

ARGUMENT

Economics' Biggest Success Story Is a Cautionary Tale

Field experiments now dominate development economics—often at the expense of the world's poor.

BY **SANJAY G. REDDY** | OCTOBER 22, 2019, 9:56 AM

When I was a graduate student of economics in the early to mid-1990s, a new idea was just starting to emerge in the field of global development: using randomized controlled trials (RCTs), of the sort that had long been common in medicine, to assess efforts to assist the poor. One of the very first of these studies tested the impact of **eradicating parasitic worms** on school attendance among children, with the researchers picking schools randomly to determine not only how children in these schools were affected but also neighboring ones. Researchers have since tested the impact of placing additional teachers in a classroom or monitoring teachers' attendance with cameras; the **effect** of access to bank or microfinance loans; and even the **effect** of specific appeals made by candidates in an election campaign on voting behavior.

The growing interest in RCTs has culminated in the awarding of the Nobel Memorial Prize in Economic Sciences last week to several of its pioneers: Esther Duflo, Abhijit Banerjee, and Michael Kremer. The RCTs they have promoted were **described** by the Royal Swedish Academy of Sciences as having come to “entirely dominate development economics.” The prize committee suggested the rise to centrality of this previously marginal idea was evidence of scientific progress and of a breakthrough that much better enabled us to “improve the lives of the worst-off people around the world.” We should all be glad if it were so simple. The fact that RCTs now so thoroughly shape development economics may be less a success story than a cautionary tale.

The RCT trend has been fueled by two factors: one from within

economics, the other from outside it. Within the discipline, RCTs promised to address a problem that had bedeviled economists' efforts to empirically assess development programs—namely, that the people who fared better when there was a change in circumstances were often those who were also more motivated or better positioned in some way to take advantage of it. There was no sure way of telling apart which interventions seemed to work because of such factors from which ones worked, well, because they worked. A premise among those who used statistical methods to address this problem had been that experiments on people were not possible. The “randomistas”—as they later came to be called—cheekily turned this idea on its head, proposing precisely to try such experiments.

The factor from outside of the discipline fueling the rise of RCTs was a widespread collapse of faith in the ability of public policies durably to change economic fates (culminating in the worldwide wave of structural adjustment policies in the 1980s and 1990s, bringing about austerity and market-oriented **policy reforms** in almost a hundred countries) and doubt too about the efficacy of international aid in the face of ongoing economic stagnation in large parts of the world. In this context, there was a growing interest in distinguishing what worked in development from what didn't, with the idea that successful small-scale interventions could be made bigger by receiving adequate support from nongovernmental organizations, aid agencies, private foundations, impact investment funds, and governments. There was too an inclination to elevate explanations of development failure and success centered on individuals making the best of their circumstances, perhaps with the help of specific interventions (hence, for instance, a growing fascination with microfinance).

Against this background, an RCT wave swept the world. There have been a **thousand or more** trials by now, and some of these have informed funding decisions and policies on a national or international scale. Indeed, some **aid agencies** and government bodies have strongly preferred funding interventions that are validated by an RCT. A movement of so-called ethical altruists also argues that it is only sensible to **give money** to interventions that have been found to be high impact according to an RCT and as a consequence focuses on a very narrow range of development initiatives, such as deworming and malaria treatment. The RCT movement has even started to touch rich countries, with RCTs being applied to determine the efficacy of different schooling methods in the **United States** and elsewhere. (In fact, RCTs had **originated** in an earlier era with experiments on the behavioral effects of income security and health insurance schemes in the United States in the 1960s and 1970s, but this had been largely forgotten.)

RCTs grew from strength to strength, celebrated by the media as a clever idea leading to a revolution in how poverty could be addressed; endorsed by politicians, who were often instrumental in enabling trials to be implemented; and receiving massive support from private and public funding agencies. RCTs received very little criticism from within the profession for perhaps a decade and a half. Then around 2010 the dam broke, as other economists, both those working on development and those interested in statistical methods, including some of great eminence within the mainstream of the discipline, began to point to weaknesses in the randomistas' arguments.

This countercharge from within the discipline has had three critiques: insight, reliability, and adequacy.

The insight critique contested the proposition that RCTs had revealed significant new facts or provided new understanding of development processes that would not have been had otherwise. RCTs take a long time and cost a lot (running quite easily into hundreds of thousands of dollars each). But closer inspection reveals that they most often merely provide a validation of common sense. Whereas at times randomization seemed to reveal something surprising (for instance, claiming to have shown that microfinance was less effective than many had assumed), in other instances it simply told us what had been long expected (for instance, that providing treatment for diseases benefits the community at large). One such finding—that providing preventative public health treatments at low or no cost, or better yet with incentives, leads to an increase in the number of people willing to accept them—is cited by the prize committee as having led to a change in the received wisdom in favor of user fees in primary health. This gets the history quite wrong, since such fees had long before that lost favor, due in part to activists, including prominent economists such as Jeffrey Sachs, who had made it a focus of his advocacy. I know because I was myself involved in this debate in the late 1990s, when the World Bank, World Health Organization, and other institutions were still promoting them.

RCTs cannot reveal very much about causal processes since at their core they are **designed** to determine whether something has an effect, not how. The randomistas have attempted to deal with this charge by designing studies to interpret whether variations in the treatment have different effects, but this requires a prior conception of what the causal mechanisms are. The lack of understanding of causation can limit the value of

any insights derived from RCTs in understanding economic life or in designing further policies and interventions. Ultimately, the randomistas tested what they thought was worth testing, and this revealed their own preoccupations and suppositions, contrary to the notion that they spent countless hours listening to and in close contact **with the poor**. It is not surprising that economists doing RCTs have therefore been centrally concerned with the effects of incentives on individual behavior—for instance, examining the idea that contract teachers who fear losing their jobs will be more effective than those with a guarantee of employment.

But valuable innovations in everyday life, whether on the small or large scale, are likely to result from explorations of a more open-ended kind. This requires that people experiment with the institutions of which they are a part, which is not the same as conducting randomized experiments on people. Policies (and reforms of policies) that go beyond one dimension are essential in a complex environment. For instance, better schools are likely to result both from measures dealing with teachers' employment and ones dealing with curriculum, community participation, and funding arrangements. RCTs simply cannot advise us on how best to combine all of these, let alone on how to think creatively about them. Better schools may also result from changes that result from improvements in other domains beyond the individual school—for instance, safer neighborhoods, better drug policy, or lessened poverty. The actions needed to achieve better outcomes may sometimes only be possible to undertake at a level going much beyond the locality. A good example is provided by the iodization of salt, which has **contributed** not only to better health but may also have improved educational outcomes.

And so it should not be a surprise that RCTs played no role at all in some of the greatest development successes of the past (including the creation of a free and universal public education system and widespread public health measures in the 19th and 20th centuries in the United States and other countries). In medicine, it has long been recognized that interactions between drugs, and treatments generally, require that the results of individual RCTs be acted on with great care. In addition, improving health requires combining medical knowledge in complex ways from both societal and individual levels (for instance, public health measures such as closing sewers and individual measures such as eating nutritiously).

Many of the most important findings in development economics have come from broad comparisons between cases. For instance, the **finding** that some countries achieved high health and educational outcomes at low incomes, and that this in turn resulted in much lower population growth and other benefits, came from comparing their

experiences, not from fine-grained statistical tests of household behavior. There is still considerable scope for applying comparative studies of this kind to gain important insights. Indeed, if one needs a sophisticated statistical method to identify an effect, then its relevance may be doubtful.

The reliability critique contests the idea that RCTs provided a sure means—indeed the **gold standard**—for inferring, although in narrow terms, what worked in development. Those who have made the reliability critique, including eminent, statistically minded economists (a few of whom are also Nobel Prize winners), have argued that RCTs suffer from two problems of reliability. The first, external validity, concerns whether the estimate of the effect of a treatment from the place that the RCT is administered, even if it is accurate there, can be transferred elsewhere, given differences in the behaviors of different populations, as well as in the prevailing environmental, institutional, and social circumstances. For instance, public health information may influence behavior more where the government is trusted than where it is not. It may even have the opposite effect from that intended if the government is held in great suspicion. The second concern, internal validity, is about whether the results from a given context are really meaningful and accurate even there. RCTs are designed to measure the average effect of a treatment in a population and cannot generally tell us how it affects different parts of that population (in an extreme case, which is encountered frequently in medical trials, it may harm some people even as it creates a benefit on average). The effect of an intervention may moreover change over time even in a single place due to learning and behavioral responses. For these and other reasons, it is necessary to take care in interpreting what RCTs have actually measured and in employing their lessons, even in the very same place that they have been implemented.

Since the ultimate justification for RCTs is to inform policies on a larger scale, these are very serious problems. If the results of RCTs cannot be generalized, or only partially, then the cost they involve becomes still harder to justify. Moreover, RCT proponents themselves admit their method can be applied only to interventions that affect individuals or local communities and which can then be scaled up—with all of the already noted difficulties that involves.

The adequacy critique notes that RCTs can't inform many, if not most, of the central questions in contemporary global development, especially those that require policies that go beyond widely replicating what works locally. For instance, although an RCT may help inform what causes a migrant to leave her home or what aids her integration elsewhere, it cannot tell us how to organize migration policy between countries. Indeed, although evidence is an important aid to policymaking, it is not sufficient. Designing appropriate policies and prioritizing among them **requires** reference to values to determine what is the appropriate trade-off between different goals and taking into account the political and social consequences of implementing them. For instance, in a post-conflict society, a policy intervention that risks being perceived as prioritizing the interests of one group, whatever an RCT may suggest would be its average impact on the persons benefitted, may be imprudent or even dangerous to implement.

To add to these concerns from within the field of development came another, largely from outside of it: an ethical worry concerning the acceptability of experimenting on people. RCTs involve treating people as means rather than ends (contrary to the spirit of the **famous dictum** of Immanuel Kant) with the idea that the knowledge gained will be of broader public benefit. As noted, it is not at all clear that such knowledge *is* gained, but even if so, the approach of RCTs is to experiment on people rather than to work with them. It is not a surprise that nearly every RCT involves treating poor people, usually also in poor countries, as their subjects (or is that objects?). It is often argued that the participants in trials receive benefits that they would not otherwise receive and therefore can have no complaint, but in an unjust world, the doling out of benefits on a random basis (when, for instance, it is known that some are poorer and more deserving or in more urgent need than others) can be hard to accept and may even lead to potential harms. For instance, a **study** on how economic incentives affect drinking by pedicab drivers in India found that they shifted their consumption from the daytime to the evening. One wonders how this may have affected intra-household violence or other outcomes not considered in the study.

Indeed, it is not clear that the randomistas have adopted the ethical protocols that have long been standard in medical research to ensure meaningful informed consent by individuals, prevent harms to subjects, and ensure that a trial is stopped when there is reason to believe either that it is causing such harms or that there are **evident benefits** that should be offered to all. These ethical concerns, and associated attitudes to the poor as suitable for experimentation, have been barely at all discussed by the randomistas (and do not figure in the discussion of their work by the prize committee). The administration of RCTs has **suffered** from more than a whiff of neocolonial attitudes. Arguably, all of the difficulties of RCTs stem from a single source: a failure to

recognize the full personhood of those who are affected by interventions. Research and policy would be improved by having a less caricatured view of people, whether in deciding when it is alright to experiment on them, what motivates them and what **mistakes** they are likely to make, how it is determined “what works,” or in acting on these conclusions.

If RCTs now “entirely dominate” development economics, or worse, provide the basis for development policymaking, that is no cause for celebration. The roaring success of the randomistas tells us most of all about the historical moment in which they came to prominence: one in which defeatism or cynicism about public initiatives on a larger scale has been replaced by a focus on what works at the level of individuals and communities. But even there, what does work, really, remains an open question. The difficult question of how to fix broken institutions and help societies function better requires going beyond a biomedical metaphor of taking the right pill. Nobel or not, the debate must continue.

Sanjay G. Reddy is an associate professor of economics at the New School for Social Research in New York. Twitter: [@sanjayreddy](https://twitter.com/sanjayreddy)

TAGS: ARGUMENT, DEVELOPMENT, ECONOMICS

[VIEW
COMMENTS](#)