

A Brief History of Evidence- Based Policy

By
JON BARON

This article provides a brief history of evidence-based policy, which it defines as encompassing (1) the application of rigorous research methods, particularly randomized controlled trials (RCTs), to build credible evidence about “what works” to improve the human condition; and (2) the use of such evidence to focus public and private resources on effective interventions. Evidence-based policy emerged first in medicine after World War II, and has made tremendous contributions to human health. In social policy, a few RCTs were conducted before 1980, but the number grew rapidly in U.S. welfare and employment programs during the 1980s and 1990s and had an important impact on government policy. Since 2000, evidence-based policy has seen a major expansion in other social policy areas, including education and international development assistance. A recent milestone is the U.S. enactment of “tiered evidence” social programs in which rigorous evidence is the defining principle in awarding government funding for interventions.

Keywords: evidence-based policy; evidence-based medicine; social policy; randomized controlled trial; welfare reform; tiered evidence

Evidence-based policy, as I define it in this article, encompasses two core elements: the application of rigorous research methods, particularly randomized controlled trials (RCTs), to build credible evidence about “what works” to improve the human condition; and the use of such evidence to focus public and private resources on programs, practices, and treatments (“interventions”) shown to be effective.

Jon Baron is vice president of evidence-based policy at the Laura and John Arnold Foundation, a nonprofit philanthropic foundation. He is also the founder and former president of the nonprofit Coalition for Evidence-Based Policy, which worked with federal officials from 2001 to 2015 to advance evidence-based reforms in government programs.

Correspondence: jbaron@arnoldfoundation.org

DOI: 10.1177/0002716218763128

Over most of modern world history, humanity did not need evidence-based policy to make important progress in improving health and life around the globe. For example, in the case of medical or public health interventions, such as penicillin for the treatment of bacterial infections, insulin for the treatment of diabetes, urban sanitation, and refrigeration of perishable foods, the treatment effects were so large that they could be detected using methods that were not rigorous, such as simply observing people's health before and after the intervention. The same is true of certain social and political interventions, such as a money economy as opposed to barter, rule of law as opposed to autocracy or anarchy, and schooling of children. Here, too, the intervention effects were so large that they could be directly observed without a rigorous evaluation.

But in the twenty-first century, many countries have already exploited the kinds of interventions that have blockbuster effects. To continue to make progress, evaluation methods need to detect effects that are more modest but still quite important—for example, a 10 percent increase in the survival rate (for a medical treatment), or a 20 percent reduction in the school dropout rate (for an education intervention). In these cases, simply observing people before and after cannot establish whether any change in their condition is due to the intervention or to other, confounding factors, such as the body's own immune defenses (in the case of a medical treatment) or demographic changes in the student population (in the case of an education intervention). To determine whether the intervention caused the outcomes we observe, we need an evaluation method that controls for confounding factors—such as an RCT that randomizes a sizable number of patients or students.

This article traces the history of evidence-based policy from its inception, after the end of World War II, through the present. My goal is to provide an overview, leaving it to other authors in this volume to fill in the specifics in particular time periods or fields of policy.

Evidence-Based Medicine in the Postwar Era, and Its Amazing Contribution to Human Health

Before World War II, medicine in the United States and other advanced countries was based mostly on anecdote and unscientific evidence. Sir Richard Doll, the British epidemiologist who became a leading figure in evidence-based medicine, recounts the state of medicine when he completed his training in the 1930s:

New treatments were almost always introduced on the grounds that in the hands of professor A or in the hands of a consultant at one of the leading teaching hospitals, the results in a small series of patients . . . had been superior to those recorded by professor B (or some other consultant) or by the same investigator previously. Under these conditions variability of outcome, chance, and the unconscious (leave alone the conscious) in the selection of patients brought about apparently important differences in the results obtained; consequently, there were many competing new treatments. . . . [W]hen I began to investigate peptic ulcers, I was soon able to prepare a list of treatments

beginning with each letter of the alphabet. Standard treatments, for their part, tended to be passed from one textbook to another without ever being adequately evaluated. (Doll 1998, 1217)

The 1940s and 1950s saw the first RCTs published in medicine, including the 1946 UK trial of streptomycin for treating pulmonary tuberculosis and the U.S. Salk polio vaccine field trials of 1954. These studies made an enormous contribution to public health. The Salk vaccine trials, for instance, demonstrated the effectiveness of a medical intervention that would, in the years that followed, eradicate paralytic polio—a devastating disease that, in the early 1950s, afflicted an average of sixteen thousand people in the United States each year, causing an average of nineteen hundred deaths (Centers for Disease Control and Prevention 1999).

Despite the success of the Salk vaccine and streptomycin studies, RCTs initially spread slowly and not without opposition until policy-makers embraced them as a requirement for licensing new drugs (Doll 1998, 1219). In 1962, Congress enacted legislation that, as implemented by the Food and Drug Administration (FDA), required that well-conducted RCTs demonstrate the effectiveness of any new pharmaceutical drug (and, as later amended, medical devices) before the FDA would approve it for marketing (FDA 2017). The FDA requirement directly embodied the second component of evidence-based policy described above—using rigorous evidence to focus public and private resources (in this case, an FDA license, and a potentially large and lucrative market) on effective interventions. The FDA policy also created a powerful new incentive for pharmaceutical companies to conduct RCTs aimed at building the number of proven-effective interventions—the first component of evidence-based policy.

The FDA policy, along with parallel support for clinical trials by the National Institutes of Health (NIH), transformed the RCT in medicine from a rare and controversial method that had first appeared in the medical literature only 15 years earlier (Medical Research Council 1948) into the widely used gold standard for assessing the effectiveness of all new drugs and medical devices. Between 1966 and 1995, the number of clinical research articles based on RCTs surged from about one hundred to ten thousand annually (Chassin 1998, 574).

Since the early 1960s, RCTs required by the FDA or funded by the NIH and other agencies have produced the conclusive evidence of effectiveness behind most major medical advances, including vaccines for measles, hepatitis B, and rubella; interventions for hypertension and high cholesterol, which helped to cut the incidence of coronary heart disease and stroke by more than 50 percent over the past half-century; and cancer treatments that have dramatically improved survival rates from leukemia, Hodgkin's disease, breast cancer, and many other cancers. Such advances have profoundly improved life and health in America (Gifford 1996).

Early RCTs in Social Policy, 1930s–1970s

The emergence of RCTs in social policy was far more gradual than it was in medicine. Several of the early social policy RCTs are discussed in more detail in other

articles; I summarize them briefly here to place them in historical context. Early RCTs included:

- the Cambridge-Somerville Youth Study, a relatively small RCT initiated in the 1930s to evaluate a program that provided counseling and group recreational activities, including summer camp, to low-income adolescent boys at risk for delinquency (Dishion, McCord, and Poulin 1999);
- the Manhattan Bail Bond project, an RCT initiated in 1961 to test the effects of releasing certain defendants without bail before trial (Botein 1965);
- the Perry Preschool study, a small RCT initiated in 1962 to evaluate a program that provided high-quality preschool education to three- and four-year-old children from low socioeconomic backgrounds (Schweinhart 2004);
- the Abecedarian study, a small RCT initiated in 1972 to evaluate a program that provided educational child care and high-quality preschool from birth to age 5 for children from disadvantaged backgrounds (Ramey, Sparling, and Ramey 2012);
- the income maintenance experiments initiated in the late 1960s and 1970s, which comprised four large-scale RCTs of “negative income tax” systems that provided a guaranteed payment to people with no earned income, gradually eliminated the payment as income rose, and converted into a graduated positive tax system as income rose further (Munnell 1986);
- the National Supported Work Demonstration, a large-scale RCT launched in 1974 to evaluate a program that offered subsidized jobs to hard-to-employ people, followed by assistance in finding an unsubsidized job (Manpower Demonstration Research Corporation 1980);
- three large-scale RCTs conducted in the 1970s to test the effects of providing housing vouchers to low-income households (Merrill and Joseph 1980); and
- the RAND Health Insurance Experiment, a large-scale RCT initiated in 1971 to test how different levels of cost-sharing (i.e., the portion of the bill the patient pays) affected health-care expenditures and patient health (Manning et al. 1987).

The policy impact of these early RCTs varied. At one end, the Cambridge-Somerville Youth Study generated significant interest among scholars because it found sizable adverse effects on participants’ life outcomes, but remained largely unknown in the policy community. The findings from the income maintenance experiments—modest adverse effects on participants’ employment and earnings—likely dampened the initial enthusiasm in the policy community for a negative income tax (a version of the idea had been proposed by the Nixon administration and passed by the House of Representatives in 1969). Similarly, the disappointing findings of the National Supported Work Demonstration on the employment and earnings of three of the four targeted populations—ex-offenders, ex-addicts, and youth (there were positive impacts for welfare mothers)—helped to diminish interest in supported work as a tool to address poverty. On the other hand, the

findings of the Perry Preschool and Abecedarian studies—large, long-term positive effects on outcomes such as educational attainment, employment, and criminal activity—have had an important influence on policy, inspiring many federal, state, and local initiatives to expand preschool, particularly for low-income children. That occurred despite the limitations of both studies, which were small demonstration projects with key departures from random assignment that somewhat reduce confidence in their findings.¹

The most important contribution that these early RCTs made to the history of evidence-based policy, however, may be unrelated to their findings. Judy Gueron and Howard Rolston (2013) suggest that it was to demonstrate that RCTs—including large, multisite trials—were indeed feasible in diverse areas of social policy. These early studies also built a community of social policy researchers at MDRC (which was created by the Ford Foundation to administer the National Supported Work Demonstration), Mathematica Policy Research, Abt Associates, and elsewhere with the expertise to carry out large-scale RCTs. And by 1980, a group of funders—most notably the Ford Foundation and U.S. Department of Labor (DOL), but also the U.S. Departments of Health and Human Services (HHS) and Housing and Urban Development (HUD)—had developed an interest and experience in sponsoring RCTs.

Welfare and Employment RCTs of the 1980s and 1990s, and Their Impact on Policy

A key development in the early 1980s was MDRC's launch of the Work/Welfare Demonstration, with core funding from the Ford Foundation. This initiative, comprising eight RCTs of different state-level welfare-to-work programs, showed for the first time that large-scale RCTs could be successfully integrated into normal state agency operations to evaluate programs developed and administered by those agencies. It was a novel approach at the time; the earlier large-scale RCTs described above had evaluated programs designed by researchers, foundations, and/or federal officials for purposes of the study (Gueron and Rolston 2013).

The Work/Welfare Demonstration's success encouraged HHS, in the mid-1980s, to begin funding large-scale RCTs of welfare-to-work and other employment, income supplementation, and related programs for the poor. These efforts had strong support from Congress, which, in the Family Support Act of 1988, required HHS to use RCTs to evaluate various welfare reforms. Support also came from the White House. During the George H. W. Bush and Bill Clinton administrations, HHS—with White House support—implemented a “demonstration waiver” policy, under which HHS allowed states to waive certain provisions of federal welfare law, using authority given to the HHS Secretary by section 1115 of the Social Security Act, so that they could test new welfare-to-work programs and other reforms. HHS approved the waivers only if states agreed to rigorously evaluate their reforms using RCTs wherever feasible (HHS 1994).

HHS thus directly funded or facilitated (e.g., through demonstration waivers) more than eighty-five RCTs in the 1980s and 1990s, building a sizable body of credible evidence with clear policy relevance. For example, these studies convincingly demonstrated that work-focused welfare reform programs that emphasized short-term job-search assistance and training, and encouraged participants to find work quickly, had larger effects on employment, earnings, and welfare dependence than reform programs emphasizing remedial education. The work-focused models were also much less costly to operate (Bloom, Hill, and Riccio 2003). Three major work-focused reform models (two in California and one in Oregon) were found to be especially effective. Each of the three increased participants' employment and earnings by 20 to 50 percent over a follow-up period of several years, and generated net government savings (e.g., from reduced welfare and food stamps payments) of \$2,500 to \$7,500 per person in 2017 dollars (Freedman et al. 1996, 2000; Hamilton et al. 2001).

The studies also found that when programs combined mandatory participation in employment-focused services (e.g., job search assistance or job training) with earnings supplements for participants who did find work, they could raise overall income and move many people out of poverty. For example, the Minnesota Family Investment Program, using such a strategy, not only produced sizable gains of 20 to 40 percent in employment and earnings for single-parent, long-term welfare recipients, but also reduced the proportion with overall income below the poverty line from 85 percent to 75 percent (Gennetian, Miller, and Smith 2005).

These findings had an important impact on policy and practice. According to federal officials and others involved in welfare reform, they helped to build political consensus for the strong work requirements in the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (i.e., federal welfare reform) and shape many of the work-first state-level reforms that followed. The findings' scientific rigor was critical to their policy impact (Haskins 2007). The sort of work-first and job club models that these studies found to be cost-effective are now the mainstay of welfare systems in the United States, Canada, the United Kingdom, Australia, and much of Europe.

Evidence-Based Policy Expands in Other Areas: 2000–Present

In early 2000s, three developments accelerated the expansion of evidence-based policy in areas of social policy beyond welfare and employment. Two of them primarily advanced the first component of evidence-based policy, that is, building credible evidence about what works through the use of rigorous research methods. The first was Congress's enactment of the Education Sciences Reform Act of 2002, which established the Institute of Education Sciences (IES) as a largely independent research organization in the U.S. Department of Education and

directed IES-commissioned evaluations of educational programs to use RCTs wherever feasible to measure program impact. As discussed in Whitehurst (this volume), over the following years and up to the present, IES has funded many large RCTs and other rigorous impact evaluations assessing the effectiveness of a wide range of education interventions—including classroom curricula, teacher professional development programs, school choice programs, educational software products, and many others.

The second development was the 2003 launch of the Jameel Poverty Action Lab (J-PAL) at the Massachusetts Institute of Technology, discussed in Chupein and Glennerster (this volume), which began conducting large-scale RCTs in developing countries to determine what works to reduce poverty and improve education, health, and other outcomes. Such studies were a major departure from prior, less rigorous research methods in the field of international development assistance, and they fundamentally transformed research in that field over the following decade. Between 2003 and as of this writing, J-PAL RCTs have evaluated a vast array of interventions in developing countries, including educational tutoring in primary schools, child immunization campaigns, financial incentives to reduce child marriage and teen parenting, and the promotion of self-employment among the very poor, to name just a few.

The third development was the advent of the Coalition for Evidence-Based Policy, a nonprofit, nonpartisan organization that I founded in 2001, which became an effective partner with Congress and the Executive Branch in advancing both components of evidence-based policy—the development and the use of rigorous evidence (Wallace 2011). According to an independent assessment in 2009, the coalition “successfully influenced legislative language, increased funding for evidence-based evaluations and programs ... and raised the level of debate in the policy process regarding standards of evidence” (Herk 2009, 1). Perhaps its most important contribution was to enlist and support the Office of Management and Budget (OMB) as a key proponent of evidence-based policy across the federal government, through a coalition-OMB partnership that spanned the administrations of George W. Bush and Barack Obama. For example, in early 2004, the coalition worked closely with OMB to develop guidance on evidence-based approaches for federal agencies, in which OMB for the first time identified well-conducted RCTs as the strongest method for evaluating program effectiveness (OMB 2004). In 2007, the coalition worked closely with OMB and the House and Senate appropriations committees to gain congressional passage of a \$10 million pilot program to fund evidence-based home visitation interventions that “have been shown, in well-designed randomized controlled trials, to produce sizable, sustained effects on important child outcomes such as abuse and neglect” (Haskins and Margolis 2015, 32–34). In January 2009, the coalition proposed a three-tiered approach to evidence-based social investment that was a basis for the Obama administration’s tiered-evidence grant programs, discussed below, and the coalition worked closely with OMB to design some of the programs (Coalition for Evidence-Based Policy 2009).

Between 2009 and 2011, Congress enacted six tiered-evidence social programs in areas such as K–12 education, early childhood home visiting, and teen

pregnancy prevention. This marked a major milestone in the history of evidence-based policy, in that rigorous evidence was the defining principle for awarding funding for interventions. The programs, which are discussed in detail in Haskins and Margolis (2015), award their largest grants—those in the top tier—to interventions that have strong or highly promising evidence of effectiveness. These grants pay for expanding the intervention and testing (through a required replication RCT) whether the effects found in prior studies can be reproduced on a larger scale. The tiered-evidence programs also award smaller grants (in a second and, in some programs, a third tier) to implement interventions with moderate or preliminary evidence of effectiveness, and to rigorously test, usually through an RCT, whether they produce the hoped-for effects. If they do, they can move into the top tier and qualify for more funding; if not, the funds are directed to other, more promising efforts.

Some of the tiered-evidence programs fell victim to congressional budget cuts starting in 2013, but three of them attracted bipartisan support and continue to this day: the Department of Education's Education Innovation and Research (EIR) program (funded at \$100 million in Fiscal Year [FY] 2017); HHS's Maternal, Infant, and Early Childhood Home Visiting Program (funded at \$400 million in FY2017); and HHS's Teen Pregnancy Prevention (TPP) Program (funded at \$101 million in FY2017).

Findings from the initial set of tiered-evidence grants (awarded in 2010–2013) are now being reported. As expected, a number of the funded interventions were found to produce small or no positive effects, as is true whenever rigorous evaluations are carried out. But some were major successes; that is, they were found in the required RCT evaluations to produce sizable, sustained effects on important life outcomes. These include the following:

- Per Scholas Job Training, a program that provides training and employment services in the information technology sector to low-income workers (in the third year after random assignment, workers in the Per Scholas group earned an average of \$4,800, or 27 percent, more than workers in the control group) (Schaberg 2017; Hendra et al. 2016);
- Knowledge Is Power Program (KIPP) elementary and middle schools, a national network of public charter schools whose mission is to help underserved students enroll in and graduate from college (two to three years after random assignment, students in KIPP schools scored 5 to 10 percentile points higher in reading and math than students in the control group) (Tuttle et al. 2015); and
- Teen Options to Prevent Pregnancy (TOPP), an intervention for low-income adolescent mothers that aims to reduce rapid repeat pregnancy and promote healthy birth spacing (20 months after random assignment, 21 percent of the TOPP group had experienced a repeat pregnancy, compared to 38 percent of the control group) (Stevens et al. 2017).

Concluding Observations

The first and most clearly successful example of evidence-based policy in action occurred in the field of medicine—particularly pharmaceutical medicine, which uses rigorous evidence of effectiveness to determine which drugs can be marketed, and which has thereby generated tremendous advances in human health over the past half century. Evidence-based policy has also scored some notable successes in social policy. Space constraints prevent me from discussing the challenges along the way in both evidence-based medicine and social policy, but as illustrative examples these have included opposition from the American Medical Association and pharmaceutical industry to the 1962 legislation requiring rigorous evidence of drug effectiveness prior to marketing (the legislation was nevertheless enacted because Congress felt urgency to respond to the thalidomide tragedy);² opposition from certain policy officials and academics to the primacy of RCTs for establishing effectiveness in medicine (e.g., Frieden 2017) and in social programs (e.g., Schorr 2012; Deaton 2010); and the dilution of the standards for rigorous evidence of effectiveness in some of the tiered-evidence social programs (e.g., Coalition for Evidence-Based Policy 2010; Baron 2014).

Despite important recent advances, tiered-evidence programs, rigorous evaluation requirements, and other evidence-based approaches have so far gained only a foothold in social policy; the majority of social spending is still allocated with little regard to rigorous evidence about what works. But the narrative of this article suggests that history is on the side of evidence. As Robert Slavin (2017) has observed, evidence-based social policy may now be in a position similar to the automobile industry in 1909: there were few cars and lots of engineering problems, but anyone with eyes to see knew that the automobile was the future. So may it be with evidence.

Notes

1. See <http://evidencebasedprograms.org/1366-2/65-2>.

2. Thalidomide is a pharmaceutical drug that, in the late 1950s and early 1960s, was marketed worldwide to alleviate sleeplessness and, for pregnant women, morning sickness. Described by its manufacturer as completely safe, it caused severe birth defects that affected thousands of children.

References

- Baron, Jon. 2 April 2014. In Support of the Reauthorization of HHS's Maternal, Infant, and Early Childhood Home Visiting Program, with Recommendations for Improvement. Testimony before the House Ways and Means Subcommittee on Human Resources. Available from <http://coalition4evidence.org/wp-content/uploads/2014/08/Coalition-testimony-4.2.2014.pdf>.
- Bloom, Howard S., Carolyn J. Hill, and James A. Riccio. 2003. Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments. *Journal of Policy Analysis and Management* 22 (4): 551–75.

- Botein, B. 1965. The Manhattan Bail Project: Its impact in criminology and the criminal law process. *Texas Law Review* 43:319–31.
- Centers for Disease Control and Prevention. 2 April 1999. Achievements in public health, 1900–1999: Impact of vaccines universally recommended for children—United States, 1990–1998. *Morbidity and Mortality Weekly Report* 48 (12): 243–48.
- Chassin, Mark R. 1998. Is health care ready for six sigma quality? *Milbank Quarterly* 76 (4): 565–91.
- Chupein, Thomas, and Rachel Glennerster. 2018. Evidence-informed policy from an international perspective. *The ANNALS of the American Academy of Political and Social Science* (this volume).
- Coalition for Evidence-Based Policy. 2009. *Suggestions for the administration's new social entrepreneurship initiative: Focus on building a body of research-proven programs, shown to produce major gains in education, poverty reduction, crime prevention, and other areas*. Washington, DC: Coalition for Evidence-Based Policy. Available from <http://coalition4evidence.org/wp-content/uploads/2009/06/ideas-for-social-entrepreneurship-initiative-12309.pdf>.
- Coalition for Evidence-Based Policy. 2010. *HHS's Evidence-Based Teen Pregnancy Prevention Program: Excellent first step, but only 2 of 28 approved models have strong evidence of effectiveness*. Washington, DC: Coalition for Evidence-Based Policy. Available from <http://coalition4evidence.org/wp-content/uploads/2010/05/Coalition-comments-HHS-Teen-Pregnancy-Prevention-May-2010.pdf>.
- Deaton, Angus. 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* 48:424–55.
- Dishion, Thomas J., Joan McCord, and Francois Poulin. 1999. When interventions harm: Peer groups and problem behavior. *American Psychologist* 54 (9): 755–64.
- Doll, Richard. 1998. Controlled trials: The 1948 watershed. *British Medical Journal* 317:1217–20.
- Freedman, Stephen, Daniel Friedlander, Winston Lin, and Amanda Schwede. 1996. The GAIN evaluation: Five-year impacts on employment, earnings, and AFDC receipt. Working Paper 96.1, MDRC, Washington, DC.
- Freedman, Stephen, Jean Tansey Knab, Lisa A. Gennetian, and David Navarro. 2000. *The Los Angeles jobs-first GAIN evaluation: Final report on a work first program in a major urban center*. Washington, DC: MDRC.
- Frieden, Thomas R. 2017. Evidence for health decision making—Beyond randomized, controlled trials. *New England Journal of Medicine* 377:465–75.
- Gennetian, Lisa A., Cynthia Miller, and Jared Smith. 2005. *Turning welfare into a work support: Six-year impacts on parents and children from the Minnesota Family Investment Program*. Washington, DC: MDRC.
- Gifford, Ray W. 1996. FDR and hypertension: If we'd only known then what we know now. *Geriatrics* 51 (1): 29–32.
- Gueron, Judith M., and Howard Rolston. 2013. *Fighting for reliable evidence*. New York, NY: Russell Sage Foundation.
- Hamilton, Gayle, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines, Sharon McGroder, Martha Zaslow, Jennifer Brooks, et al. 2001. *National evaluation of welfare-to-work strategies: How effective are different welfare-to-work approaches? Five-year adult and child impacts for eleven programs*. Washington, DC: MDRC and Child Trends.
- Haskins, Ron. 2007. *Work over welfare: The inside story of the 1996 welfare reform law*. Washington, DC: Brookings Institution Press.
- Haskins, Ron, and Greg Margolis. 2015. *Show me the evidence: Obama's fight for rigor and results in social policy*. Washington, DC: Brookings Institution Press.
- Hendra, Richard, David H. Greenberg, Gayle Hamilton, Ari Oppenheim, Alexandra Pennington, Kelsey Schaberg, and Betsy L. Tessler. 2016. *Encouraging evidence on a sector-focused advancement strategy: Two-year impacts from the WorkAdvance demonstration*. Washington, DC: MDRC.
- Herk, Monica. 2009. *The Coalition for Evidence-Based Policy: Its role in advancing evidence-based reform, 2004–2009*. Available from <http://coalition4evidence.org/wp-content/uploads/Indep-assessment-of-Coalitions-work-2009.pdf>.
- Manning, Willard G., Joseph P. Newhouse, Nailhua Duan, Emmett B. Keeler, Arleen Leibowitz, and M. Susan Marquis. 1987. Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review* 77 (3): 251–77.

- Manpower Demonstration Research Corporation. 1980. *Summary and findings of the National Supported Work Demonstration*. Pensacola, FL: Ballinger Publishing Company.
- Medical Research Council Streptomycin in Tuberculosis Trials Committee. 1948. Streptomycin treatment for pulmonary tuberculosis. *British Medical Journal* 2 (4582): 769–82.
- Merrill, Sally R., and Catherine A. Joseph. 1980. *Housing improvement and upgrading in the housing allowance demand experiment*. Cambridge, MA: Abt Associates Inc.
- Munnell, Alicia, ed. 1986. *Lessons from the Income Maintenance Experiments: Proceedings of a conference*. Boston, MA: Federal Reserve Bank of Boston, Conference Series 30.
- Office of Management and Budget. 2004. *What constitutes strong evidence of a program's effectiveness*. Washington, DC: Office of Management and Budget.
- Ramey, Craig T., Joseph J. Sparling, and Sharon Landesman Ramey. 2012. *Abecedarian: The ideas, the approach, and the findings*. Los Altos, CA: Sociometrics Corporation.
- Schaberg, Kelsey. 2017. *Can sector strategies promote longer-term effects? Three-year impacts from the WorkAdvance demonstration*. Washington, DC: MDRC.
- Schorr, Lisbeth B. 2012. Broader evidence for bigger impact. *Stanford Social Innovation Review* 10 (4): 50–55.
- Schweinhart, Lawrence J. 2004. *The High/Scope Perry Preschool Study through age 40: Summary, conclusions, and frequently asked questions*. Ypsilanti, MI: High/Scope Press.
- Slavin, Robert. 15 June 2017. The age of evidence. *Huffington Post*. Available from <http://www.huffingtonpost.com>.
- Stevens, Jack, Robyn Lutz, Ngozi Osuagwu, Dana Rotz, and Brian Goesling. 2017. A randomized trial of motivational interviewing and facilitated contraceptive access to prevent rapid repeat pregnancy among adolescent mothers. *American Journal of Obstetrics and Gynecology* 217 (4): 423.e1–423.e9.
- Tuttle, Christina Clark, Philip Gleason, Virginia Knechtel, Ira Nichols-Barrer, Kevin Booker, Gregory Chojnacki, Thomas Coen, and Lisbeth Goble Tuttle. 2015. *Understanding the effect of KIPP as it scales, vol. I, Impacts on achievement and other outcomes final report of KIPP's Investing in Innovation Grant evaluation*. Cambridge, MA: Mathematica Policy Research.
- U.S. Department of Health and Human Services. 27 September 1994. Demonstration proposals pursuant to section 1115(a) of the Social Security Act; Policies and Procedures. *Federal Register* 59 (186).
- U.S. Food and Drug Administration. 1 April 2017. Adequate and well-controlled studies. 21 Code of Federal Regulations §314.126.
- Wallace, John W. 2011. *Review of the Coalition for Evidence-Based Policy*. Washington, DC: Coalition for Evidence-Based Policy. Available from <http://coalition4evidence.org/wp-content/uploads/Report-on-the-Coalition-for-Evidence-Based-Policy-March-2011.pdf>.
- Whitehurst, Grover J. (Russ). 2018. The Institute of Education Sciences: A model for federal research offices. *The ANNALS of the American Academy of Political and Social Science* (this volume).