

How low-cost, lightweight RCTs can improve program evaluation

by **Vetan Kapoor, Michael Taylor, and Ana Boltik**

In 2010, the United Kingdom's Ministry of Justice faced the daunting challenge of collecting millions of pounds in late court fines. Traditional methods of collection, such as dispatching fine collectors, were proving costly and ineffective. Partnering with the UK Cabinet Office's Behavioural Insights Team, they designed a simple experiment: use personalized text message reminders to inform delinquent citizens of their amount owed. To evaluate the effect of the text messages, they selected a random group of delinquent citizens to receive the text messages and compared their response rates to those that were subject to traditional enforcement measures. Stunningly, they saw a 33% increase in fine repayment amongst those that received a personalized text message reminder.

The White House's call to arms on RCTs and Evidence-Based Policy Making

In a July 2013 memo to the heads of all federal agencies, the Office of Management and Budget encouraged the use of experimentation and innovation to test new approaches to program delivery.

The memo identified RCTs for public sector use, stating that "many innovative companies use rapidly conducted randomized field trials to identify high impact innovations and move them quickly into production."¹

The experiment that they used is called a randomized control trial (RCT) and is a widely accepted, effective way to measure the impact of an intervention. Though RCTs are often thought to involve costly evaluations requiring deep technical expertise, it is also possible, under certain conditions for lay managers in government to deploy RCTs and harness their benefits cheaply. This

paper introduces the idea behind low-cost, light weight RCTs, similar to the trial conducted by the UK's Ministry of Justice, and identifies the conditions under which a manager might be able to successfully employ one to determine which changes to a policy, program or communication yield the best results.

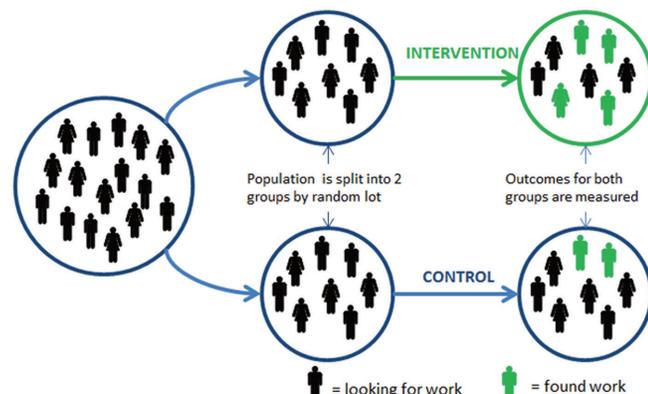
What exactly are RCTs?

RCTs are experiments in which:

1. Individuals from a population of interest are randomly assigned to either a control or intervention group. In its simplest form, the intervention group receives the

intervention whose impact is being measured—which might be a program, particular option or type of communication etc.

2. The control group does not receive the intervention.
3. After the implementation of the intervention the outcome that the intervention is designed to affect, such as a behavior, are measured separately in the intervention and control groups.
4. The difference in the outcome between these two groups is attributed to the intervention being measured.
5. In some cases, the results may be better in the intervention group, showing that the intervention increases effectiveness. If the control group yields better results, the intervention has not reached its intended level of effectiveness.



Source: Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials, Behavioral Insights Team, UK Government, 2012.

In the example illustrated in the diagram, individuals from a population of job seekers are randomly assigned into two groups. One group is given targeted job search assistance through a program designed to help the unemployed find work (intervention group), while the other group continues to make use of existing resources (control group). After the intervention is delivered, each group is measured to see how many individuals found work. As seen in this example, the group receiving the targeted job assistance has more people who found work, and the size of that difference is considered to be the effect of the program.

from data to impact
helping agencies
harness data to
transform government
effectiveness

How RCTs are helping to combat Malaria

There are over 300 million cases of Malaria each year, and about 35% of the world's population lives in risk of contracting severe types of the disease. Studies have shown that the use of insecticide treated bednets (ITNs) while sleeping can greatly reduce Malaria-related deaths. Yet use of ITNs is low, even when the benefits of bednets are explained to households and bednets are given out for free.

The Poverty Action Lab (J-PAL) at MIT conducted an RCT in India to determine why bednet usage was low despite many education campaigns. They found that the problem was not information or bednet provision, but that many villagers were reluctant to make subsequent purchases or spend money to have bednets re-treated with insecticide. The RCT provided the intervention group with microloans to help buy and re-treat bednets. The program successfully increased usage by over 12%.

Many similar randomized experiments have been conducted around the optimal pricing of ITNs. For example, researchers at Stanford University experimented with cost-sharing and differentiated pricing to determine their impact on demand for bednets.⁹

According to National Academy of Sciences, and the Office of Management and Budget, RCTs are among the best ways to test program effectiveness when executed properly.² Specifically, RCTs are considered to have high internal validity. That is to say, when they are designed and executed correctly, they excel at answering questions about causality—the connection between a particular intervention and its results.

The medical field has used RCTs for decades to evaluate the effectiveness of drugs. Recently, RCTs have gained momentum as a method for testing other policies and programs in fields such as international development.

For example, economists Esther Duflo and Abhijit Banerjee used RCTs to

test a low cost alternative to a large fertilizer subsidy program in Kenya. Their findings indicated that the Kenyan government could actually achieve similar levels of success at dramatically lower costs simply by changing the timing at which fertilizer is marketed and sold to the farmers.³

Another example of ways in which governments can use evaluation to have a greater impact in the context of resource scarcity comes from state and non-profit run Scared Straight programs. These programs expose teenagers at risk of committing crimes to the life of an inmate currently in prison, seeking to demonstrate the harsh and brutal nature of prisons and hence deter these teenagers from criminal offenses. Scared Straight programs are relatively cheap, costing about \$60 per participant. Given the high costs of trials and incarcerations, such programs might seem like a very small price to pay to help deter crime. However, research into the efficacy of these programs shows that they just don't work. For example, studies have indicated that participants in Scared Straight programs are actually *7% more likely* to commit crimes afterward than those who do not participate. This may be due to pressure from peers to demonstrate that these participants weren't in fact scared and could fit in, or may simply be that some teenagers identify with the inmates they visit. Either way, randomized trials can show what works and what does not: researchers have found several effective programs on teenagers at risk of committing crimes, including the WayOut program. Run in Snohomish

County, Washington, the program works with teenagers and their parents through a two-day educational seminar focused on helping young people build better futures. The program costs roughly \$200 per participant, but reduces chances of engaging in crime by more than 2%.⁴

Things to consider when choosing RCTs

Although RCTs have the potential to provide powerful insights, they also feature some potential risks, such as:

- Ethical issues associated with withholding beneficial programs from a control group. For example, imagine a school district seeking to test the impact of a new healthy lunch program on test score outcomes that withholds the healthy lunches from the students in control group.
- Design complexity in effectively isolating a control group. An RCT testing the impact of a new alcohol tax on total consumption in one district would likely be very ineffective, for instance, as consumers could simply choose to shop for their alcohol in a nearby district.
- High costs associated with collecting outcome data and creating and maintaining a control group. The most significant costs come from collecting good data from the two groups, and ensuring consistent and effective participation from individuals involved in the trial.

These are all valid concerns that must be considered when determining whether an RCT is right for an organization. But where it is possible to address these concerns effectively, a low-cost and lightweight experiment design can position agency leaders to gain deep insight into what works so they can direct energy and resources to effective interventions.

Using "Randomized Encouragement" to Overcome Ethics Challenges

Michael Greenstone, a professor at MIT, used an RCT and an approach known as "randomized encouragement" to evaluate the impact of low-income weatherization assistance programs in Michigan.

Greenstone and his partners did not want to deny access to a program for which families below a certain income threshold qualified. The team identified several families who were eligible for assistance but had not yet signed up. From this group, the researchers randomly selected two sub-groups, and encouraged one set of families to sign up for the program via post-cards. They were thus able to ensure randomized take up of the benefit without denying anyone services.

Low-cost, lightweight RCTs could be a game changer

Advances in technology are making low-cost lightweight experiment designs more viable for the first time. Many platforms allow for easy separation of the control and intervention groups. In addition, mobile technology and web-based interaction allow for quick and inexpensive data collection. Rapid experimentation using low-cost, lightweight RCTs can help agency leaders address tough challenges, evaluate programs, and improve the effectiveness of government programs. So what when might a low-cost and lightweight be viable for federal agencies?

The data is already being collected and/or easily accessible

Many agencies run programs for which they already collect or have access to large amounts of administrative data. If this data aligns to desired outcome measures, experimenters may be able to make comparisons with minimal added cost.⁵ The City of New York was able to take advantage of this when using a low-cost light weight RCT to evaluate their \$75 million teacher incentive program. The City was already collecting relevant outcome measures such as standardized test scores, GPA, attendance etc. The trial demonstrated that the incentive program did not lead to improved outcomes the City was able to allocate funds more effectively elsewhere.⁶

Control groups can be easily isolated and ethical concerns can be minimized

Current polling suggests that while 9 out of 10 people support organ donation, fewer than 1 in 3 people actually register. The UK Behavioral Insights team used an RCT to test the effects of different messages encouraging citizens to register on high traffic webpages on the gov.uk website. They found that if the best performing message was used over the entire year, it would have led to roughly 96,000 extra registrations as opposed to the control condition.⁷ The control and intervention groups were easily isolated by simply varying which message was displayed on the website. In this case, there were minimal ethical concerns associated with showing various messages.

The organization can measure outcomes rapidly

Wikipedia's fundraising committee sought to conduct an RCT to increase overall donations to its Foundation. The donation window was relatively small, and so the committee needed results that could meaningfully impact contribution levels quickly. They sent people visiting their site to different landing pages to determine which messages would best motivate donors. Wikipedia monitored donation levels to see which versions were most persuasive. They found that donations to Wikipedia increased significantly among those people who landed on a version of the webpage that displayed a picture of founder Jimmy Wales.⁸ Because conclusions were derived within days, not years, Wikipedia was able to effectively use the RCT to improve program performance and quickly demonstrate results.

There is buy-in from stakeholders and a genuine desire to learn from findings

RCTs are most effective when they are being used to identify insights that agency leaders or program managers genuinely want to deploy. For example, if an agency is considering different options to expand a program or increase take-up, RCTs can help test the efficacy of these options through a pilot. Agencies can then invest their scarce resources on approaches that appear successful at pilot stage. Low-cost RCTs are particularly useful in these circumstances, as they allow for rapid results. In addition, over time, it is possible to run a series of back-to-back low cost RCTs that build on earlier results and allow approaches to evolve and improve. Without the commitment from program managers to learn from and adapt based on findings, such experiments may not be as effective.

RCTs and the move toward data-driven analysis of programs are creating a multitude of possibilities for federal agencies. In many cases, these agencies are already equipped with treasure troves of data and a desire to identify the most effective methods for accomplishing key objectives. While RCTs may not be appropriate in many cases, when they are, they have the capacity to help program managers build an evidence base of what works and what does not. Such evidence is particularly important in an era of tightening budgets. In situations where there is frequent customer interaction, administrative data is readily available, the program's goals are measurable, and a simple method for creating and maintaining control groups exists, low-cost, lightweight RCTs can be very useful. If these contexts apply, it may be the right time to consider conducting an RCT to improve program performance.

¹ "Memorandum to the Heads of Departments and Agencies," Office of Management and Budget, July 2013, <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>

² See, for example, "What Constitutes Strong Evidence of a Program's Effectiveness?" Office of Management and Budget, 2004, http://www.whitehouse.gov/sites/default/files/omb/part/2004_program_eval.pdf.

³ Duflo, Esther and Abhijit Banerjee, *Poor Economics*, 2012, Public Affairs Publishing. See also <http://www.nber.org/papers/w15131>.

⁴ "Doing What Doesn't Work." Jitinder Kohli, *Center for American Progress*, <http://www.americanprogress.org/issues/open-government/news/2012/02/08/11108/doing-what-doesnt-work/>

⁵ Interview with Mary Ann Bates (J-PAL), Coalition for Evidence brief

⁶ Coalition for Evidence—Rigorous Program Evaluations on a Budget

⁷ UK Behavioral Insights Team Applying Behavioural Insights to Organ Donation: preliminary results from a randomised controlled trial

⁸ http://meta.wikimedia.org/wiki/Fundraising_2010/Banner_testing

⁹ Free distribution or cost-sharing? Evidence from a Randomized Malaria Prevention Experiment, *The Quarterly Journal of Economics*, 2010, <http://www.stanford.edu/~pdupas/CohenDupas.pdf>

To learn more about Deloitte's approach to government performance management, visit our From Data to Impact practice site & read our most recent reports: www.deloitte.com/us/fromdatatoimpact

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, "the Deloitte Network") is, by any means of this publication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

Copyright © 2014 Deloitte Development LLC. All rights reserved. Member of Deloitte Touche Tohmatsu Limited

To learn more, contact:

Jitinder Kohli

Director

Deloitte Consulting LLP

jkohli@deloitte.com

+1 571 858 0821

Leslie Fierro

Specialist Master

Deloitte Consulting LLP

lfierro@deloitte.com

+1 213 553 1269