

CONCEPT PAPER:

Low-cost RCTs are a powerful new tool for building scientific evidence about “what works” to address major social problems

- I. **Background:** Well-conducted RCTs are regarded as the strongest method of evaluating the effectiveness of programs, practices, and treatments (“interventions”), per evidence standards articulated by the Institute of Education Sciences (IES) and National Science Foundation (NSF),<sup>1</sup> National Academy of Sciences,<sup>2</sup> Congressional Budget Office,<sup>3</sup> U.S. Preventive Services Task Force,<sup>4</sup> Food and Drug Administration,<sup>5</sup> and other respected scientific bodies.

Uniquely among study methods, random assignment of a sizable number of individuals<sup>6</sup> to either a treatment group (which receives a new intervention) or a control group (which receives services-as-usual) ensures, to a high degree of confidence, that there are no systematic differences between the two groups in either *observable* characteristics (e.g., income, ethnicity) or *unobservable* characteristics (e.g., motivation, psychological resilience, family support). Thus, any difference in outcomes between the two groups can be confidently attributed to the intervention and not to other factors. For this reason, recent IES and NSF research guidelines recommend that “generally and when feasible, [studies that measure program effectiveness] should use designs in which the treatment and comparison groups are randomly assigned.”<sup>1</sup>

- II. **Breakthrough:** Researchers have shown it is possible, in many instances, to conduct sizable RCTs at low cost, addressing a major obstacle to their widespread use, and building valuable evidence.

A. The low cost is achieved by –

1. **Embedding random assignment in initiatives that are being implemented anyway as part of usual program operations.** Government and foundations fund a vast array of strategies and approaches 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.

- B. **Such leveraging of 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.

- C. **Low-cost RCTs thus offer a powerful new vehicle for evidence-building, and an important complement to traditional, more comprehensive RCTs as part of a larger research agenda.** For example, low-cost RCTs can be a highly cost-effective tool for identifying interventions that show impacts and are therefore strong candidates for traditional RCTs. Traditional RCTs can then be used to generate valuable additional evidence about whether, under what conditions, and how to scale up the intervention so as to achieve optimal impact.<sup>7</sup>

III. **Examples:** The following are five sizable, 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. These studies all produced valid evidence of practical importance for policy decisions and, in some cases, identified program strategies that produce budget savings. (More details and citations for these studies are [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. **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, over a three-year period, 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).

C. **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 South Carolina counties. The study found that the program reduced 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).

#### 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).

### IV. 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 [Top Tier Evidence](#), [Blueprints for Healthy Youth Development](#)), 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:** In Department of Labor-commissioned RCTs that have reported results since 1992, about 75% of tested interventions 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 important 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. For this reason, we believe such studies may be most valuable when focused on interventions backed by promising prior evidence that suggests impacts will be found (*e.g.*, findings from low-cost RCTs, as noted above).

- 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.

## Endnotes

---

<sup>1</sup> Institute of Education Sciences (of the U.S. Department of Education) and National Science Foundation, *Common Guidelines for Education Research and Development*, August 2013, [linked here](#).

<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. 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>5</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>6</sup> In some RCTs, whole groups (such as schools or counties) – rather than individuals – are randomly assigned to treatment versus control conditions, but the same principle applies.

<sup>7</sup> Examples of additional evidence supplied by traditional RCTs include: (i) corroboration of the earlier impact findings in different samples and settings, thus building strong, replicated evidence of effectiveness; (ii) estimates of 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) detailed information on the services received by intervention participants, and how they differ from any services received by the control group (so as to assess how much of a contrast in services is needed to generate a meaningful impact); (v) possible reasons why the intervention produced its effect; and (vi) how the intervention's benefits compare to its costs.

<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 shortly.

<sup>10</sup> 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. Michael L. Maitland, Christine Hudoba, Kelly L. Snider, and Mark J. Ratain, "Analysis of the Yield of Phase II Combination Therapy Trials in Medical Oncology," *Clinical Cancer Research*, vol. 16, no. 21, November 2010, pp. 5296-5302. Jens Minnerup, Heike Wersching, Matthias Schilling, and Wolf Rüdiger Schäbitz, "Analysis of early phase and subsequent phase III stroke studies of neuroprotectants: outcomes and predictors for success," *Experimental & Translational Stroke Medicine*, vol. 6, no. 2, 2014.

<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](#).