Request for Proposals:

Low-Cost Randomized Controlled Trials to Drive Effective Social Spending

December 2015
The Laura and John Arnold Foundation is expanding its investment in low-cost randomized controlled trials (RCTs) designed to build policy-important evidence about “what works” in U.S. social spending. We are pleased to be partnering again in this effort with the Annie E. Casey Foundation. This document contains:

- Brief background on our investment in low-cost RCTs and the current expansion – below (2 pages);
- Request for Proposals (RFP) – first attachment (3½ pages); and
- A short concept paper on low-cost RCTs as a breakthrough in policy research that can help drive effective social spending – second attachment (3½ pages plus references).

**BACKGROUND AND PURPOSE**

In 2013, the Laura and John Arnold Foundation (LJAF), in partnership with the Annie E. Casey Foundation, funded the launch of a low-cost RCT competition. The competition was administered by the nonprofit, nonpartisan Coalition for Evidence-Based Policy, which developed the project in response to a White House and Office of Management and Budget (OMB) call to action for evidence-based reforms across the federal government. The competition was set up as a high-visibility initiative – e.g., leading to a White House conference, as discussed below.

The competition sought to demonstrate, to a broad government and philanthropic audience, that low-cost RCTs are a powerful new tool for building scientific evidence about “what works” in social spending. Well-conducted RCTs are widely regarded as the most credible method of evaluating whether a social program is effective, but are often assumed to be inherently too expensive and burdensome for practical use in most areas. Recently, however, researchers have shown that, in many instances, high-quality RCTs can be conducted at low cost and minimal burden, addressing a key obstacle to their widespread use. This is achieved by:

- **Embedding random assignment in initiatives that are being implemented anyway as part of usual program operations.** RCTs can be embedded in many new or ongoing programs, for example, by using a lottery process – i.e., random assignment – to determine who among those eligible will be offered program services (since available funds are often insufficient to serve everyone who qualifies).

- **Measuring key study outcomes with administrative data that are already collected for other purposes** (e.g., student test scores on state exams, criminal arrest records, and health care expenditures), rather than engaging in original – and often expensive – data collection through interviews or testing.

Such studies make it possible now, as never before, for policy officials to use scientific evidence about what works to increase government effectiveness (as discussed further in the attached concept paper).

The competition succeeded in funding large RCTs with strong designs, carried out by highly-capable researchers, and measuring outcomes of self-evident policy importance. The studies awarded in the first competition cycle, which are now underway, have all met their initial benchmarks – recruitment of large samples, successful randomization resulting in treatment and control groups that are highly similar in observable characteristics, and access to low-cost administrative data to measure policy-important outcomes over a sustained period of time (between two and seven years). These studies were highlighted at a White House conference on low-cost RCTs in July 2014, and received significant press and policy attention. They are summarized below along with an example of a study awarded in the second year (last bullet), illustrating the diverse areas of award:
• **A large, multi-site RCT of Bottom Line**, a program that provides one-on-one guidance to help low-income, first-generation students get into and graduate from college. This study is measuring college enrollment, persistence, and completion outcomes for a sample of about 2,400 students over a seven-year period, using administrative data from the National Student Clearinghouse. The total study cost is approximately $159,000, of which $100,000 was awarded through the competition.

• **A large RCT of Durham Connects**, a postnatal nurse home visiting program designed to improve child and mother health and well-being. The study – which seeks to replicate the positive findings from a prior RCT of Durham Connects – is using hospital administrative records to measure program impacts on families’ emergency department use and related healthcare costs through child age 24-months, for a sample of 937 families in Durham County, North Carolina. The total study cost is approximately $183,000, of which $96,000 was awarded through the competition.

• **A large, multi-site RCT of workplace health and safety inspections** conducted by the federal Occupational Safety and Health Administration (OSHA). For a sample of about 13,000 business establishments eligible for a randomized inspection, the study is testing whether being randomly chosen for inspection affects establishments’ subsequent injury rates and business outcomes (e.g., sales, business closures) over a three-to-four year period – all measured through administrative data from OSHA and other sources. The total study cost is approximately $153,000, of which $96,000 was awarded through the competition.

• **A large RCT of Multi-Tiered Systems of Supports (MTSS)**, a schoolwide reform initiative, designed to increase academic achievement and reduce behavioral problems. It is being scaled-up district-wide in Wake County, North Carolina, through a phased-in approach that will embed an RCT. Forty-four schools (the treatment group) began implementing MTSS this year, and 44 schools (the control group) will implement it two years later. The study will measure math and English test scores, behavioral suspensions, and other outcomes over the two-year period using district administrative data. The total study cost is approximately $150,000, of which $81,000 was awarded through the competition.

In spring 2015, LJAF assumed the administration of the competition, as the Coalition for Evidence-Based Policy – which launched the effort – wound down its operations as an independent organization. The Coalition’s leadership, and core elements of the group’s work, have been integrated into LJAF.

With the attached RFP, LJAF is significantly expanding its investment in low-cost RCTs. Specifically:

• The RFP increases the targeted award amount from $100,000 (in prior RFPs for the competition) to $150,000, and will allow awards of up to $300,000 if the additional cost can be justified, as described in the RFP.

• LJAF plans to fund all proposals that receive a strong rating from the proposal review panel, based on the criteria in the RFP.

• We plan to issue more than one RFP each year, to enable interested parties to apply for funding as opportunities arise. We will issue at least one additional RFP over the coming year, and plan to transition to more frequent RFPs thereafter.

The Annie E. Casey Foundation plans to contribute to the current RFP cycle in its mission area (child and family policy), as described in the RFP.
REQUEST FOR PROPOSALS

I. Overview:

A. This RFP invites grant applications to conduct low-cost RCTs in any area of domestic social policy. The targeted award amount is $150,000, but we will potentially award up to $300,000 if the additional cost can be justified, as described below. Among the grants awarded, we anticipate funding at least one RCT in the area of child and family policy that preferably utilizes integrated data systems to measure study outcomes, consistent with the interests of the Annie E. Casey Foundation as a contributor to this initiative.

B. We plan to fund all proposals that receive a strong rating from the expert review panel, based on the criteria in this RFP. The panel will be similar in composition to that used in evidence review initiatives of the Coalition for Evidence-Based Policy.

II. Application Process and Selection Criteria:

A. The following table shows the requested application materials and timeline:

<table>
<thead>
<tr>
<th>Stage of application process</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>All prospective applicants are asked to submit a letter of interest (maximum three pages)</td>
<td>Deadline: March 1, 2016</td>
</tr>
<tr>
<td>Applicants will be notified whether they are invited to submit a full proposal (full proposals must be invited)</td>
<td>By April 15, 2016</td>
</tr>
<tr>
<td>Invited applicants submit a full proposal (maximum six pages)</td>
<td>Deadline: June 1, 2016</td>
</tr>
<tr>
<td>Applicants will be notified whether they have been selected for award</td>
<td>June 2016</td>
</tr>
<tr>
<td>Grants will be awarded</td>
<td>July – August 2016</td>
</tr>
</tbody>
</table>

B. Letters of interest and invited full proposals should address each of the selection criteria below, within three pages (for the letter) and six pages (for the invited full proposal). Applicants may use their own format, with single or double spacing, and a font of 11 or larger. The page limit does not include attached letters or other documents specifically requested in this RFP. Please submit all items via email to LowCostRCT@arnoldfoundation.org.

C. Selection Criteria:

For the letter of interest: While we ask applicants to address all four selection criteria below, we do not expect applicants to have finalized all aspects of the study design and partnership agreements; therefore, reviewers will focus primarily on the other two criteria – “importance” and “experienced researcher” – in determining which applicants to invite to submit a full proposal.

For the invited full proposal: Reviewers will consider whether all four criteria are satisfied.
➢ **IMPORTANCE**: Is the applicant proposing to evaluate an intervention –

- That is backed by highly-promising prior evidence, suggesting it could produce sizable impacts on outcomes of recognized policy importance – such as educational achievement, workforce earnings, criminal arrests, hospitalizations, child maltreatment, and government spending. For example, we specifically encourage proposals seeking to replicate findings from prior rigorous evaluations that are especially promising but not yet conclusive (e.g., due to only short-term follow-up, a single-site study design, and/or well-matched comparison groups but not randomization). As a threshold condition for “highly promising” evidence, proposals should show that the intervention can be or (preferably) has been successfully delivered under real-world implementation conditions.

- or -

- For which there are other compelling reasons to evaluate its effectiveness – e.g., it is, or soon will be, widely implemented with significant taxpayer investment, and its impact on its targeted outcomes is currently unknown.

Please note that, to meet this criterion, it is not sufficient to establish that the study addresses an important problem; applicants must also present compelling reasons to evaluate the specific intervention.

➢ **EXPERIENCED RESEARCHER**: Does the applicant’s team include at least one researcher in a key substantive role who has previously carried out a well-conducted RCT (even if not low cost)? A well-conducted RCT is characterized, for example, by low sample attrition, sufficient sample size, close adherence to random assignment, and valid outcome measures and statistical analyses. To address this criterion, applicants should provide reports from prior RCTs that the researcher has conducted (please send the full study reports as email attachments – no more than two reports in all). Reviewers will rely primarily on these reports in assessing this selection criterion.

➢ **STUDY DESIGN**: Is the applicant’s proposed RCT design –

- Valid? In other words, does it have a sufficiently large sample (as shown through a power analysis) and other elements needed to generate credible evidence about the intervention’s impact on one or more targeted outcomes of high policy importance? We strongly encourage designs that measure such outcomes in both the short and longer term, as appropriate for the type of intervention and study, to determine whether the effects endure long enough to constitute meaningful improvement in people’s lives. Reviewers, in assessing an applicant’s proposed design, will use Key Items to Get Right When Conducting RCTs of Social Programs as a reference.

- and -

- Low cost? Such low cost may be achieved, for example, by (a) embedding random assignment in an intervention that government or philanthropic organizations are already funding or planning to fund; and/or (b) measuring key outcomes using administrative data that are already collected for other purposes and are of reasonable quality.

Applicants, as part of their discussion of this criterion, should specify the study’s primary outcome(s) of interest, how the outcome(s) will be measured and over what length of time, and what analyses are planned (e.g., any subgroups to be examined, regression methods to be used).
PARTNERS: Does the applicant’s team include all parties needed to conduct the RCT? Examples of necessary parties include: researcher(s), an agency delivering the intervention, and an agency housing the administrative data. To verify the existence of such partnership, the reviewers will look for attached letters or other communication showing, for example, that (a) a social service agency that delivers the intervention has agreed to participate in the study, including random assignment; and (b) a data agency has agreed to provide the researcher(s) with access to the administrative data needed to measure study outcomes.

D. Other items to include in the letter of interest and invited full proposal:

1. Applicants should specify the amount of funding requested. Our targeted award amount is $150,000. Applicants requesting a larger award – up to the limit of $300,000 – should submit an attachment (no more than one page) to the letter of interest and full proposal that provides the reasons for the higher request.

Invited full proposals (but not letters of interest) should include a one-page project budget as an attachment. In the budget, LJAF allows for project-related overhead expenses such as salaries and benefits of administrative staff, equipment, supplies, and travel to be included in direct costs; however, LJAF will only provide funding for indirect costs, up to a 10% limit, if extenuating circumstances exist and have been approved by LJAF prior to submission of an invited full proposal. (Please contact David Anderson, at the email address or phone number shown below, for additional information.)

If additional funding from other sources is needed to carry out the study, we request that the applicant’s budget show (a) the total study cost, and (b) the portion of that cost to be covered by LJAF; and include an attached letter or other communication showing that the additional funding will be in place prior to LJAF’s grant award. In such cases, the total study cost – including the additional funding – should still meet the spirit of a “low-cost RCT.”

2. Applicants should specify the proposed recipient of the grant award, which we generally expect to be a tax-exempt organization (e.g., nonprofit organization, university, or governmental unit). If an organization is not a tax-exempt organization and is interested in applying, please contact David Anderson (see contact information below).

3. Applicants should briefly address how their study meets recognized ethical standards for research with human subjects.

III. What To Expect in the Grant Agreement: Awardees will be asked, as a condition of award, to –

- Pre-register the study on the Open Science Framework (OSF) website and, prior to commencement of the study, upload a copy of the research and analysis plan described in their proposal.

- Provide us with brief phone or email updates on the study’s progress on a quarterly basis, and before making any key decisions that could materially affect the study’s design or implementation.

- Submit concise reports on the impact findings at appropriate intervals. These reports should make it easy for readers to see the study’s main results and gauge their credibility (e.g., by showing the
similarity of the treatment and control groups in pre-program characteristics, the amount of sample attrition, and the statistical significance of the impact findings).

- and -

- Make their datasets and related materials (e.g., survey instruments, code used to clean and analyze datasets) publicly available on the OSF site, unless doing so would materially hinder study implementation or raise its cost. Applicants will be asked to do this within one year of the last data collection, and only to the extent allowed under any confidentiality/privacy protections.

[Note: The above list previews the main items in the streamlined grant agreements that LJAF uses for low-cost RCTs, but is not an exhaustive list of the conditions of the award.]

IV. Questions? Please contact David Anderson, Director of Evidence-Based Policy (danderson@arnoldfoundation.org, 202-239-1248).
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), National Academy of Sciences, Congressional Budget Office, U.S. Preventive Services Task Force, Food and Drug Administration, and other respected scientific bodies.

Uniquely among study methods, random assignment of a sizable number of individuals 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.”

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.

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.  

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.9

- **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.10

- **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”).11
• **Business:** Of 13,000 RCTs of new products/strategies conducted by Google and Microsoft, 80-90% have reportedly found no significant effects.¹²

**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.
The Annie E. Casey Foundation has supported the advancement of integrated data systems, which can be a valuable resource for carrying out low-cost RCTs. Such systems can enable studies to measure sample members’ outcomes across multiple areas – such as (in a youth development RCT) high school achievement, college enrollment, and workforce employment and earnings – by linking repositories of administrative data in each of these areas.

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.


7 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.

8 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.

9 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.

10 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.
