Viewpoint
Rethinking RCTs
By Jana Smith & Sara Flanagan
Rethinking RCTs

Randomized controlled trials have limited value for program implementers without better theories of change and broader sources of data. Behavioral science can help.

BY JANA SMITH & SARA FLANAGAN

Rigorous impact evaluations have become increasingly important to guide the direction and scaling of social impact programs. In 2000, only 39 impact evaluations of work performed in low- and middle-income countries were published, according to the development evidence portal maintained by the International Initiative for Impact Evaluation (3ie). In 2020, 1,526 were published.

The gold standard for such evaluations is the randomized controlled trial (RCT). The use of RCTs in development research has grown significantly in the past two decades, gaining public prominence with the awarding of the 2019 Nobel Prize in Economic Sciences to Esther Duflo, Abhijit Banerjee, and Michael Kremer for their use of field experiments in anti-poverty work. Applying the rigors of traditional laboratory RCTs to test interventions in health care, education, agriculture, and other fields has helped policy makers and NGOs understand what works and does not work in international development.

Implementers of interventions often complain that program monitoring and evaluation need more investment if they are to learn lessons and boost impact. Rigorous impact evaluations cost money—large-scale RCTs can be especially expensive. Complaints aside, organizations and funders often do earmark a significant amount of money for such expenses. However, even large investments of money, time, and expertise often do not yield the results that satisfy what program implementers especially need: valuable and actionable feedback.

Despite the accolades showered on using RCTs in development, many in the field question whether RCTs are appropriate for evaluating complex interventions. They argue that randomization is unfeasible in many cases, that generalizing from RCT findings is difficult, that randomization alone does not mean results are unbiased, and that the design of RCTs sheds little light on why results happen.

Putting such criticisms aside (though many are valid), we think opportunities are available to make the findings of RCTs more actionable for implementers. As economist Angus Deaton and philosopher Nancy Cartwright have written, “[F]or a great many questions where RCTs can help, a great deal of other work—empirical, theoretical, and conceptual—needs to be done to make the results of an RCT serviceable.”

Behavioral science is at the forefront of promising creative paths to better measurement, evaluation, and adaptive learning. In our capacities as behavioral designers at the nonprofit ideas42, we use behavioral science every day to understand how context shapes decision-making to address complex social problems around the globe. We have extensive experience designing interventions with partners in more than 45 countries and have conducted many rigorous evaluations to strengthen these interventions.

We have also conducted evaluations to inform decision-making and improvement of external programs and helped partners to apply the results they received from third-party evaluations in more actionable ways. Drawing from our work, we propose two ways for program designers and funders to get the most from their investments in RCTs so that they produce the answers needed to improve programming and optimize impact.

RETHINKING THEORIES OF CHANGE

First, we recommend rethinking theories of change to aid evaluation design and decision-making. A theory of change often frames an evaluation by providing a coherent narrative for how a program’s activities are expected to generate impact. Formulating a theory of change is a useful exercise for aligning stakeholders and helping them to develop a shared understanding of what results they expect programming to achieve and how. Far too often, however, theories of change contain granular detail of programming approaches and required inputs but fail to articulate the specific mechanisms through which these inputs will lead to intended outcomes.

Fortunately, insights from behavioral science can enrich theories of change and elucidate why and how programs generate results. For example, adoption of a service, product, or process ultimately depends on human behavior. Evidence from behavioral science can help identify when achieving outcomes may require shifting perceptions, beliefs, or norms and whether proposed programming could plausibly contribute to those shifts. Furthermore, many outcomes stated in theories of change (e.g., improved nutrition, higher levels of educational attainment, or more favorable birth outcomes) result from a host of different positive behaviors from a range of stakeholders. Theories of change informed by behavioral science can pinpoint the relevant behaviors of frontline providers, policy makers, clients, managers, and others who may be critical to achieving those outcomes and ensure that they are measured. In this way, behavioral science can help implementers develop a more nuanced understanding of how programs generate impact. In addition,
it can highlight ways to enhance program design before investing what can be many years in data collection for an evaluation.

Theories of change can also benefit from including external mechanisms and evidence-based pathways that may be relevant but that the program designers may not have conceived as within the purview of the program. How can we identify untapped opportunities to have more impact if we focus only on a narrow suite of indicators based on our preexisting notions of how to generate change? To be sure, we do not mean that data-collection efforts should balloon to capture every conceivable pathway through which change can occur. Instead, we envision using formative research—research that is targeted, hypothesis-driven, and qualitative—or already-available evidence-based behavioral models to highlight factors that may have gone unrecognized but that may be critical to driving change. A behaviorally informed evaluation may reveal that the program is having some impact on unexpected pathways or that it is not achieving the expected impact because the untargeted pathways are more critical to the outcome. Both would be valuable insights to inform program decision-making.

For example, we recently evaluated a social-marketing program aimed at preventing smoking among adolescent girls in Ghana. Formative research suggested that girls’ perspectives on friendships; social environments—specifically the social relationships and settings outside of school or work in which they were more likely to be offered a smoke—was a critical determinant of their likelihood to smoke. The program designers did not target the social environment and did not include elements of it in their original theory of change. When we included indicators of social environment in our behaviorally enhanced theory of change, we were able to validate their relevance to girls’ smoking; demonstrate the ways in which prior programming may have already been influencing relevant pathways, such as girls’ perspectives on friendships; and identify promising opportunities for programmers to strengthen their impact via social environments.

**REEXAMINING ASSUMPTIONS**

Our second recommendation is to reexamine assumptions about what data is useful for rigorous evaluations, by better incorporating publicly available data sources and qualitative data.

Practitioners rightly emphasize using both quantitative and qualitative sources of data to monitor program implementation as a complement to more rigorous impact evaluations. But they typically focus too narrowly on the process—how intervention components are being delivered or how they are being received—and fail to delve into how the program and its broader context may be influencing outcomes. Practitioners and researchers often tout the importance of qualitative methods in the context of these evaluations and behavioral studies more broadly. But evaluators often deploy qualitative methods and data only to inform quantitative measures or to bring more color to quantitative findings, rather than as distinct sources of evidence in their own right.

We recognize that qualitative research is critical to informing a theory of change and to developing hypotheses around mechanisms to test with quantitative methods. But qualitative methods can also be used to support this testing, especially in cases when quantitative methods may not be reliable. For instance, qualitative approaches may be necessary to identify and explore contextual nuances that are influencing how the program is working. We must abandon our bias toward quantitative data as the only reliable source of truth in evaluations.

RCT’s rigorous, controlled conditions. At times, however, unexpected trends may arise in the data that cannot be understood with data collected for the evaluation alone. While data sources external to the evaluation itself cannot be used to establish causal impact, they can help inform hypotheses about why certain trends may be observed, especially changes in the macro-context of a study.

Take, for example, our Ghana smoking evaluation. When we observed an increase in smoking rates, we hypothesized that this increase may have been due to a seasonal increase in social activity. We were then able to confirm intensified physical movement during that time period by using aggregated cell-phone mobility data made public by tech companies during the COVID-19 pandemic. In another example from the same evaluation, we observed a decrease over time in the proportion of adolescents believing that most of their peers have tried smoking. We hypothesized that high national inflation during the study period may have influenced this perception: Effective cuts to pocket money may have changed social activity and spending behavior in ways that reduced the visibility of smoking.

To be sure, RCTs offer a powerful approach to clarifying assumptions and ensuring that resources are invested in the most effective programs and policies. In cases where RCTs are the right approach to answer a research question, we can still do more to harness their strengths by designing them to generate more actionable findings. By more effectively integrating behavioral science into formulating more specific theories of change and measuring a wide range of evidence-informed mechanisms, we will answer questions we did not even know to ask. Elevating qualitative data and leveraging publicly available data can add color and richness to those answers and help us more fully realize the potential of evaluations to catalyze the impact of future programming and policies.

JANA SMITH is managing director of global health at ideas42, a nonprofit that uses insights from human behavior—why people do what they do—to help improve lives, build better systems, and drive social change. She is also a current member of the World Health Organization’s Behavioral Insights & Sciences for Health Technical Advisory Group.

SARA FLANAGAN is a principal behavioral designer at ideas42.