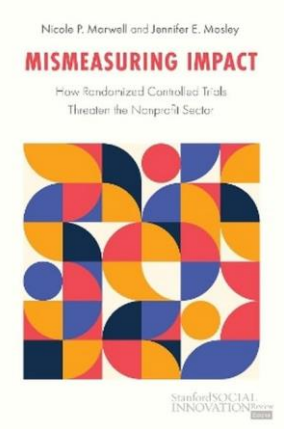
Christian Relief, Development, and Advocacy 7(1), 2026
https://doi.org/10.65538/crda.v7i1.715

BOOK REVIEW

Mismeasuring Impact: How Randomized Controlled Trials Threaten the Nonprofit Sector

by Nicole P. Marwell and Jennifer E. Mosley

Reviewed by Jeffrey R. Bloem and Bruce Wydick



Stanford, CA. Stanford University Press. 2025. $35.00

Those of us who work in research, evaluation, or nonprofit organizational management are familiar with the claim that randomized control trials (RCTs) are the best way for nonprofit organizations to tell us whether a program "works" (i.e. truly helps people). This is a compelling argument to many, even many in our field of economics, in which RCTs have dominated development research over the last two decades, and were the subject of a Nobel Prize in 2019. In *Mismeasuring Impact*, co-authors Nicole P. Marwell and Jennifer E. Mosley contend that this claim is unambiguously false.[1] While they do not dispute the power of well-executed RCTs to identify the causal effects of programs, they provide a multitude of reasons why RCTs are not a good fit for most nongovernmental organizations (NGOs). The arguments the authors make may hold particular sway for Christian development NGOs, who may place a particularly strong emphasis on ensuring access to programming is not allocated randomly, but is rather according to specific missional goals.

Summary

Mismeasuring Impact is a short book with nine chapters. The first is introductory and lays out five problems with RCTs and why they threaten the nonprofit sector. These five problems consist of: (i) the "false certainty" problem, (ii) the "programs need organizations" problem, (iii) the "communities need organizations" problem, (iv) the "rich get richer" problem, and (v) the "agility" problem. The existence of these problems, combined with the power and allure of RCTs as the "gold standard" for program evaluation within nonprofits, sets up an environment that Marwell and Mosley convincingly characterize as threatening to the nonprofit sector.

Providing context representing the basis for the book's central thesis, the second chapter summarizes the rise of RCTs as the "gold standard" method for program evaluation. Marwell and Mosley describe the "gold standard movement" that led to the present reality in which RCTs have achieved an elevated status in terms of the quality of evidence and the worthiness of

[1] For those who have studied the philosophy of science, the idea that RCTs (or any empirical method for that matter) show us what "works" represents a misunderstanding of the scientific process. As Karl Popper notes in *The Logic of Scientific Discovery* (Popper 1959), "[A] theory of induction is superfluous. It has no function in a logic of science… The best we can say of a hypothesis is that up to now it has been able to show its worth, and that it has been more successful than other hypotheses although, in principle, it can never be justified, verified, or even shown to be probable" (315). In other words, a theory can never be proven by empirical science, it can only be falsified. This critical reasoning is foundational to the vocabulary of hypothesis testing but is often overlooked by claims that RCTs or other methods allow us to know what "works."

funding among nonprofits, governments, and philanthropies.

Chapters three through seven dig into each of the five problems with RCTs identified in the introduction. Chapter three discusses the myriad implementation challenges that can inhibit RCTs from credibly estimating program impacts on measurable outcomes, leading to the false certainty problem. Chapter four highlights the realization among many nonprofit organizations that implementing an RCT is typically harder than initially imagined, pulling scarce resources away from other necessary activities conducted by the organization. Chapter five critiques the idea that RCTs help build the credibility and legitimacy of an organization, especially from the perspective of funders. The authors argue that if RCTs indeed do this, they often do so at the expense of the clients and communities the organization serves. Chapter six contends that the managerial logistics of RCTs typically overwhelm the capacity of nonprofit organizations, and are hence carried out poorly. Chapter seven discusses how RCTs restrict nonprofit adaptability while the study is ongoing, fundamentally compromising the responsiveness and agility that make nonprofits an attractive organizational model to address socio-economic challenges in the first place.

Chapter eight provides an alternative to the "RCTs as gold standard" view with three principles for how to think about evaluation differently. The first principle is to fit evaluation methods with specific organizational strategies and community needs. By this, Marwell and Mosley emphasize the idea that there is no best method for evaluation, only an evaluation approach that best fits an organization's needs at a particular time. This idea echoes a call from Mary Kay Gugerty and former USAID Chief Economist Dean Karlan (one of the leading advocates of RCTs himself) in their 2018 book *The Goldilocks Challenge: Right-Sized Evaluation and Monitoring for Social Sector Organizations*, in which they argue that most nonprofit organizations cannot competently implement RCTs. The second principle is to center participant perspectives. Rather than designing evaluation methods around a single question ("does the program work?"), evaluations should prioritize generating information that allows organizations to be more responsive to the needs and perspective of their clients. The third principle is to focus on improvement via an agile and iterative process rather than through standardizing models of program implementation. Together these three principles allow nonprofit organizations to fully adopt an orientation of continuous improvement.

The final chapter summarizes the book by advocating for a move "beyond RCTs." Marwell and Mosley do see a role for RCTs as one possible evaluation approach. But they suggest RCTs are appropriate only under a long and quite restrictive set of conditions, that truthfully few nonprofit organizations are likely to satisfy. *Mismeasuring Impact* will bring sighs of relief from nonprofit leaders who have previously judged RCTs to be necessary for fundraising but logistically unappealing. Although they may understate the necessity for causal inference in program evaluation, Marwell and Mosley rightfully put RCTs in their place as merely one tool in the toolbox of evaluation methods.

Are RCTs Really the 'Gold Standard?'

A significant contribution of *Mismeasuring Impact* is that it points to a fundamental problem in how we talk about RCTs. Because of their elegance in identifying causal effects and their transparency, RCTs have transformed academic development economics, but (unlike much of the work done by economists) RCTs are understandable to the rest of the world. As a result, they have been tagged with the "gold standard" moniker, not because of what they can uniquely do, but often for the transparency in what they do. Too often this is misinterpreted to mean that RCTs are the best way to evaluate any program at any time and in any place. A more accurate understanding of the "gold standard" framing is that RCTs represent an aspirational type of comparison between a treated group and a comparison group that is the same on average in every imaginable way except for program participation. Notice that this latter articulation is more about the conceptual idea of the comparison being made and less about the practical methodological approach for the evaluation.

The 2019 Nobel Prize in Economics was awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer for pioneering this approach to program evaluation in development economics. But RCTs are only one means of causal identification of treatment effects. And they are capable of yielding unbiased estimates of treatment effects only when they can be implemented cleanly and ethically. But it is not always possible for nonprofit organizations to implement clean and ethical RCTs. In fact, Marwell and Mosley actually omit or underdiscuss a great number of other disadvantages presented by RCTs, including a range of externality issues, impact dynamics, reference populations and external validity, and a variety of pernicious experimental effects on behavior, among many other critiques and possible shortcomings.

In such scenarios, different types of quasi-experimental approaches enable comparisons that are able to achieve causal identification without purposeful random assignment. Indeed, there is nothing an RCT can do that a solid quasi-experimental design can't do. The 2021 Nobel Prize in Economics was awarded to David Card, Joshua Angrist, and Guido Imbens for

pioneering quasi-experimental approaches to program evaluation. Modern day practitioners of empirical social science are trained to use both experimental and quasi-experimental tools. Together they both represent the "gold standard," in the sense that the "gold standard" means obtaining unbiased estimates of program effects.

This misunderstanding of the role of RCTs is puzzling. Timothy Ogden's 2017 book *Experimental Conversations: Perspectives on Randomized Trials in Development Economics* features interviews with key players in the RCT movement. These interviews include RCT pioneering researchers Michael Kremer, Abhijit Banerjee, Esther Duflo, Dean Karlan, and Rachel Glennerster along with RCT skeptics such as Angus Deaton and Lant Pritchett. What is notable from this compendium of verbatim conversations is the absence of the idea that RCTs are necessarily the best way for nonprofits to evaluate their programs. Moreover, even the skeptics do not subscribe to the idea that RCTs should never be implemented. Instead, the universally shared view is that RCTs are one available tool in the toolbox of evaluation methods. Some of the methods that Marwell and Mosely describe as alternatives to RCTs, such as writing down theories of change and qualitative interviews, fall short of this gold-standard mark. Ineffective NGOs may be excellent at writing down compelling theories of change. Development initiatives often represent life-and-death interventions. Would a physician use a theory of change, a focus group, or any variety of qualitative method to decide which drug to adopt to fight a patient's cancer? One would hope not. There is a gold standard, and it is unbiased estimation of causal effects. It omits some methods that are often used to evaluate programs, but it is not limited to RCTs.

A New Alternative to RCTs for Nonprofit Organizations

Marwell and Mosley write well about the problems and challenges of RCTs—an experimental evaluation method—for nonprofit organizations, but they underestimate the power of the quasi-experimental methods for addressing most of the pitfalls of RCTs for NGOs while maintaining a rigorous approach for estimating causal effects. We conclude this review with a discussion of the benefits of quasi-experimental evaluation methods and why they overcome many of Marwell's and Mosely's objections to RCTs. Our own view is that embedding quasi-experimental methods maintains the gold standard of causal inference with a practical, ethical, and cost-effective alternative to RCTs.

A unique feature of quasi-experimental evaluation approaches is that they can be seamlessly integrated into the ongoing program operations of a nonprofit via a method we at CEIDS (the Collaborative for Econometrics and Integrated Development Studies) have come to call a "controlled access to treatment" approach. This approach does not require randomly assigning participation in the program while simultaneously randomly withholding participation from those the organization exists to serve. It does not need to distract organizational staff from their core duties, and indeed it makes the collection of counterfactual data a routine part of these duties.

The controlled access to treatment approach is simple and harnesses the quasi-experimental methods developed by those receiving the 2021 Nobel Prize. Due to budget and logistical constraints, most nonprofit organizations have a clear and measurable way to define program eligibility. This eligibility might be based on geography, socio-economic characteristics of households (i.e., income levels, etc.), characteristics of individuals (i.e., age, health outcomes, etc.), or simply first-come-first-served. These eligibility criteria are often directly linked to an organization's mission. This treatment assignment problem is a major issue Marwell and Mosley cite when they convey the degree to which nonprofits cringe at RCT implementation, and it is a problem faced by Christian NGOs who want to make sure access to program is governed by missional values. The controlled access to treatment approach lets organizations do what they exist to do and asks them to do it deliberately. By providing access to a program in a deliberate, controlled, and ethical manner, the organization generates a setting in which evaluators can credibly achieve the gold-standard level of causal inference that is all-too-often thought to materialize only in the experimental setting of an RCT.

To take a few examples, imagine an organization that operates a child sponsorship program. The organization can serve 50 children in a village at a time, but families with a total of 100 children have applied. Given the organization's mission to serve the poor, they decide to use poverty scores as the eligibility criterion, admitting the poorest 50 children into the program. Consider a second example of a reforestation program, for which rollout is planned five years ahead of time across subsequent villages controlling program rollout north to south along a main road. In a third example, a development NGO controls participation in a tutoring program for girls of families with more than three children living within four kilometers of a tutoring center.

In each of these examples, controlling access to the program by poverty status, location, or number of children assuages ethical concerns associated with RCTs. The allocation rule does not distress members of the community because the criteria fit reasonable social customs and norms. It is congruent with a priori need, logistical efficiency, or expectations of impact. And in each example, the program eligibility rules ensure that access is granted in a way that is congruent

with an organization's values. Crucially, by allocating the program in this kind of transparent, controlled way, comparisons of key outcome measures can be made using quasi-experimental methodology. In order of the above examples, these would be regression discontinuity, difference-in-differences, and instrumental variables. This kind of approach enables evaluators to achieve gold standard identification of causal effects of program participation with the same statistical rigor of an RCT yet is consistent with an NGO's capacity and ethics.

The technical foundations for this work are well-established and such approaches are currently implemented within some of the largest for-profit companies in the world such as Microsoft, Google, Amazon, and Meta. As clearly documented by Marwell and Mosley, credible alternatives to RCTs are both currently missing and critical for the future success of the nonprofit sector. We wholeheartedly agree.

References

Gugerty, Mary Kay, and Dean Karlan. 2018. *The Goldilocks Challenge: Right-Fit Evidence for the Social Sector*. New York: Oxford University Press.

Ogden, Timothy N. 2017. *Experimental Conversations: Perspectives on Randomized Trials in Development Economics*. Boston: MIT University Press.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. New York: Basic Books.

Jeffrey R. Bloem is a Research Fellow at the International Food Policy Research Institute (IFPRI) and Research Affiliate within the Ford Program at the University of Notre Dame as a member of the Collaborative for Econometrics and Integrated Development Studies (CEIDS). He holds an MS in Agricultural, Food, and Resource Economics from Michigan State University, and a PhD in applied economics from the University of Minnesota.

Author email: bloem.jeff@gmail.com

Bruce Wydick is Professor of Economics at the University of San Francisco where he specializes in experimental and quasi-experimental tools to study the impact of development programs. He is also affiliated with the Collaborative for Econometrics and Integrated Development Studies (CEIDS) at Notre Dame. In addition to recent work on microenterprise and child sponsorship, he has studied the nature of charitable giving and the role of hope and aspirations in escaping poverty traps. His most recent book is *Shrewd Samaritan* (Thomas Nelson - HarperCollins).

Author email: wydick@usfca.edu