

EVIDENCE-BASED POLICYMAKING REQUIRES A PORTFOLIO OF TOOLS

Testimony Submitted for the Record to the
Subcommittee on Human Resources
Committee on Ways and Means
United States House of Representatives
July 17, 2013

Margery Austin Turner
The Urban Institute

Margery Austin Turner is Vice President for Program Planning and Management at the Urban Institute. Opinions expressed are those of the author and do not necessarily reflect those of the Urban Institute, its trustees, or its sponsors.

Today more than ever, policymakers need evidence to help inform major decisions about program design, implementation, and funding. Whether assessing the likely effectiveness of a new initiative, comparing competing approaches to a given problem, figuring out where to cut, or refining a program's rules to make it more cost effective, decisions based on rigorous evidence make better use of scarce public dollars and improve outcomes for people.

A Case Study of Evidence-Based Policymaking. The evolution of federal low-income housing policy from its initial reliance on supply-side, project-based subsidies to include demand-side subsidies was significantly influenced by successive rounds of research and evidence. The origins of today's housing choice voucher program can be found in a 1969 President's Committee on Urban Housing report, *A Decent Home*, which recommended that HUD experiment with cash allowances made to families according to their need. Congress authorized just such an experiment in 1970. This study was a precursor to the influential Experimental Housing Allowance Program (EHAP), conducted in the 1970s and 1980s. In 1973, based on emerging evidence from these experiments, President Nixon asked Congress to reorient housing policy toward providing cash assistance instead of increasing production, based on preliminary experiments. This request led to the 1974 Housing Act that created the Section 8 voucher program.

By the early 1980s, findings from three housing allowance experiments (summarized in Struyk and Bendick 1981) showed that participants enjoyed improved quality of housing, that the program did not cause rent inflation, and that vouchers were actually cheaper than production programs. Based on this evidence, the Reagan administration implemented a dramatic shift from supply-side interventions focused on construction of new subsidized housing to demand-side certificate and voucher interventions. Although interest groups on the traditional left had long been the staunchest supporters of government-supported construction (and previous Democratic administrations had expanded construction), the Democratic Congress appropriated funds for voucher demonstrations from 1983 to 1987 and authorized the Reagan administration's proposed voucher program in 1988.

In the years that followed, evidence from random control trials, administrative data, and implementation research has been used to strengthen the housing voucher program's design and administration. More specifically, changes in how subsidies are calculated, rules imposed on participating landlords, allowable rent levels in different types of neighborhoods, and performance standards for the administering local agencies have all been informed by rigorous evidence conducted over many years (National Research Council 2008).

Roles for Evidence in the Policy Process

Often, the conversation about "evidence-based policy" focuses too narrowly on a single question and a single step in the policymaking process. In this narrow context, research determines whether a particular program "works" or not—whether it achieves the intended outcomes—and that evidence is

used as the basis for either funding the program or killing it. In other words, if an initiative or program has not been proven effective, it is not “evidence based” and it should not be implemented.

In reality, policy development occurs in multiple stages and extends over time. New policies emerge in response to problems and needs, possible approaches are advanced and debated, new policies are adopted and implemented, and established policies are critiqued and refined. As the figure to the right illustrates, evidence can add value at every stage, but the questions decisionmakers need to answer differ from one stage to the next. More specifically, policymakers need evidence to help them



- diagnose problems and underlying causes;
- design new policy options and assess the likely effects of alternatives;
- demonstrate and evaluate the impacts of new, model programs;
- monitor program implementation, measuring costs and performance and their sensitivity to different settings; and
- evaluate the long-term impacts and cost-effectiveness of existing programs.

No single research tool or methodology can deliver the evidence policymakers need to make informed decisions at all these stages. Instead, policymakers and practitioners need a portfolio of rigorous research tools to effectively advance evidence-based policy.

Random Control Trials

Random control trials—in which people are randomly assigned to participate in a program or serve as controls—are often referred to as the “gold standard” for evidence about whether a program is effective. And indeed, this approach is extremely powerful because it compares outcomes for a program’s participants to the outcomes comparable people achieve without the program.

To illustrate, the Urban Institute is currently evaluating a Youth Alliance initiative that places low-income youth in internships to teach them job skills and improve their résumés for college and careers. Often, programs of this kind appear effective on the surface, because many participants go on to college or get jobs. But what if the young people who were sufficiently motivated and goal-oriented to learn about and apply for the special services would have succeeded anyway? To find out whether these programs work, outcomes for participating youth must be compared to a control group of equally motivated

youth, the perfect recipe for a random control trial. In the case of the Youth Alliance initiative, only a limited number of internships were available, so the (much bigger) pool of eligible applicants was divided by lottery into a treatment group, which was offered internships, and a control group, which was not. These two groups of young people are similar in every other respect, including their drive and motivation to succeed. So, the evaluation, which is currently under way, will be able to measure the impact of the intervention relative to what participants would achieve without it.

While random control trials constitute the best tool in many circumstances, some programs cannot be effectively evaluated in this way. In particular, complex “place-based” interventions (like the new Promise Neighborhoods Initiative or the Choice Neighborhoods program) are not good candidates for random control trials. These interventions are designed to improve outcomes—educational success, health, employment and income—throughout an entire community, not just for individuals who participate in a defined program. Part of the approach is to saturate an entire area with new services, benefits, or incentives, so that even people who are not directly targeted or enrolled will experience spillover effects. These spillovers mean that the statistical framework underlying random assignment does not apply.

In theory, researchers could randomly assign whole communities to receive a place-based intervention or not. But place-based programs do not simply implement a simple prescription formulated the same way everywhere. Instead, these programs evolve organically in the communities where they are implemented and draw different elements from a broad menu of possible services. Each intervention is tailored to conditions on the ground. They are also continuously improved using data in an ongoing development effort that adapts to local circumstances, successes, and failures, with constant feedback from outcome data. The combination of spillover effects on people not receiving services, locally tailored designs, and continuous improvement makes a simple random assignment design the wrong choice for evaluating these types of interventions. But alternative methods can produce credible estimates of outcomes in the targeted communities compared to what would have happened without the intervention (Nichols 2013).

Other tools constitute the “gold standard” for delivering the evidence policymakers need to answer other questions.

Microsimulation Models

Often, policymakers want to anticipate the likely effects of policy changes, rather than waiting to measure their impacts after the fact. Microsimulation models can forecast outcomes under a wide range of “what if” scenarios. Although the development of credible models is complex and time-consuming, once a model is in place, it can quickly and efficiently analyze a wide range of alternative policies.

The Urban-Brookings Tax Policy Center has built a painstakingly detailed model that applies the provisions of the tax code to a 270,000-unit sample of “taxpayers” that statistically represents the US tax base.¹ The model reports the revenue and distributional effects of elements of and changes to the tax

¹ For a more complete description of the Urban-Brookings tax policy simulation model, see <http://taxpolicycenter.org/taxtopics/Brief-Description-of-the-Model-2013.cfm>.

code. For example, recent analyses applied the model to questions of equity and efficiency of the home mortgage interest deduction, demonstrating that its biggest beneficiaries are households making more than \$100,000. This analysis also predicted whose taxes would rise, whose would fall, and how total tax revenues would change under several alternatives to the mortgage interest deduction currently under discussion (Eng et al. 2013).

Administrative Data Analysis

Public agencies at every level of government collect a tremendous volume of information about benefit recipients, market transactions, and enforcement actions. These data can be systematically linked, monitored over time, and analyzed to produce reliable evidence for policymakers. In many cases, administrative data provide the building blocks for microsimulation modeling or the outcome measures in a random control trial. But rigorous analysis of administrative data can also provide answers to immediate questions about program design and implementation.

The Urban Institute is using administrative data to help design streamlined enrollment procedures for health insurance coverage under the Affordable Care Act. Publicly funded health insurance programs like Medicaid and CHIP require enrollees to complete vast amounts of paperwork to demonstrate their eligibility. This barrier is costly for both applicants and processors, and it prevents some eligible people from receiving coverage. With 21st century data and technology, it should be possible to automate processes to determine applicant eligibility, improving accuracy and saving public dollars. To test this proposition, Urban Institute researchers assembled administrative data from tax returns, state workforce eligibility records, and other sources to determine what existing records could verify current eligibility for Medicaid. This study will produce potential business rules for initial applications and renewal to the Medicaid program that would improve the system's efficiency and reach (Dorn et al. 2013).

Performance Measurement

Performance measurement is a familiar concept throughout the federal government and a key element of evidence-based policymaking. The Government Performance and Results Act of 1993 required federal agencies to report annually on performance indicators beginning in fiscal year 1999. Agencies set targets for each performance indicator and report progress against those targets. Several federal agencies have responded to the need for data-driven performance reviews by developing an approach that consists of regularly held, structured, data-driven performance review meetings. This strategic leadership approach, often referred to as PerformanceStat, was initially developed by the New York City Police Department in 1994, and has since been adapted for use in other local, state, and federal government agencies (Hatry and Davies 2011).

One of the key challenges of performance measurement is defining indicators that are both meaningful and measurable. If agencies focus exclusively on easy-to-collect process measures (like number of applications reviewed or service referrals provided), they fail to reflect outputs and outcomes of ultimate importance. But if they focus instead on long-term outcomes, they sacrifice near-term feedback on program performance. To help policymakers and practitioners develop effective performance measures, the Urban Institute, Child Trends, and Social Solutions are collaborating to

develop PerformWell, a web portal providing measurement tools and practical knowledge that human services professionals can use to manage their programs' day-to-day performance. Information in PerformWell leverages research-based findings that have been synthesized and simplified by experts in the field (Urban Institute 2011).

Qualitative Research

Sometimes, fully diagnosing a complex problem, designing an innovative solution, or understanding exactly how a program should be implemented requires in-depth, nuanced information that cannot be obtained from conventional surveys or administrative data sources. Qualitative research uses in-person observation, in-depth interviews, and focus groups to dig deep and explore the behavior of people and institutions. Qualitative methods can explain and enrich findings from statistical data, highlight key issues that might otherwise be missed, and reveal hypotheses for further testing.

The Urban Institute recently completed an evaluation of the federal New Markets Tax Credit program, which has allocated over \$3 billion to public-private investments in low-income community development. In conjunction with quantitative analyses of administrative and original survey data, researchers drew a small random sample of assisted businesses and analyzed each in depth, learning about the type of business and the loan it received, the nature of its neighborhood and market, the importance of the federal subsidy, and the subsequent economic performance of the business. This in-depth information played a critical part in a “triangulation” process that comprehensively addressed a complex market intervention (Abravanel et al. forthcoming).

Qualitative methods can also be invaluable in crafting the specific provisions of a new intervention before it is implemented or evaluated on a large scale (and at high cost). For example, behavioral economics has taught us that our financial decisionmaking is deeply influenced by subtle contextual factors that frame our choices. Programs designed to encourage savings among low-income earners are unlikely to succeed unless they take advantage of the “choice architecture” among their intended beneficiaries. The StabilityFirst pilot test, conducted in 2010 by Harvard’s “ideas42” center on applied behavioral economics, enrolled 20 students at Central New Mexico Community College in Albuquerque into a prepaid debit card program. The students were interviewed at length both before and after to gauge their reactions to the program. A range of issues surfaced, including difficulty resolving customer service matters. Participants were reluctant to make calls to the customer service line, not wanting to commit their scarce cellphone minutes to a possibly lengthy call with time spent being transferred or on hold (Mills 2011). Information like this enables program administrators to make adjustments that correct these “blocking factors,” making the subsequent randomized trial far more useful and the outcomes sought more likely to come about.

It almost goes without saying today that policymaking should be evidence based. Scarce public dollars should go to programs that target real problems, operate efficiently, and demonstrably achieve their intended goals. But policymaking is a messy, iterative process, and the opportunities for evidence to inform and strengthen decisions are numerous and varied. Instead of relying on a single tool,

policymakers and practitioners should draw from a “portfolio” of tools to effectively advance evidence-based policy. Using the wrong tool may produce misleading information or fail to answer the questions that are most relevant when a decision is being made. Applying the right tool to the policy question at hand can inform public debate, help decisionmakers allocate scarce resources more effectively, and improve outcomes for people and communities.

References

- Abravanel, Martin D., Nancy M. Pindus, Brett Theodos, Kassie Bertumen, Rachel Brash, Zach McDade. Forthcoming. “New Markets Tax Credit (NMTX) Program Evaluation Final Report.” Washington, DC: The Urban Institute.
- Dorn, Stan, Laura Wheaton, Paul Johnson, and Lisa Dubay. 2013. “Using SNAP Receipt to Establish, Verify, and Renew Medicaid Eligibility.” Washington, DC: The Urban Institute. <http://www.urban.org/publications/412808.html>.
- Eng, Amanda, Harvey Galper, Georgia Ivsin, and Eric Toder. 2013. “Options to Reform the Deduction for Home Mortgage Interest.” Washington, DC: Urban-Brookings Tax Policy Center. <http://www.urban.org/publications/412768.html>.
- Hatry, Harry, and Elizabeth Davies. 2011. *A Guide to Data-Driven Performance Reviews*. Washington, DC: IBM Center for the Business of Government. <http://www.businessofgovernment.org/report/guide-data-driven-performance-reviews>.
- Mills, Gregory. 2011. “Alpha Testing and Randomized Control Trials: Improving on the Gold Standard.” *MetroTrends* (blog), October 19. <http://blog.metrotrends.org/2011/10/alpha-testing-randomized-control-trials-improving-gold-standard/>.
- National Research Council. 2008. *Rebuilding the Research Capacity at HUD*. Washington, DC: The National Academies Press.
- Nichols, Austin. 2013. “Evaluation of Community-Wide Indicators.” Washington, DC: The Urban Institute. <http://www.urban.org/publications/412855.html>.
- President's Committee on Urban Housing (Edgar F. Kaiser, chairman). 1969. *A Decent Home: The Report of the President's Committee on Urban Housing*. Washington, DC: US Government Printing Office.
- Struyk, Raymond, and Marc Bendick Jr., eds. 1981. *Housing Vouchers for the Poor: Lessons from a National Experiment*. Washington, DC: Urban Institute Press.
- Urban Institute, The, and Venture Philanthropy Partners. 2011. “Tough Times, Creative Measures: What Will It Take to Help the Social Sector Embrace an Outcomes Culture?” Washington, DC: The Urban Institute and Venture Philanthropy Partners. <http://www.urban.org/publications/412467.html>.