GDN's Next Horizons Essay Contest 2014*
THE FUTURE OF DEVELOPMENT ASSISTANCE
Supported by the Bill and Melinda Gates Foundation

Winning Entry

ADAPTIVELY DEPLOYING AND EVALUATING AID: AN INTEGRATED APPROACH

Abstract
Recent advances in the design and analysis of randomized controlled trials have resolved many concerns about the fairness, efficiency and limitations of experimentation. In *adaptive controlled trials*, who and how many recipients get an experimental treatment is determined in part by the observed performance of the intervention thus far. When an intervention is shown to be promising in a particular place or for a particular institution in a given study, resources are shifted to optimize expected welfare in a manner that is anticipated by the experimental design. Adaptive controlled trials thus provide a principled way to resolve the tension between maximizing aid impact and preserving scientific rigor in evaluation. We argue that, insomuch as we seek to apply the same standards to aid that we apply to medicine, adaptive controlled trials provide a state-of-the-art approach for optimizing human well-being and knowledge.

Authors
Will Portman, Peter M. Aronow, and Allison Carnegie

**The GDN Next Horizon Essay Contest was launched globally by the Global Development Network (GDN) in 2014, with the support of the Bill & Melinda Gates Foundation. The contest invited contributions related to the future of development assistance to inform the ongoing discourse on development assistance with fresh thinking, and revamp policy debates with new voices. This essay is one of 13 winning entries selected by a high-profile Jury of aid policy makers, experts and practitioners chaired by Nancy Birdsall.*

Introduction

The Old Testament states that King Nebuchadnezzar "ordered a group of his subjects to eat rich meat and drink wine while another group was made to adhere to vegetarianism [and drink water] in order to evaluate the merits of the two diets."[1] In this study, Nebuchadnezzar altered one variable, diet, between two groups. He waited for a period of time, 10 days. Then he made observations about each group's outcomes (the water-drinking vegetarians looked "healthier and better nourished"[2] than the wine-drinking meat-eaters), concluding that the discrepancy in outcomes between the two groups must be attributable to the one difference in treatment conditions applied to them. Therefore, he reasoned that less-rich foods were less corrupting to the body and vice-versa.

Perhaps Nebuchadnezzar should have conducted a follow-up evaluation in order to separate and gauge the relative effects of the components of each diet in the trial—wine and meat for one group, and water and vegetables for the other. Perhaps he should have randomly assigned people to the treatment and control groups (he did not) and made the groups larger (they were quite small) in order to ensure that the difference in outcomes between them could be attributed to the treatment variable alone and not to selection bias, confounding variables, or random chance.

Despite the imperfections in his study, Nebuchadnezzar did manage to set up the basic elements of a valid experimental trial: a treatment variable, treatment and control groups, and observations over time in order to measure the treatment's effects.

Randomized Controlled Trials

The sort of experimental design Nebuchadnezzar carried out, improved by the additions of random assignment and appropriate controls and hence called a *randomized controlled trial* (RCT), is often associated today with medical research. In order to test, say, the extent to which a new cancer drug is effective and safe, medical researchers assemble a sample that is representative of the population they're interested in and large enough to obtain the desired confidence level and confidence interval. They randomly assign people to treatment and control groups, maintaining equipoise between the two groups except for the key variable of the drug in question. Then they make observations in order to measure the effects of the drug on the treatment population.

---

[1] Eble, Boone and Elbourne, *Risk and Evidence of Bias in Randomized Controlled Trials in Economics*, 1.

[2] Daniel 1: 15 (New International Version).

Since the publication in the 1940s of a series of influential articles about research methodology in the *Journal of the American Medical Association*, medical researchers have conducted hundreds of thousands of RCTs.[3] Because RCTs have been "shown by several studies to yield less biased treatment effect estimates than observational studies," they have come to represent "the 'gold standard' of evidence" in medical research.[4]

Despite Nebuchadnezzar's RCT prototype and RCTs' prevalence in modern medical research, the technique has only begun to emerge as a research tool in development economics in the past two decades. A group of economists dubbed the "randomistas" by *Wired* magazine, including Abhijit Banerjee and Esther Duflo at the Massachusetts Institute of Technology, Michael Kremer at Harvard, and Dean Karlan at Yale, have begun to call for, and to implement in their own research, a similar level of empirical rigor as, for example, that which new pharmaceuticals are subjected to in the Food and Drug Administration (FDA)'s drug approval process.[5]

Application to Development Economics

The randomistas have striven to transform development economics, a field long dominated by "macroeconomic theories, anecdotal evidence, and good intentions," into one where public policies are subjected to the same exacting standards for evidence as medical interventions.[6] As Esther Duflo puts it, "These economics I'm proposing, it's like 20th century medicine. It's a slow, deliberative process of discovery. There is no miracle cure, but modern medicine is saving millions of lives every year, and we can do the same thing."[7]

RCTs have several features to recommend them as an effective means of establishing a fuller and more nuanced understanding of the complex causal chains between inputs and outcomes across various contexts in development assistance. First, quite simply, they are able to generate credible empirical data that can be used to validate or invalidate causal inferences.

As Henry E. Brady writes in "Causation and Explanation in Social Science," of three common methods used to develop causal inferences—correlational evidence, counterfactual reasoning, and experimental designs—the first two are insufficient. Correlation can identify relationships between variables but cannot establish whether

---

[3] Eble, Boone, and Elbourne, *Risk and Evidence of Bias*, 1.

[4] Ibid.

[5] Benko, "The Hyper-Efficient, Highly Scientific Scheme to Help the World's Poor."

[6] Ibid.

[7] Duflo, "Social Experiments to Fight Poverty."

and to what extent one variable affects another. Counterfactual thought experiments lack evidence to verify their conclusions. The third method for getting at causality, however, experimental designs, offers a compelling synthesis of the real-world applicability of correlations and the logic of counterfactual reasoning.[8]

RCTs are one type of experimental design. There are also less-contrived, observational experiments called natural or quasi-natural experiments; however, these depend on identifying random or as-if-random treatment assignment somewhere in the world at some point in time. Additionally, they lack the deliberately designed and implemented controls that are a hallmark of RCTs. Comparing RCTs to natural or quasi-natural experiments, Thad Dunning, a practitioner and proponent of observational studies in some contexts, writes that "Because they are not so much planned as discovered, using natural experiments … involves an element of luck. ... The causes that Nature deigns to assign at random may not always be the most important causal variables for social scientists."[9]

Commenting on RCTs' superior ability to be systematically applied to questions in the social sciences, Jasjeet Sekhon writes that "the only designs I know of that can be mass produced with relative success rely on random assignment. Rigorous observational studies are important and needed. But I do not know how to mass produce them."[10] On the topic of RCTs' credibility advantage over natural or quasi-natural experiments, Abhijit Banerjee and Esther Duflo write that "the identifying assumptions [for natural experiments] are not directly testable, and the validity of any particular study depends instead on how convincing the assumptions appear."[11] Donald Green and Alan Gerber write that "only randomization provides a procedure for generating instrumental variables that are valid on their face."[12] Joshua Angrist and Jörn-Steffen Pischke write that randomization represents "the experimental ideal."[13] Clearly, there are compelling arguments and strong support for the use of RCTs in development economics.

Drawbacks and Criticisms

---

[8] Brady, "Causation and Explanation in Social Science."

[9] Dunning, *Natural Experiments in the Social Sciences: A Design-Based Approach*, 3.

[10] Sekhon, "Opiates for the Matches: Matching Methods for Causal Inference," 503.

[11] Duflo, Glennerster, and Kremer, "Using Randomization in Development Economics Research: A Toolkit," 10.

[12] Green and Gerber, "Reclaiming the Experimental Tradition in Political Science," 813.

[13] Angrist and Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 11–24.

However, there are also some disadvantages associated with the technique. Critics argue, among other things, that RCTs emphasize the accuracy of answers over the relevance of questions—empirical validity over explanatory power.[14] "Randomized evaluations do pretty well when they are targeted closely at the policy change under consideration," Dani Rodrik writes, "but less so when they require considerable extrapolation."[15] An *Economist* article has questioned whether someday, RCTs' advocates "might even find a way to apply them to the sweeping assertions of macroeconomists," implying that they currently add little insight to the big debates in the field.[16]

Even when clearly relevant RCTs can be designed, concerns remain. Critics have questioned the affordability and the ethical implications of RCTs. Martin Ravallion writes that "the scale of the randomized trials needed to test even one large national program could well be prohibitive."[17] Casey Mulligan writes that "randomization is often times neither the most economical nor the most ethical way to learn how the world works,"[18] adding:

> Prof. Jeffrey Sachs's Millennium Villages Project, an ambitious effort to help African villages escape poverty, has been criticized for, among other things, failing to randomly assign its treatments.
>
> But Professor Sachs didn't accidentally forget to randomize his assistance. He thinks that it's wrong to withhold from poor people assistance that he's confident can help. The patients who get placebos in randomized F.D.A. trials would probably agree.
>
> The problems with randomized trials cannot be dismissed as mere philosophical challenges, because people react to the poor treatment they get from experimenters. Why should a patient agree to let a dice or random number generator decide his fate?[19]

Ravallion and Mulligan are not alone in their concerns. Just as randomization has, in recent years, developed a strong following among development economists, so too has

---

[14] Sartori, "Concept Misformation in Comparative Politics."

[15] Rodrik, "The New Development Economics: We Shall Experiment, but How Shall We Learn?" 5.

[16] *Economist*, "Random Harvest."

[17] Quoted in Rodrik, "The New Development Economics," 22.

[18] Mulligan, "The Economics of Randomized Experiments."

[19] Ibid.

it accumulated its share of skeptics and detractors, including Jeffrey Sachs, Lant Pritchett and Angus Deaton.[20] However, it might be possible in some cases to ameliorate concerns about RCTs' ability relevance, cost, and ethical implications by turning to a recent spin-off of RCTs called *adaptive clinical trials* (ACTs).

Adaptive Clinical Trials

ACTs have emerged over the past two decades as an effective, credible and innovative method of conducting medical research; both the FDA and the European Medicines Agency have recognized them as valid alternatives to traditional RCTs in their drug approval processes.[21] ACTs are similar to RCTs in that they feature treatment and control groups as well as a randomly assigned treatment variable or variables. They differ in that they take into account interim data from the trial in order to modify the characteristics of the treatment variables and/or the composition of the treatment groups during the course of the trial.[22]

A Brookings Institution report summarized the technique like this:

> One promising approach for modernizing clinical trials and maximizing their efficiency is using data accumulated during the trial to inform their design. While traditional trials have fixed parameters that are determined in advance and held constant throughout the trial, "adaptive" trials allow for certain parameters—such as treatment regimen, study population, and sample size—to be modified based on interim results. These preliminary analyses conducted during an ongoing trial can also be used to stop a trial early if the product is unlikely to meet its target endpoint. Or, to drop certain treatment arms that appear less effective, which helps to avert failure, additional costs, and unnecessary risk to patients further down the line.[23]

The dynamism inherent in ACTs presents challenges for preserving equipoise between treatment and control groups. But it is possible to preserve ACTs' internal validity using

---

[20] Sachs, "Millennium Villages Project"; Pritchett, "RCTs in Development: Lessons from the Hype Cycle"; Deaton, "Instruments, Randomization, and Learning about Development."

[21] Food and Drug Administration, *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics*; European Medicines Agency, "European Medicines Agency Launches Adaptive Licensing Pilot Project."

[22] Berry, "Adaptive Clinical Trials: The Promise and the Caution."

[23] Daniel et al., *Right Drug, Right Patient: Streamlining Clinical Trials to Speed Drug Development.*

Bayesian statistical techniques, which allow significantly greater flexibility within the parameters of the trial than would be possible using a traditional fixed randomization approach.[24]

Potential Benefits over Randomized Controlled Trials

Just as development economists have been inspired by the successful use of RCTs in medical research and have begun incorporating the technique into their own research methodology, so too might they now consider adding ACTs to their toolkit, using them to address problems and questions for which traditional RCTs may be not be feasible, scalable or ethical—or at least not perceived as ethical.

ACTs' responsiveness to interim evidence is appropriate for an evaluation method intended to be integrated into aid delivery programs, in which providing the most effective humanitarian assistance based on the current consensus is often a far more urgent priority than learning about what might be a better way of doing things in the future. The combination of validity and flexibility in ACTs makes them an optimal evaluation system to be integrated into real-world aid delivery programs.

In many cases, ACTs are more cost-effective than traditional RCTs. First, embedding them into aid delivery programs obviates the need for separate disbursements for stand-alone evaluations, enabling larger-scale evaluations than have been feasible in the past. Second, the speed with which ACTs can generate useful evidence as compared to that of traditional RCTs, due to their ability to pursue promising treatments and drop nonpromising treatments during the course of a study rather than only after its conclusion, means that adaptive trials frequently conclude significantly sooner than fixed randomization trials, resulting in lower operating costs and the ability to begin implementing findings sooner.

Cost effectiveness was one of the primary reasons the FDA decided to develop guidelines for ACTs and recognize them as an option for drug-testing, as "the number of drugs invented per billion dollars invested in R&D has been nearly cut in half every nine years for the last fifty years."[25] ACTs' flexibility means that if it becomes apparent before a trial has concluded that a treatment is working, or that a certain form of a treatment is more effective than another, then study participants don't need to be deprived of the treatment until after the study's conclusion simply to preserve a control group.

Criticism of RCTs by Mulligan, Sachs and others often assumes that the control group would receive the same aid the treatment group receives were it not for the trial. While

---

[24] Ibid. For a helpful overview of the use of Bayesian statistics in ACTs, see "Bayesian Adaptive Designs for Clinical Trials" by Yi Cheng and Yu Shen.

[25] Cheng and Shen, "Bayesian Adaptive Designs."

that assumption is inaccurate and undervalues such trials' long-term benefits, it is widespread, and ACTs provide a compelling alternative. With adaptive designs, people who need humanitarian aid can receive it, and research can still go forward. The relatively seamless combination of aid delivery and evaluation that ACTs offer—a combination of policy and research, of doing and learning, and of alleviating suffering now while discovering how to more effectively alleviate suffering in the future—is one of their key advantages.

Francis Giles, MD, then affiliated with the M.D. Anderson Cancer Center in Houston, said in 2006, "I see no rationale to further delay moving to [adaptive] designs. They are more ethical, more patient-friendly, more conserving of resources, more statistically desirable."[26] Giles has conducted multiple Bayesian-based leukemia studies.

The FDA, for its part, has written:

> Compared to non-adaptive studies, adaptive design approaches may lead to a study that (1) more efficiently provides the same information, (2) increases the likelihood of success on the study objective, or (3) yields improved understanding of the treatment's effect (e.g., better estimates of the dose-response relationship or subgroup effects, which may also lead to more efficient subsequent studies).[27]

The Brookings Institution noted that ACTs "could have particular value in enhancing the development of personalized therapies. They allow investigators to learn much more about products during development and to identify the most responsive patient subpopulations and the best drugs for these individuals."[28] While all of these perspectives concern adaptive trials in medical research, it stands to reason that if economists have successfully adapted medical RCTs for the social sciences, they could do the same with ACTs.

Application to Development Economics

While ACTs do not resolve all of the concerns associated with the use of RCTs to study the effects of development assistance, in many cases they make it possible to improve on existing strategies. For example, a widely debated question in development economics is which types of cash transfers best secure desired socioeconomic outcomes. Mexico's Oportunidades program and Brazil's Bolsa Família program have emerged as prominent models for conditional cash transfer programs, which the *New York Times* has praised for aiming to "combat poverty today while breaking the cycle of

[26] Quoted in McCarthy, "Is It Time to Change the Design of Clinical Trials?"

[27] Food and Drug Administration, *Guidance for Industry.*

[28] Daniel et al., *Right Drug, Right Patient*.

poverty for tomorrow."[29] A recent World Bank report concluded that conditional cash transfers "generally have been successful in reducing poverty and encouraging parents to invest in the health and safety of their children."[30]

However, there is support for nonconditional cash transfers as well, such as the finding by Christopher Blattman, Nathan Fiala, and Sebastian Martinez that "unsupervised grants" of several hundred dollars significantly increased business assets, work hours, and earnings among youth in Uganda.[31] In the area of cash transfers, ACTs can help identify which types of transfers are most effective for which segments of the population across varied contexts.

The standard RCT approach to the question of cash transfers, such as that carried out by Blattman, Fiala, and Martinez, does generate learning: New knowledge is gained with the successful completion of each trial. The process of gaining this knowledge, though, is often slow, expensive, unresponsive to interim data, and limited in the number of policy variations that can be tested. An alternative to the standard RCT model, and largely the norm in common practice, it simply to roll out one version of a policy to an entire population. However, it is difficult if not impossible to carry out a valid and meaningful evaluation of this approach.

An ACT, on the other hand, would begin with a range of treatments, each thought to be potentially viable within the range of uncertainty concerning cash transfers. Treatment proportions would be adjusted based on preliminary experimental results. Because ACTs can be rolled out in a manner that allows experimenters to assess where treatments are having the greatest effect, experimenters would be able to identify and target, over the course of the study, groups that would most benefit from cash transfers.

Conclusion

ACTs, thus, would ultimately facilitate valuable learning about when and under which circumstances cash transfers should be used, and in which form they should be deployed, facilitating a more effective and ethical distribution of resources—while gathering the information needed to do this in a more timely and cost-effective way.

Esther Duflo has said that "you can put social innovation to the same rigorous, scientific tests that we use for drugs. And in this way, you can take the guesswork out of

---

[29] Rosenberg, "To Beat Back Poverty, Pay the Poor."

[30] Fiszbein and Schady, C*onditional Cash Transfers: Reducing Present and Future Poverty*, xi.

[31] Blattman, Fiala and Martinez, "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda."

policymaking by knowing what works, what doesn't work and why."[32] The logical continuation of development economics' embrace of the empirical rigor of medical trials is for economists to take advantage of the integration of deployment and evaluation found in adaptive clinical trials.

Bibliography

[32] Duflo, "Social Experiments to Fight Poverty."

Angrist, Joshua D., and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2009.

Banerjee, Abhijit, and Esther Duflo. "The Experimental Approach to Development Economics." *Annual Review of Development Economics* 1, no. 1 (2009): 151–178. http://economics.mit.edu/files/3158.

Benko, Jessica. "The Hyper-Efficient, Highly Scientific Scheme to Help the World's Poor." *Wired*, November 12, 2013. http://www.wired.com/2013/11/jpal-randomized-trials/.

Berry, Donald A. "Adaptive Clinical Trials: The Promise and the Caution." *Journal of Clinical Oncology* 29, no. 6 (February 2011): 606–609. http://jco.ascopubs.org/content/29/6/606.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics* 129, no. 2 (2014): 697–752.

Brady, Henry E. "Causation and Explanation in Social Science." In *The Oxford Handbook of Political Methodology,* edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 2010. Abstract available at http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-10.

Cheng, Yi, and Yu Shen. "Bayesian Adaptive Designs for Clinical Trials." *Biometrika* 92, no. 3 (September 2005): 633–646. http://biomet.oxfordjournals.org/content/92/3/633.abstract.

Daniel, Gregory W., Derek Griffing, Erica Socker, and Sophie Mayer. *Right Drug, Right Patient: Streamlining Clinical Trials to Speed Drug Development.* Washington, DC: Brookings Institution, 2014. http://www.brookings.edu/blogs/up-front/posts/2014/07/02-clinical-trials-pharmaceutical-drug-development.

Deaton, Angus. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48, no 2 (June 2010): 424–455. https://www.aeaweb.org/articles.php?doi=10.1257/jel.48.2.424.

Duflo, Esther. "Social Experiments to Fight Poverty." Recorded February 2010. TED Talk video, 16:37. http://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty/transcript#t-929000.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics,* vol. 4, edited by T. Schultz and John Strauss, 3896–3957. Amsterdam and New York: North Holland, 2008 .

Dunning, Thad. *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge University Press, 2012.

Eble, Alex, Peter Boone, and Diana Elbourne. *Risk and Evidence of Bias in Randomized Controlled Trials in Economics*. Discussion Paper No. 1240, Centre for Economic Performance, London, September 2013. http://cep.lse.ac.uk/pubs/download/dp1240.pdf.

Economist. "Random Harvest." *Economist*, December 14, 2013. http://www.economist.com/news/finance-and-economics/21591573-once-treated-scorn-randomised-control-trials-are-coming-age-random-harvest.

European Medicines Agency. "European Medicines Agency Launches Adaptive Licensing Pilot Project." Press release, March 19, 2014. http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2014/03/news_detail_002046.jsp&mid=WC0b01ac058004d5c1.

Fiszbein, Ariel, and Norman Schady. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC: World Bank, 2009. https://openknowledge.worldbank.org/bitstream/handle/10986/2597/476030PUB0Cond101Official0Use0Only1.pdf?sequence=1.

Food and Drug Administration. *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics.* Washington, DC: Food and Drug Administration, 2010. http://www.fda.gov/downloads/Drugs/Guidances/ucm201790.pdf.

Green, Donald P., and Alan S. Gerber. "Reclaiming the Experimental Tradition in Political Science." In *Political Science: The State of the Discipline*, 3rd ed., edited by Helen V. Milner and Ira Katznelson, 805–832. New York: W.W. Norton, 2002.

McCarthy, Alice. "Is It Time to Change the Design of Clinical Trials?" *Cure*, June 8, 2006. http://www.curetoday.com/index.cfm/fuseaction/article.show/id/2/article_id/371.

Mulligan, Casey B. "The Economics of Randomized Experiments." *New York Times*, March 5, 2014. http://economix.blogs.nytimes.com/2014/03/05/the-economics-of-randomized-experiments/.

Pritchett, Lant. "RCTs in Development: Lessons from the Hype Cycle." Center for Global Development blog, November 14, 2013. http://www.cgdev.org/blog/rcts-development-lessons-hype-cycle.

Rodrik, Dani. *The New Development Economics: We Shall Experiment, but How Shall We Learn?* Harvard Kennedy School Faculty Working Paper, October 2008. http://www.hks.harvard.edu/fs/drodrik/Research%20papers/The%20New%20Development%20Economics.pdf.

Rosenberg, Tina. "To Beat Back Poverty, Pay the Poor." *New York Times*, January 3, 2011. http://opinionator.blogs.nytimes.com/2011/01/03/to-beat-back-poverty-pay-the-poor/.

Sachs, Jeffrey D. "Millennium Villages Project." *New York Times*, September 6, 2013. http://www.nytimes.com/2013/09/07/opinion/millennium-villages-project.html.

Sartori, Giovanni. “Concept Misformation in Comparative Politics.” *American Political Science Review* 64, no. 4 (1970): 1033–1053. http://www.sant.ox.ac.uk/people/knicolaidis/sartori.pdf.

Sekhon, Jasjeet S. “Opiates for the Matches: Matching Methods for Causal Inference.” *Annual Review of Political Science* 12 (2009): 487–508. http://sekhon.berkeley.edu/papers/opiates.pdf.