax	0	1 148
State Income Tax		420
Federal Income Tax		0
Aid to Families with Dependent Children	5 052	0
Food Stamps	2 722	1 538
Medicaid	Yes	Partial ^a
Total (net of taxes)	7 774	17 812

^a The mother had one year of Medicaid coverage, and her children were covered as long as she had low income.

Note: Even if the mother has \$4000 or \$5000 of work expenses, she would still be better off working, especially because states had extensive funds available to pay for child care.

Source: Committee on Ways and Means, (1996 *Green Book*, p. 399).

Research played a major role in setting the stage for welfare reform by convincing nearly everyone that good welfare-to-work programs could help mothers find employment and save federal and state welfare dollars. But the impact of evidence on the specific arguments used by those who opposed or supported the federal reform legislation in 1996 was modest at best. Most members did not use evidence, even when it was available, to buttress their case, let alone carefully consider research evidence in arriving at their position on welfare reform and its various features. Plato's republic of philosophers who make wise policy decisions for society has not yet arrived.

Using muscle to overcome a negative evaluation: the case of after-school programs

If welfare reform provides a mostly positive example of evidence playing an important role in creating agreement on the efficacy of welfare-to-work programs, the 2003 Mathematica evaluation of the 21st Century Community Learning Centers program provides a far less positive example of evidence influencing policy. Initiated by the federal government in 1998 and funded at \$1 billion per year, the 21st Century program provides after-school care that includes the opportunity for homework and recreational activities for elementary school and middle school students in the afternoon after regular school hours. The typical 21st Century program is open five days a week for around three hours a day. Eighty or so children are enrolled in most of the programs, which are operated by certified teachers and teachers' aides. A typical schedule is that students arrive after school, receive a snack, participate in supervised homework, and then participate in organised activities. Some centres offer activities in martial arts, fitness, dance, art and music. The goals most frequently cited for the program are to provide a safe place for children in the afternoon and to help students improve their academic skills. Several reports based on a large-scale random-assignment evaluation by Mathematica showed that students in the program, as compared with controls, were more likely to be under adult supervision in the afternoon but did not improve their academic performance in school. Worse, students in the after-school group were more likely to get into trouble in school than control children, and teachers were less likely to report that program group students got along well with peers (Dynarski et al. 2002; James-Burdumy et al. 2005). Thus, the study found at least as many negative as positive impacts.

When the initial Mathematica study was released in 2003, the Bush administration's Office of Management and Budget (OMB) was preparing the President's budget for 2004. Under strong pressure to cut spending in order to reduce the projected 2004 deficit, OMB decided to use the study as justification to reduce spending on the 21st Century program by about 40 per cent, from \$1 billion to \$600 million. The President approved this proposal, and OMB put the cut in the President's budget. Here was a clear case in which a random-assignment evaluation found that a program produced poor outcomes and was therefore used as a justification to reduce funding for the program.

Under the budget process followed by the federal government, the President proposes a budget in February of each year, Congress passes its own budget resolution around April, and then committees write legislation to meet the numbers in the congressional budget resolution. If the President and Congress are from the same political party, which they were in this case, the President's budget and the

congressional budget are often close cousins. The 21st Century program, however, was popular among many members of Congress and enjoyed an effective group of child advocates arguing against the cut proposed by President Bush. In addition, as improbable as it might sound, the famous body builder and actor, Arnold Schwarzenegger, entered the action just as the congressional committees were making decisions about spending programs. Schwarzenegger was widely known for his advocacy of after-school programs and had led the fight on a state-wide referendum, enacted by California voters in 2002, which provided funding for a state after-school program (Rivera 2002, p. 6). Adding to the fascinating political situation, Schwarzenegger was in the final stages of deciding to run for governor of California as a Republican and was receiving encouragement from the Bush White House to run. Thus, the White House may have been somewhat conflicted about its recommendation to cut funding for the 21st Century program. In May 2003, the Senate Appropriations Committee held a hearing on its 2004 budget and, knowing full well what he would say, invited Schwarzenegger to testify on the administration's proposed cut in after-school programs. When an official representing the administration testified in favour of the cut by citing the findings from the Mathematica study, Schwarzenegger told the committee:

It would be a mistake, let me repeat, a big mistake, to use that study as justification to reduce current funding levels for after-school programs. Instead of cutting back the funding for after-school programs, we should begin to work together to focus on finding ways to improve them. (O'Keefe 2003)

Schwarzenegger did not feel it necessary to refute the study with any data or cogent arguments. Rather, he simply said it would be a mistake to use the study to cut the program. This is analysis by assertion, but it killed the Bush administration's proposal to cut the program. The Republican Congress ignored both the Mathematica study and the Bush administration and went ahead with full funding for the program. The lesson here, yet again, is that good evidence does not speak for itself in the policy process and is only one — sometimes a rather puny — element in a policy debate. Perhaps a few more studies and a continuing string of claims that after-school programs are not working would eventually cause Congress to take notice. But there was no chance in 2003 that a single study, even a high-quality study like Mathematica's, would overcome the political forces supporting after-school programs. In this case, star power and political exigency trumped evidence — and this particular data-based justification for a cut in the 21st Century program has not been heard since.

Home visiting programs and child abuse

The United States has many programs that aim to reduce child maltreatment and improve parenting by having trained staff visit pregnant mothers or mothers of newborns in their homes and offer them instruction about healthy living, financial management, child rearing and similar issues.⁸ There are several prominent home-visiting models, many with written curriculums, trained staff and elaborate financing arrangements. Individual programs vary with respect to children's and mothers' age, the risk status of families served, the range of services offered, and the intensity of the intervention as measured by the frequency and duration of the home visits. A recent review of these studies for the journal *The Future of Children* found that seven home-visiting programs have evidence from random-assignment studies showing that they produced at least one significant effect on parenting behaviour or child outcomes (Howard and Brooks-Gunn 2009).

Arguably, the most notable of these programs is the Nurse–Family Partnership program developed by David Olds of the University of Denver. First tested by a random-assignment design in rural New York on a sample of poor white teen mothers beginning in 1977, the program was later evaluated by random assignment in Memphis and Denver. In both replications, some characteristics of the original program, as well as the types of participating families, were varied. The original and both replications produced significant impacts on several maternal and child outcomes and have been reported in refereed journals. In 1996, Olds began expanding the program by working with state officials and others while trying to ensure fidelity to his program model. By 2008, Nurse–Family Partnership programs were being conducted in 25 states. Seldom has an intervention program been so carefully tested and expanded with such serious attention to getting new sites to maintain fidelity to the program model (Olds et al. 1998).

The success of the Olds program did not go unnoticed by senior officials in the Obama campaign and subsequently the Obama administration. President Obama's 2010 budget Blue Print, released in February 2009, included funds for 'Nurse Home Visitation' modelled on the Olds program (OMB 2009). With this proposal, the administration served notice that it intended to fund only programs based on the Olds model.

The president's intention to fund only the Olds program startled the worlds of early childhood education in general and home visiting in particular, because it meant that other nationally prominent programs, such as Parents as Teachers and Healthy Families America, would be left out. The concerns of these groups were not without

⁸ The summary used here follows Haskins et al. (2009, especially pp. 4–5).

merit. Some of them had, as we have seen, been subject to random-assignment evaluations. Furthermore, within the scholarly world, some believed that the Olds program required further evaluation: there were inconsistencies in the results from the three evaluations; the programs had not been subject to evaluation by researchers outside of Olds's team; and the program focused on a narrow group of mothers — notably, low-income first-time mothers who agreed, while pregnant, to participate in a two-year program.

With the emphasis on 'nurse home visiting' in Obama's budget Blue Print, the debate left the pristine confines of academic journals and conferences and leaped into the rough and tumble forum of federal policy making. In this venue, the home-visiting programs that felt slighted by the President's budget Blue Print initiated a lobbying campaign to broaden the President's proposal to include additional home-visiting programs. Many of the programs not singled out by the President were part of a long-established coalition of influential and effective Washington child advocacy groups that included the Center for Law and Social Policy, the Children's Defense Fund, the Child Welfare League of America and others. The general line taken by these programs and their advocates was that Obama's emphasis on home visiting was an important advance for children and families, but that his proposal to single out one program for support was ill-advised. All high-quality, evidence-based programs, they argued, should be eligible for funding. Not surprisingly, groups favouring the Olds program started lobbying, too. All this is standard fare for federal policy making.

Two entries in the debate are especially worthy of note. The Coalition for Evidence-Based Policy, an influential Washington lobby for high-quality program evaluation, declared its support for the President's decision to fund research-proven home-visitation programs such as the Nurse–Family Partnership. Run by Washington veteran Jon Baron, the coalition has assembled an advisory board that includes several noted scholars and others with an interest in applying high-quality evidence to policy choice, including a Nobel laureate. In April, the coalition issued a well-reasoned brief that emphasised its nonpartisan nature as an organisation focused on promoting the development of rigorous evidence. Indeed, Baron and his coalition have almost single-handedly succeeded in getting many pieces of federal legislation to designate funds for program evaluation. Citing an authoritative evidence review from *The Lancet*, a respected medical journal, that found the Olds program to have the 'best evidence for preventing child abuse and neglect', the Coalition for Evidence-Based Policy expressed unqualified support for funding of programs, such as the Nurse–Family Partnership, that meet the highest standards of evidence. A six-page attachment to the brief reviewed evidence from the three randomised controlled trials (RCTs) by which the Nurse–Family Partnership had shown its

strong impacts, while pointing to deficiencies in the RCTs by which five other home-visiting programs had been evaluated.

Perhaps spurred by the coalition's brickbat against the non-Olds programs, four highly respected scholars, including Deborah Daro of the University of Chicago, Kenneth Dodge of Duke, Heather Weiss of Harvard and Edward Zigler of Yale, sent a public letter to President Obama. Their soundly argued letter praised his proposal for investing in home-visitation programs, but criticised the focus on one program model. The impressive quartet argued that a single program model would leave out too many at-risk parents. They also cautioned against a sole reliance on evidence generated from RCTs, which do not provide guidance on how to scale up a model program to serve national needs. Finally, they expressed the view that although at-risk families merit the most intensive services, all families should have access to early child development programs. The world of social science, it appears, does not speak with one voice, and even the best evidence can lead to multiple—and sometimes directly opposing—conclusions.

By the time Congress approved its budget resolution in late April, the forces supporting the broader language appeared to be making headway, because the budget supported home-visiting programs that 'will produce sizable, sustained improvements in the health, well-being, or school readiness of children or their parents' and contained no mention of nurse visiting. Similarly, the Obama language on nurses was gone from the final administration budget released in early May.

The next and critical step was for congressional committees to begin writing the new program into law. The chairman of the Human Resources Subcommittee of the House Ways and Means Committee, Jim McDermott (a Democrat), was the first out of the box. He circulated draft legislation in early June 2009, and then held a hearing on his bill on 9 June. Like the budget resolution, the McDermott draft bill represents a compromise between the contending forces. Specifically, it would give priority funding to programs that 'adhere to clear evidence-based models of home visitation that have demonstrated significant positive effects on important program-determined child and parenting outcomes, such as reducing abuse and neglect and improving child health and development'. Preferred programs must also have 'well-trained and competent staff' and include training, technical assistance and evaluation.

Perhaps the most important sign of the central role being played by evidence in this debate is the 8 June blog posting by Peter Orszag, the director of the federal Office of Management and Budget and President Obama's closest adviser on budget policy. Orszag asserts that he and the President are placing evidence of program success from 'rigorous' evaluations (by which he appears to mean RCTs) at the

centre of decision making. He states emphatically that the Obama administration will evaluate as many programs as possible, cut off funding for those that are not working, and expand those that are. In the case of home-visiting programs, he endorses the two-tier approach of giving more money to the programs with the strongest evidence of success and some but less money to programs that have ‘some supportive evidence but not as much’. Orszag also cites several examples of how the Administration is expanding funds for conducting rigorous program evaluations and then using the evidence to make funding decisions.

As this episode unfolds, there is a lot to like for those who want to see quality evidence play a more prominent role in policy choice. It must be counted as a victory for the forces that favour evidence-based policy that the federal policy process on home visiting hinges importantly on evidence, a clear sign that both the Administration and Congress are giving a prominent place to high-quality evidence on successful programs. It also augurs well for evidence that the McDermott bill requires continuing evaluation of programs that receive the bill’s funding. Indeed, the bill sets aside \$10 million in guaranteed funding, mostly for program evaluation.

Regardless of the outcome, social scientists have taken an important step towards the goal of getting policy makers to consider high-quality evidence when making program funding decisions. That is a signal achievement for the research community — and, in the long run, for the improvement of public programs for children and families. As we have seen, the policy process does not often show any special deference to arguments based on evidence, but the US Congress and executive branch are showing new appreciation for how information can be used to fund the best programs that are likely to produce the best benefits for taxpayer investments.

Even so, it cannot be concluded that in recent years social science evidence has suddenly become a dominant force in legislative debates, but its status does seem to be improving. If the Obama administration actually delivers on the promise by the President and his budget director to fund programs that have strong evidence of success and to end programs that fail to produce impacts, the importance of evidence in political decisions in Washington will take a major leap forward. But before fans of evidence-based decisions get too excited, I suggest that they follow one round of the annual congressional appropriations process and see how many decisions are based on any appeal to evidence.

Applying evidence to program management

I now arrive at the less assertive part of my paper's split personality. I am by no means an expert on management issues. Rather, I have dipped into the management literature from time to time over the years, usually when I was curious about how data could inform program management. It is also impossible to be in Washington, DC, and not notice the many efforts to improve government efficiency, especially the Government Performance and Results Act passed in 1993 and the Bush Administration's Program Assessment Rating Tool implemented in 2002. As befits a man with only a modest amount to say, I have made this section much shorter than the first. My intent is to discuss five straightforward ideas for creating and applying program performance data to management decisions in order to build a system of continuous program monitoring and improvement.

Methods other than random assignment

In all three examples discussed above, RCT evaluations conducted under field conditions played a prominent role. There is impressive — albeit not unanimous — agreement in the social science world that RCTs are the most reliable method of program evaluation. The Coalition for Evidence-Based Policy, referred to above, was established specifically to promote the use of rigorous evidence in assessing program effectiveness, especially through the use of RCTs.⁹ Recently, the prestigious National Academies concluded that 'the highest level of confidence' in evidence of program effectiveness 'is provided by multiple, well-conducted randomized experimental trials'. Indeed, the National Academies went so far as to assert that when evidence from randomised trials is not available, 'evidence for efficacy or effectiveness cannot be considered definitive, even if based on the next strongest designs' (O'Connell et al. 2009, p. 371).

It is difficult not to agree that RCTs provide the strongest evidence of program effectiveness. However, it would be naive not to recognise that there are several problems with RCTs. The first is cost. It is not at all unusual for a multi-year, multi-site RCT evaluation on a single intervention to cost \$3 million or more.¹⁰ The US Government, sometimes in cooperation with the states, operates more than 1000 programs, including around 70 that are major social intervention programs (Brodsky

⁹ See the Coalition for Evidence-Based Policy's website (http://coalition4evidence.org/wordpress/?page_id=6/).

¹⁰ The cost of an RCT depends in large part on the type of data used to measure outcomes. If participants must be interviewed or tested, the costs increase sharply. If, on the other hand, administrative outcome data such as data from unemployment records, school records or government program records can be used, costs decline.

2008). A single evaluation of all 1000 programs could cost as much as \$3 billion. Even if someone were to present a strong argument that the \$3 billion would be well spent on evaluations, the current fiscal difficulties of the US federal government, which are sure to last for a decade or more, all but preclude finding \$3 billion to conduct RCT evaluations of all 1000 federal programs. Even if only the 70 major social programs identified by the Congressional Research Service were evaluated, the price tag could be \$210 million (Burke 2003). It follows that there should be some set of rules for picking the programs that promise to produce the highest return from RCT evaluations.

Another reason that an exclusive focus on RCTs would be unwise is that there are many other designs that have produced results that many social scientists think are worthy of attention (Besharov 2009a). Even if RCTs are the most reliable design, other designs and methods might be used more cheaply to identify promising initiatives that could then be subject to RCTs. Given the difficulty and expense of conducting RCTs, it might be argued that they should be saved for only the most promising intervention programs. How do we know the most promising programs? By using second-best designs and methods. At a minimum, these include regression discontinuity, propensity matching, difference-in-difference, fixed effects, instrumental variables and interrupted time series, all of which have been productively used in evaluating the effects of social programs in the United States and abroad (Besharov 2008; Smith 2009).

Although I remain a huge fan of RCTs, cost considerations plus the availability of worthy alternatives leads me to conclude that non-optimum methods will find a role in any broad system of evaluating large sets of social programs.

Test more than one program

Agencies can almost always identify several potential programs that could address a problem under their purview. Almost all preschool and K-12 programs, for example, have more than one curriculum designed to help children reach major goals such as proficient reading and math or cooperative social behaviour. Both RCTs and most of the other methods listed above can handle the simultaneous testing of more than one program. The great advantage of simultaneous testing is that more can be learned in a fixed period of time, and probably at a lower average price. Another advantage is that testing more than one intervention raises the odds of finding one that works. Neither politicians nor the public are happy to sit around waiting a decade or so while evaluators determine whether a particular program can produce reliable impacts. Concurrent testing of differing program models with the same goals should be pursued whenever possible.

Test current programs against alternative programs

One of the great problems with RCT designs is that they can leave program operators and managers without obvious next steps. In high-stakes evaluations, the most reliable information is simply whether the programs produce better results than are produced by no program. If there are no significant program-control differences, we are left knowing that the particular approach used by one intervention program does not work, but with little else. By contrast, if we compare a current program with several improved versions of the program or with an entirely new approach, the odds of learning something constructive are improved. In programs that are already operating at multiple sites, continuous monitoring of data on outcomes of interest will often reveal programs or program features seen to be producing superior results. Careful study of these outliers that produce impressive results can lead to the generation of hypotheses about effective program variations that can be tested against less successful programs. In the final stages of testing superior programs, after experience with second best but less expensive and time-consuming methods shows good outcomes, RCTs can be used to provide definitive evidence.

Know the program; study its implementation

Managers and their staff are often located in government buildings far away from the field in which their programs are implemented. In talking with bureaucrats about their programs, I often have the feeling that they know very little about the specific circumstances under which the programs are implemented, the people carrying out the implementation, or the children or adults the program is intended to help. If managers are to play a role in identifying ways to improve programs, they must do more than read evaluation reports. They must get out into the countryside and get up close and personal with their programs and the people implementing and participating in the programs. It is the combination of direct observation and evaluation reports that will help managers make good use of evaluation results and participate in searching for ways to alter the intervention and then test the new innovations to determine whether they produce better results.

Evaluation expert Richard Nathan has recently argued that program evaluations should devote more attention to program implementation and the institutions responsible for program implementation (Nathan 2009). In the case of national reforms, Nathan thinks researchers should visit and study multiple sites using whatever methods seem appropriate to find out exactly how field offices are implementing reforms. He cites as an example a remarkable study of welfare reform implementation following the 1996 legislation by his colleague Irene Lurie (2006).

Lurie and her team observed over 1000 interactions between caseworkers and clients in 12 local sites to determine how caseworkers conveyed the work requirement to families applying for welfare. Summarising a complex set of results, a central finding was that, unlike previous iterations of welfare reform that typically failed to have impacts on local welfare institutions, caseworkers observed by Lurie strongly enforced the work requirement of the 1996 law by telling applicants that they had to look for work as a condition of applying for welfare. Nathan's point is that trying to understand policy means trying to understand how the policy is actually implemented — if at all — at the local level. He argues further that knowledge provided by implementation studies of this type can be a useful complement to evaluation studies in allowing program managers to determine the effectiveness of programs under their purview and to continuously monitor and improve their programs.

3.2 Summary

The United States and other Western democracies are developing a worthy tradition of subjecting many social intervention programs to RCTs to determine their effectiveness. The prestigious National Academies in the United States has determined that only RCTs provide 'definitive' evidence of program effects. Not all researchers agree that RCTs are definitive, but there does appear to be all but universal agreement that social science has now developed effective methods of determining whether social intervention programs are having their intended impacts.

It is useful to think of using this evidence about program impacts in at least two ways. The first is to inform legislative decisions about whether to fund programs. This paper shows that there are interesting examples from the United States of the application of evidence from program evaluations to legislative decisions. All the examples show that many factors besides evidence inevitably play an important role in legislative debates, but in some (albeit not all) cases the role of evidence in the political debate can be of great importance in determining the outcome.

The second productive use of evaluation evidence is to provide information to program managers about whether their programs are being implemented effectively. In this case, since democratic governments tend to be organised by large administrative agencies with responsibility for many programs, the question arises of whether evaluation information can be built into a centralised system of continuous evaluation for the large number of social intervention programs under the control of specific agencies. While remaining somewhat sceptical of building large, bureaucratic mechanisms that encompass many programs, this paper offers

several generalisations about how managers can use evaluation information, both from RCTs and from other methods, to improve their programs.

Program evaluation, especially the RCT, is a reliable and valuable weapon in the ongoing effort by democratic governments to provide their citizens with effective social programs. There is no question that evaluations of single programs provide valid information about program impacts that can be effectively used to continuously improve programs. Whether evaluation information can be efficiently incorporated into broader, system-wide schemes for monitoring and improving many programs remains an open question.

References

- Ainslie, R. (ed.) 1984, *The Child and the Day Care Setting*, Prager, New York.
- Antos, J. et al. 2008, 'Taking back our fiscal future', *Brookings–Heritage Fiscal Seminar*, Washington, DC, April 2008, Brookings Institution and Heritage Foundation, Washington, DC.
- Auerbach, A.J. and Gale, W.G. 2009, *The Economic Crisis and the Fiscal Crisis: 2009 and Beyond, An Update*, Brookings, Washington, DC.
- Besharov, D.J. 2008, 'From the great society to continuous improvement government', *2009 APPAM Presidential Address*, Washington, DC, 7 November, College Park, Maryland, University of Maryland, Welfare Reform Academy.
- 2009a, 'From the great society to continuous improvement government: shifting from "Does it work?" to "What would make it work better?"', *Journal of Policy Analysis and Management*, vol. 28, no. 2, pp. 199–220.
- (ed.) 2009b, *Poverty and Welfare: Readings from Journal of Policy Analysis and Management*, Wiley, New York, in press.
- Blank, R.M. 1995, 'Outlook for the US labor market and prospects for low-wage entry', in Nightingale, D.S. and Haveman, R.H. (eds), *The Work Alternative: Welfare Reform and the Realities of the Job Market*, Urban Institute, Washington, DC, pp. 33–69.
- Bowlby, J. 1969, *Attachment and Loss*, vol. 1: Attachment, Basic Books, New York.
- Brodsky, R. 2008, 'Commanding performance', *National Journal*, 19 April, pp. 65-7.
- Burke, V. 2003, *Cash and Noncash Benefits for Persons with Limited Income: Eligibility Rules, Recipient and Expenditure Data, FY2000-FY2002*, CRS Report

-
- to Congress, RL32233, Congressional Research Service, 25 November, Washington, DC.
- Burtless, G. 1995, 'Employment prospects of welfare recipients', in Nightingale, D.S. and Haveman, R.H. (eds), *The Work Alternative: Welfare Reform and the Realities of the Job Market*, Urban Institute, Washington, DC, pp. 71–106.
- CBO (Congressional Budget Office) 2007, *Historical Effective Federal Tax Rates: 1979–2005*, <http://www.cbo.gov/ftpdocs/88xx/doc8885/12-11-HistoricalTaxRates.pdf>, (accessed 4 September 2009).
- CNN, 1996, 'The era of big government is over', http://www.cnn.com/US/9601/budget/01-27/clinton_radio/ (accessed 3 September 2009).
- Coe, N.B., Acs, G., Lerman, R.I. and Watson, K. 1998, 'Does work pay? A summary of the work incentives under TANF', Series A, No. A-28, Urban Institute, Washington, DC, <http://www.urban.org/uploadedPDF/anf28.pdf>, (accessed 2 March 2010).
- Crawford, C.C. 1994, *Multiple Employment Training Programs: Major Overhaul Is Needed*, T-HEHS-94-109, Government Accountability Office, Washington, DC.
- CWM (Committee on Ways and Means) 2005, *Compilation of the Social Security Laws Including the Social Security Act, as Amended, and Related Enactments through January 1, 2005*, vol. I, WMCP 109–2, US Government Printing Office.
- Dynarski, M. et al. 2003 *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program; First Year Findings*, PR02-82, Mathematica Policy Research, Princeton.
- Gruber, C.P., Clarke-Stewart, K.A. and Fitzgerald, L.M. 1994, *Children at Home and in Day Care*, Erlbaum, Hillsdale, NJ.
- Gueron, J.M. 2003, 'Presidential address: fostering research excellence and impacting policy and practice: the welfare reform story', *Journal of Policy Analysis and Management*, vol. 22, no. 2, pp. 163–74.
- and Pauly, E. 1991, *From Welfare to Work*, Russell Sage, New York.
- Haskins, R. 1991, 'Congress writes a law: research and welfare reform', *Journal of Policy Analysis and Management*, vol. 10, no. 4, pp. 616–32.
- 2006, *Work over Welfare: The Inside Story of the 1996 Welfare Reform Law*, Brookings, Washington, DC.
- and Sawhill, I. 2009, 'Supporting and encouraging work', *Creating an Opportunity Society*, Brookings, Washington, DC.

-
- Paxson, C. and Brooks-Gunn, J. 2009, ‘Social science rising: a tale of evidence shaping public policy’, *The Future of Children*, Policy Brief, Princeton University – Brookings Institution, Princeton, New Jersey.
- Haveman, R.H. 1995, ‘The Clinton alternative to “Welfare as we know it”: is it feasible?’, in Nightingale, D.S. and Haveman, R.H. (eds), *The Work Alternative: Welfare Reform and the Realities of the Job Market*, Urban Institute, Washington, DC.
- Howard, K.S. and Brooks-Gunn, J. 2009, ‘The role of home visiting programs in preventing child abuse and neglect’, *The Future of Children*, vol. 19, no. 2, Brookings – Princeton, Princeton, forthcoming.
- James-Burdumy, S. et al. 2005, *When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program*, Final Report, PR05-12, Mathematica Policy Research, Princeton, NJ.
- LaLonde, R.J. 1995, ‘The promise of public sector-sponsored training programs’, *Journal of Economic Perspectives*, vol. 9, no. 2, pp. 149–68.
- Long, D.A. 1988, ‘The budgetary implications of welfare reform: lessons from four state initiatives’, *Journal of Policy Analysis and Management*, vol. 7, no. 2, pp. 289–99.
- Lurie, I. 2006, *At the Front Lines of the Welfare System: A Perspective on the Declines in the Welfare Caseload*, Rockefeller Institute Press, Albany, New York.
- Montgomery, L. and Murray, S. 2009, ‘Lawmakers warned about health care costs’, *The Washington Post Online*, 17 July, <http://www.washingtonpost.com/wp-dyn/content/article/2009/07/16/AR2009071602242.html> (accessed 29 July 2009).
- Morris, P.A., Gennetian, L.A. and Duncan, G.J. 2005, ‘Effects of welfare and employment policies on young children: new findings on policy experiments conducted in the early 1990s’, *Social Policy Report*, vol. 21, no. 2, MDRC, New York.
- Munnell, A.H. 1986, *Lessons from the Income Maintenance Experiments: An Overview*, Conference Series No. 30, Brookings, Washington, DC.
- Nathan, R.P. 2009, ‘Social science methods and government effectiveness’, paper prepared for the National Association for Welfare Research and Statistics Annual Conference, Albany, New York, 12–15 July.
- O’Connell, M.E., Bost, T. and Warner, K.E. (eds) 2009, *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*, National Academies, Washington, DC.

-
- O’Keefe, E. 2003, ‘Terminator touts after-school programs, *ABC News Online*, 14 May, <http://abcnews.go.com/US/Story?id=90638&page=1>.
- Olds, D. et al. 1998, ‘Long-term effects of nurse home visitation on children’s criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial, *JAMA: The Journal of the American Medical Association*, vol. 280, no. 14, pp. 1238–44.
- OMB (Office of Management and Budget) 2009, *A New Era of Responsibility: Renewing America’s Promise*’.
- Pear, R. 2009, ‘Housing committee approves health care bill’, *The New York Times Online*, 16 July, <http://www.nytimes.com/2009/07/17/us/politics/17cbo.html> (accessed 29 July 2009).
- Radin, B.A. 2000, ‘The Government Performance and Results Act and the tradition of federal management reform: square pegs in round holes?’, *Journal of Public Administration Research and Theory*, vol. 10, no. 1, pp. 111–35.
- Rivera, C. 2002, ‘Schwarzenegger’s star power drives prop. 49’, *Los Angeles Times*, 22 October.
- Smith, J. 2009, Putting the evidence in evidence-based policy, paper presented at Productivity Commission Roundtable on Strengthening Evidence-based Policy in the Australian Federation, Canberra, 17–18 August, 2009, Productivity Commission, Melbourne.
- Tax Policy Center 2009, ‘“Making work pay” tax credit’, Brookings – Urban Tax Policy Center, Washington, DC, http://www.taxpolicycenter.org/taxtopics/conference_makingworkpay.cfm (accessed 4 September 2009).
- Weaver, R.K. 2000, *Ending Welfare as we Know it*, Brookings, Washington, DC.