



## Demonstrating How Low-Cost Randomized Controlled Trials Can Drive Effective Social Spending

The Coalition for Evidence-Based Policy, in cooperation with the White House and the Office of Management and Budget (OMB),<sup>1</sup> is developing a high-profile competition for low-cost randomized controlled trials (RCTs) that are designed to generate valid, actionable evidence about “what works” in U.S. social spending. The first three sections of this paper summarize the concept of low-cost RCTs, provide concrete examples of such studies, and discuss why they are important. The fourth section describes the competition we are launching to select and fund at least six such studies over the next three years (two to three per year).

### I. **The Breakthrough: Low-cost RCTs are a recent innovation in social policy research that can rapidly build the body of evidence about “what works” to address major social problems.**

A. **Background: Well-conducted RCTs are regarded as the strongest method of evaluating the effectiveness of social programs/strategies (“interventions”),** per evidence standards articulated by the National Academy of Sciences,<sup>2</sup> Congressional Budget Office,<sup>3</sup> Institute of Education Sciences,<sup>4</sup> U.S. Preventive Services Task Force,<sup>5</sup> Food and Drug Administration,<sup>6</sup> and other respected scientific bodies.

**Other, more common methods of evaluation, while providing valuable information, cannot by themselves produce definitive evidence.** Quasi-experiments, outcome-based performance metrics, and RCTs with small samples or other limitations are valuable for identifying *potentially* effective interventions that merit more rigorous testing, but alone cannot produce strong evidence. Studies show that, too often, their findings are not confirmed in sizable, high-quality RCTs, because they are unable to reliably control for all factors other than program participation.<sup>7</sup> Thus, a recent National Academy of Sciences report recommends that evidence of program effectiveness generally “cannot be considered definitive” without ultimate confirmation in well-conducted RCTs, “even if based on the next strongest designs.”<sup>1</sup>

### B. **Recently, researchers have shown it is often possible to conduct RCTs at low cost and burden, addressing what has been a major obstacle to their widespread use. This is done by –**

1. **Embedding random assignment in initiatives that are being implemented anyway as part of usual program operations.** Federal social programs fund a vast array of strategies and approaches at the state and local level and, over time, new initiatives and reforms are often launched. Credible evaluations can be embedded in many of these efforts – for example, by (i) using a lottery process – i.e., random assignment – to determine who will be offered program services (since programs often do not have sufficient funds to serve everyone who is eligible); or (ii) randomly assigning some individuals to the program’s usual approach (e.g., transitional jobs for ex-offenders) versus a revised model that is being piloted (e.g., transitional jobs plus drug treatment), to see if the new model produces better outcomes.

- and -

2. **Using administrative data that are collected already for other purposes to measure the key outcomes,** rather than engaging in original – and often costly – data collection (e.g., researcher-administered interviews, observations, or tests). In many jurisdictions, administrative data of reasonable quality are available to measure outcomes such as child maltreatment rates, employment and earnings, student test scores, criminal arrests, receipt of government assistance, and health care expenditures.

**Such piggybacking on ongoing efforts/resources enables many more RCTs to go forward, by reducing their cost as much as tenfold.** Specifically, this approach reduces or eliminates what are typically the most costly and complex components of an RCT: collecting original outcome data from each sample member; delivering the intervention that is to be evaluated; and recruiting a sample of individuals or other units (such as schools) to participate in the study.

**II. Illustrative Examples: The following are five well-conducted RCTs, in diverse program areas, that cost between \$50,000 and \$300,000 – a fraction of the usual multimillion-dollar cost of such studies.** Through the rigorous low-cost approaches described above, these studies all produced valid evidence of practical importance for policy decisions and, in some cases, identified program strategies that generate budget savings. (A detailed summary of each study, along with the study citations, is [posted here](#).)

**A. Child Welfare Example: Recovery Coaches for Substance-Abusing Parents**

- **Overview of the study:** This Illinois program provided case management services to substance-abusing parents who had temporarily lost custody of their children to the state, aimed at engaging them in treatment. The program was evaluated in a well-conducted RCT with a sample of 60 child welfare agencies, working with 2,763 parents. The study found that, over a five-year period, the program produced a 14% increase in family reunification, a 15% increase in foster care cases being closed, and net savings to the state of \$2,400 per parent.
- **Cost of measuring program impact: About \$100,000.** The low cost was achieved by measuring study outcomes using state administrative data (e.g., data on foster care case closures).

**B. Early Childhood Example: The Triple P (Positive Parenting Program) System**

- **Overview of the study:** This program is a system of parenting interventions for families with children ages 0-8, which seeks to strengthen parenting skills and prevent child maltreatment. A well-conducted RCT evaluated the program as implemented county-wide in a sample of 18 rural and semi-urban South Carolina counties. The study found that the program reduced county-wide rates of child maltreatment, hospital visits for maltreatment injuries, and foster-care placements by 25-35%, two years after random assignment.
- **Cost of measuring program impact: \$225,000-\$300,000.** The low cost was achieved by measuring study outcomes using state administrative data (e.g., child maltreatment records).

**C. K-12 Education Example: New York City Teacher Incentive Program**

- **Overview of the study:** This program provided low-performing schools that increased student achievement and other key outcomes with an annual bonus, to be distributed to teachers. It was evaluated in a well-conducted RCT with a sample of 396 of the city's lowest-performing schools, conducted over 2008-2010. The study found that, during the three years after random assignment, the program produced no effect on student achievement, attendance, graduation rates, behavior, or GPA. Based in part on these results, the city ended the program, freeing up resources for other efforts to improve student outcomes.
- **Cost of measuring program impact: About \$50,000.** The low cost was achieved by measuring study outcomes using school district administrative data (e.g., state test scores).

**D. Criminal Justice Example: Hawaii's Opportunity Probation with Enforcement (HOPE)**

- **Overview of the study:** HOPE is a supervision program for drug-involved probationers that provides swift and certain sanctions for a probation violation. It was evaluated in a well-conducted

RCT with a sample of 493 probationers, with follow-up one year after random assignment. The study found that the program reduced probationers' likelihood of re-arrest by 55%, and the number of days incarcerated by 48%, during the year after random assignment.

- **Cost of measuring program impact: About \$150,000.** The low cost was achieved by measuring study outcomes using state administrative data (e.g., arrest and incarceration records).

#### **E. Criminal Justice Example: Philadelphia Low-Intensity Community Supervision Experiment**

- **Overview of the study:** This was a program of Low-Intensity Community Supervision for probationers or parolees at low risk of committing a serious crime (compared to the usual, more intensive/costly supervision). The program's purpose was to reduce the cost of supervision to Philadelphia County without compromising public safety. The program was evaluated in a well-conducted RCT with a sample of 1,559 offenders, with follow-up one year after random assignment. The study found that the program caused no increase in crime compared to the usual, more-intensive supervision of such offenders, indicating that program is a viable way to reduce costs in the criminal justice system. Based on the findings, the county adopted this approach for all low-risk offenders.
- **Cost of measuring program impact: Less than \$100,000.** The low cost was achieved by measuring study outcomes using county administrative data (e.g., arrest records).

### **III. Why It Matters:**

**A. Progress in social policy, as in other fields, requires strategic trial and error – i.e., rigorously testing many promising interventions to identify the few that are effective.** Well-conducted RCTs, by measuring interventions' true effect on objectively important outcomes such as college attendance, workforce earnings, teen pregnancy, and crime, are able to distinguish those that produce sizable effects from those that do not. Such studies have identified a few interventions that are truly effective (e.g., see [topierevidence.org](http://topierevidence.org), [colorado.edu/cspv/blueprints/](http://colorado.edu/cspv/blueprints/)), but these are exceptions that have emerged from testing a much larger pool. Most, including those thought promising based on initial studies, are found to produce few or no effects – underscoring the need to test many. For example:

- **Education:** Of the 90 interventions evaluated in RCTs commissioned by the Institute of Education Sciences (IES) since 2002, approximately 90% were found to have weak or no positive effects.<sup>8</sup>
- **Employment/training:** Of the 13 interventions evaluated in Department of Labor RCTs that have reported results since 1992, about 75% were found to have found weak or no positive effects.<sup>9</sup>
- **Medicine:** Reviews have found that 50-80% of positive results in initial (“phase II”) clinical studies are overturned in subsequent, more definitive RCTs (“phase III”).<sup>10</sup>
- **Business:** Of 13,000 RCTs of new products/strategies conducted by Google and Microsoft, 80-90% have reportedly found no significant effects.<sup>11</sup>

**B. The current pace of RCT testing is far too slow to build a meaningful number of proven interventions to address our major social problems.** Of the vast diversity of ongoing and newly-initiated program activities in federal, state, and local social spending, only a small fraction are ever evaluated in a credible way to see if they work. The federal government, for example, evaluates only 1-2 dozen such efforts each year in RCTs that are usually specially-crafted projects, with research or evaluation funds often paying for delivery of the intervention, recruitment of a sample population, site visits, implementation research, and data collection through researcher-administered interviews, observations, or tests. The cost of such studies is typically several million dollars.

These studies produce valuable and comprehensive information, but – because of the cost and organizational effort – are far too few to build a sizable body of proven-effective interventions, especially since most find weak or no effects for the interventions being studied. As discussed below, we believe such studies are most valuable when focused on interventions whose effectiveness has been initially established through lower-cost methods.

- C. Embedding low-cost RCTs in the myriad of ongoing social spending activities can dramatically accelerate the process, enabling hundreds of interventions to be tested each year, rather than a few.** Often the key ingredient is creative thinking – i.e., figuring out how to embed a lottery or other randomization process into a particular activity, and measure key outcomes with an existing data source.

An important collateral benefit of such large-scale, low-cost RCT testing is that it can identify interventions that are excellent candidates for further study in traditional, more comprehensive RCTs – because effectiveness has been initially established. Traditional RCTs are especially useful when focused on such interventions, as they can supply evidence about whether and how to scale up the intervention that is of direct value for decision-making – such as: (i) corroboration of the initial finding of effectiveness in a second study; (ii) the intervention’s effect on outcomes other than those measurable with administrative data; (iii) the subgroups and conditions in which the intervention is most effective; (iv) the nature of services received by intervention participants, and how they differ from any services received by the control group; (v) possible reasons why the intervention produced its effect; and (vi) how the intervention’s benefits compare to its costs.

- D. The end goal – a sizable body of proven-effective social interventions – is of critical importance. In recent decades, the U.S. has failed to make significant progress in key areas such as –**

- **Poverty:** The U.S. poverty rate – now at 15% – reached its low in 1973. It has shown little change (whether by official or alternative National Academy measures) since the 1970s.<sup>12</sup>
- **K-12 education:** Reading and math achievement of 17-year-olds – the end product of our K-12 education system – is virtually unchanged over the past 40 years, according to official measures,<sup>13</sup> despite a 90% increase in public spending per student (adjusted for inflation).<sup>14</sup>
- **Well-being of low to moderate income Americans:** The average yearly income of the bottom 40% of U.S. households, now at \$20,600, has changed little since the early 1970s.<sup>15</sup>

**IV. Next Step: The Coalition, in cooperation with the White House and OMB, is developing a competition to demonstrate the value of low-cost RCTs.**

**A. Specific purpose:**

- **Low-cost RCTs are a recent – and not widely-known – phenomenon, carried out thus far by a small group of entrepreneurial university researchers partnered with public officials.**
- **By demonstrating the value of such studies to a wider audience of researchers and officials, this effort aims to engage many others in advancing/institutionalizing the concept.** Toward that end, we are working in consultation with the White House, OMB, and others to identify federal agency officials who could help move the concept forward; and to engage such officials in advising the initiative, exploring possible use of low-cost RCT approaches in their agencies, and advancing such approaches where appropriate.
- **This demonstration effort will complement – and provide external reinforcement for – other Administration efforts to advance low-cost RCTs.** For example, OMB prominently featured the concept of low-cost RCTs in its May 2012 memo to the heads of the federal agencies on [Use of Evidence and Evaluation](#), and cited the [brief](#) we developed on such studies. The President’s FY

2014 budget does the same (as shown [here](#), page 94). This demonstration also falls within the Administration's broader effort to spur public sector innovation and solve pressing national problems through the use of prizes, challenges, and similar competitions (described [here](#)).

## **B. Overview of the process:**

- **We are working in cooperation with the White House and OMB to ensure a high level of visibility for this demonstration effort both within and outside the federal government.**
- **The Coalition will develop and administer the competition**, including: (i) developing the competition's announcement, including the selection criteria, in consultation with federal officials and leading researchers; (ii) using an expert panel (such as the Coalition's [Top Tier Initiative](#) panel) to evaluate the proposals and select the awardees; (iii) making awards to the competition winners to fund their RCTs, using Coalition funding provided by philanthropic foundations; and (iv) monitoring implementation of the studies as they go forward.
- **The competition will seek applications proposing to carry out an RCT in any area of domestic social policy,<sup>16</sup> that:**
  - **Is low cost** – because, for example, it would (i) evaluate an initiative or intervention that another entity (e.g., federal, state, or local agency, philanthropic foundation) is already funding or planning to fund; and (ii) use administrative data to measure key study outcomes;
  - **Evaluates an intervention that, based on prior evidence and/or a compelling rationale, could produce sizable improvement in important social outcomes, at reasonable cost** – (e.g., educational achievement, workforce earnings, criminal arrests, hospitalizations, child maltreatment, government spending);
  - **Has a valid study design that will generate credible evidence about the intervention's effect on such outcomes;**
  - **Has a project team that includes at least one researcher who has previously carried out a well-implemented RCT** (even if not at low-cost);
  - **Includes an agreement with the agency funding the intervention, that it will facilitate the RCT** – for example, by allowing the researchers to carry out random assignment of individuals to treatment and control groups; and
  - **Includes assurance from the appropriate data agency that the researcher(s) will have access to the needed administrative data.**
- **We will structure the competition in stages, so as to help facilitate partnerships among the various parties that need to come together to submit a final application** – i.e., (i) researcher, (ii) agency whose initiative will be evaluated, and (iii) data agency. The initial stage(s) of the competition will seek letters of interest from individuals or organizations in each of these three groups that are seeking to partner with others to conduct a low-cost RCT – as well as promising ideas for low-cost RCTs from researchers or others who do not seek to conduct the study themselves but are hoping that others will do so. We will then seek to connect potential partners. The final stage of the competition will invite complete applications, as described above.

- C. Timeline and number of awards:** One competition will be conducted per year over the next three years. Each will make two to three awards, averaging about \$100,000 in size, to enable the awardees to carry out their low-cost RCTs. We expect to release the first request for applications at the end of 2013.

## References

---

<sup>1</sup> The offices we are working with are: White House Domestic Policy Council / Office of Social Innovation; White House Office of Science and Technology Policy; and Office of Management and Budget.

<sup>2</sup> National Research Council and Institute of Medicine, *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*, Mary Ellen O’Connell, Thomas Boat, and Kenneth E. Warner, Editors (Washington DC: National Academies Press, 2009), recommendation 12-4, p. 371, [linked here](#).

<sup>3</sup> *CBO’s Use of Evidence in Analysis of Budget and Economic Policies*, Jeffrey R. Kling, Associate Director for Economic Analysis, November 3, 2011, page 31, [linked here](#).

<sup>4</sup> U.S. Department of Education’s Institute of Education Sciences, *What Works Clearinghouse - Procedures and Standards Handbook*, Version 2.1, September 2011, pp. 11-12, [linked here](#). U.S. Department of Education, “Scientifically-Based Evaluation Methods: Notice of Final Priority,” *Federal Register*, vol. 70, no. 15, January 25, 2005, pp. 3586-3589, [linked here](#).

<sup>5</sup> U.S. Preventive Services Task Force, “Current Methods of the U.S. Preventive Services Task Force: A Review of the Process,” *American Journal of Preventive Medicine*, vol. 20, no. 3 (supplement), April 2001, pp. 21-35.

<sup>6</sup> The Food and Drug Administration’s standard for assessing the effectiveness of pharmaceutical drugs and medical devices, at 21 C.F.R. §314.126, [linked here](#).

<sup>7</sup> Howard S. Bloom, Charles Michalopoulos, and Carolyn J. Hill, “Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects,” in *Learning More From Social Experiments: Evolving Analytic Approaches*, Russell Sage Foundation, 2005, pp. 173-235.

Steve Glazerman, Dan M. Levy, and David Myers, “Nonexperimental Replications of Social Experiments: A Systematic Review,” Mathematica Policy Research discussion paper, no. 8813-300, September 2002. The portion of this review addressing labor market interventions is published in “Nonexperimental versus Experimental Estimates of Earnings Impact,” *The American Annals of Political and Social Science*, vol. 589, September 2003, pp. 63-93.

Thomas D. Cook, William R. Shadish, and Vivian C. Wong, “Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons,” *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724-50.

John P. A. Ioannidis, “Contradicted and Initially Stronger Effects in Highly Cited Clinical Research,” *Journal of the American Medical Association*, vol. 294, no. 2, July 13, 2005, pp. 218-228.

Mohammad I. Zia, Lillian L. Siu, Greg R. Pond, and Eric X. Chen, “Comparison of Outcomes of Phase II Studies and Subsequent Randomized Control Studies Using Identical Chemotherapeutic Regimens,” *Journal of Clinical Oncology*, vol. 23, no. 28, October 1, 2005, pp. 6982-6991.

John K. Chan et. al., “Analysis of Phase II Studies on Targeted Agents and Subsequent Phase III Trials: What Are the Predictors for Success,” *Journal of Clinical Oncology*, vol. 26, no. 9, March 20, 2008.

<sup>8</sup> Coalition for Evidence-Based Policy, *Randomized Controlled Trials Commissioned by the Institute of Education Sciences Since 2002: How Many Found Positive Versus Weak or No Effects*, July 2013, [linked here](#).

<sup>9</sup> This is based on a count of results from the Department of Labor RCTs that have reported results since 1992, as identified through the Department’s research database ([link](#)). We are preparing a short summary of these findings, to be released in the next few weeks.

<sup>10</sup> Ioannidis 2005, Zia et. al., 2005, Chan et. al., 2008 (see footnote 6 for full citations).

---

<sup>11</sup> Jim Manzi, *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society*, Perseus Books Group, New York, 2012, pp. 128 and 142. Jim Manzi, *Science, Knowledge, and Freedom*, presentation at Harvard University's Program on Constitutional Government, December 2012, linked [here](#).

<sup>12</sup> Carmen DeNavas-Walt, Bernadette D. Proctor, and Jessica C. Smith, U.S. Census Bureau, Current Population Reports, P60-245, *Income, Poverty, and Health Insurance Coverage in the United States: 2012*, U.S. Government Printing Office, Washington, DC, September 2013. U.S. Census Bureau, [Official and National Academy of Sciences \(NAS\) Based Poverty Rates: 1999 to 2011](#), 2012. Kathleen Short, U.S. Census Bureau, HHES Division, *Estimating Resources for Poverty Measurement, 1993 – 2003*, 2005. Panel on Poverty and Family Assistance, National Academy of Sciences, *Measuring Poverty: A New Approach*, 1995, pp. 31-36.

<sup>13</sup> Rampey, B.D., G.S. Dion, and P.L. Donahue, P.L. *NAEP 2008 Trends in Academic Progress*, NCES 2009-479, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C., 2009.

<sup>14</sup> Cornman, S.Q., and A.M. Noel, *Revenues and Expenditures for Public Elementary and Secondary School Districts: School Year 2008-09 (Fiscal Year 2009)* (NCES 2012-313). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2011. Richard H. Barr, *Revenues and Expenditures for Public Elementary and Secondary Education, 1973-74* (NCES-76-140). U.S. Department of Health, Education & Welfare, National Institute of Education. Washington, DC: National Center for Education Statistics, 1976.

<sup>15</sup> U.S. Census Bureau, Current Population Reports, September 2013, op. cit., no. 12. This refers to inflation-adjusted income. It includes income from the economy (such as earnings) but not government transfers (such as Food Stamps). However, the evidence suggests that the overall story of income stagnation for the bottom 40% of households changes little even when one adjusts income for government transfers and other items that affect household living standards. Specifically, the Census Bureau's alternative, National Academy of Sciences-based poverty measures make adjustments for government transfers, as well as factors such as state and local taxes, work expenses such as child care, out-of-pocket medical expenses, and geographic differences in housing costs. These adjustments change the poverty rate in any given year, as well as the composition of those in poverty, but do not change the overall trend in the poverty rate over time – i.e., no overall progress since the 1970s. (The relevant citations are in endnote 12.) Although the National Academy-based poverty measures only apply to a subset of the bottom 40% of U.S. households, their corroboration of no meaningful improvement for that key subset suggest that similar findings would be obtained for the larger group.

<sup>16</sup> For most of the RCTs that are awarded, there will be no restriction on the area of domestic social policy in which they fall. Some of the RCT awards, however, may be limited to certain broad policy areas (e.g., child and family policy), consistent with the mission of the foundations funding these awards.