CENTER FOR GLOBAL DEVELOPMENT

Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science

David K. Evans

Abstract

Randomised controlled trials in development economics, political science, and other social science fields have been on the rise in recent decades. Recent awards and trials in development economics have re-ignited active discussions of the ethics of these trials. This article surveys common ethical concerns with such trials and proposes a series of practical suggestions to help researchers and policymakers be more mindful of and transparent about ethics as they consider, design, implement, and report randomised controlled trials and other impact evaluations in development settings.

This paper was first published in February 2021. The original version is available at https://www.cgdev.org/sites/default/files/WP565-Evans-Ethical-issues-and-RCTs.pdf

KEYWORDS
randomised controlled trials, ethics, evaluation

JEL CODES
D04, H43

Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science

David K. Evans
Center for Global Development

This paper benefited from helpful comments from Ṣẹ̀yẹ Abímbọ́lá, Michael Clemens, Lee Crawfurd, Janeen Madan Keller, Carleigh Krubiner, David McKenzie, and Justin Sandefur, and from excellent research assistance by Amina Mendez Acosta. The Bill & Melinda Gates Foundation provided generous funding. All views are my own.

David K. Evans. 2022. "Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science." CGD Working Paper 565. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/towards-improved-and-more-transparent-ethics-randomised-controlled-trials-development

CENTER FOR GLOBAL DEVELOPMENT
2055 L Street, NW Fifth Floor
Washington, DC 20036

1 Abbey Gardens
Great College Street
London
SW1P 3SE

www.cgdev.org

Contents

1. Motivation **1**

2. Methods and definitions **2**

3. Suggestions for RCTs at three stages **3**

3.1. Suggestions for planning potential RCTs 3

3.2 Suggestions for conducting RCTs 8

3.3. Suggestions for writing up RCTs 9

4. Conclusion **10**

Works cited **11**

1. Motivation

Randomised controlled trials (RCTs) have been increasing in international development research in recent decades (Cameron, Mishra, and Brown 2016; McKenzie 2016; Sabet and Brown 2018). Against that backdrop, researchers, practitioners, and citizens continue to discuss the ethics and the efficacy of RCTs in development, including development economics (Green 2020). This discussion is not new (Angell 1997; Heckman 1991), and many of the critiques of RCTs are not unique to RCTs, rather applying to prospective field research more broadly (Evans 2021).[1] These debates are also not unique to economics or even to the social sciences (Concato, Shah, and Horwitz 2000; Farmer 2013). But the lack of novelty or uniqueness does not exempt implementers of social science RCTs from facing these ethical issues.

These debates can (and have) engendered much thought (Ogden 2017; Bédécarrats, Guérin, and Roubaud 2020; Rodgers et al. 2020). Defenders of RCTs argue for methodological advantages and critics challenge them (Deaton and Cartwright 2018; Banerjee and Duflo 2009). Critics highlight the ethical problems with excluding individuals from potentially beneficial treatments; defenders highlight that initial coverage for many social programs is incomplete anyway, and that random assignment of beneficiaries may be fairer than assignment through other, existing mechanisms (Mulligan 2014; Goldberg 2014; Duflo, Glennerster, and Kremer 2006). Critics propose that RCTs distract development economics from big, transformative questions; defenders argue that RCTs are still a minority of development research and that many important policy questions can be influenced by RCTs (McKenzie 2016; Chelwa 2020; McKenzie 2020). Critics highlight that informed consent (for RCT participation, not just survey participation) can be problematic with cluster randomised trials; ethicists suggest that it may not be necessary in the case of certain services (Hoffmann 2020; MacKay and Chakrabarti 2019). Critics highlight that some social science RCTs are used to test interventions that are already known to be effective; defenders propose that if an intervention can be successfully implemented in a given context or if it represents the best use of resources in that context, the intervention may still merit testing with an RCT (McKenzie 2013).

The focus of this piece is on practical ways to improve the ethics—and the discussion of ethics—of randomised controlled trials in the social sciences. Again, many of the issues discussed are not unique to RCTs, and so many of the proposed practices would likewise be relevant to other research involving data collection.

1 For a broader look at ethics in international development research, see van den Berg et al. (2021)—or a summary in Sturdy (2022).

2. Methods and definitions

This is not a systematic review, but rather a discussion of ethical issues that regularly come up in the planning, implementation, and write-up of RCTs, together with suggestions to address these issues, both based on my own experience implementing RCTs in low- and middle-income countries and on the literature. Because the objective is to provide practical, actionable suggestions for those who actively implement RCTs rather than an overview of all possible issues and all possible solutions, I used the snowball technique—identifying key articles known to me on the topic and then looking at relevant articles both cited in those key articles and that cite those articles (Sayers 2007). Three basic ethical principles for consideration in research that involves people as subjects, which includes social science RCTs, are respect for persons, beneficence (i.e., researchers should not only "do no harm" but should "maximize possible benefits and minimize possible harms"), and justice (i.e., researchers should not deny a benefit without "good reason") (The National Commission for the Protection of Human Subjects of and Biomedical and Behavioral Research 1979). The issues and suggestions that I raise in this article all relate to these three principles, as encountered in social science RCTs.

The suggestions I provide in this note are not the only possible solutions to ethical challenges; other articles also offer practical advice (Glennerster and Powers 2016; Cronin-Furman and Lake 2018; Asiedu et al. 2021). I hope that readers will consider not only the suggestions offered here, but also read the deeper discussions of challenges and solutions in the research cited. For a comprehensive list of resources on the ethics of field experiments, see the bibliography prepared by Cohn and MacKay (2022). There are also resources on ethics of RCTs in low- and middle-income countries in the context of health research, much of which may be relevant to social science RCTs (e.g., Emanuel et al. (2004); Hyder et al. (2014)).

In this review, a randomised controlled trial refers to the evaluation of a policy or program in which beneficiaries of the policy or program are selected randomly and then compared to a set of non-beneficiaries who are also selected randomly. In some cases, randomization assigns participation to individuals and in other cases, randomization assigns to participation to schools or communities as a whole (e.g., a scholarship is offered to all girls in a school or a workfare program is offered to all poor adults in a community); the latter are called cluster randomised trials. RCTs are a specific type of impact evaluation. Impact evaluations, as defined by Gertler et al. (2016), "are a particular type of evaluation that seeks to answer a specific cause-and-effect question: What is the impact (or causal effect) of a program [or policy] on an outcome of interest? This basic question incorporates an important causal dimension. The focus is only on the impact: that is, the changes directly attributable to a program, program modality, or design innovation." Impact evaluations include RCTs but also include other evaluation methods, such as regression discontinuity or difference-in-differences designs.

3. Suggestions for RCTs at three stages

3.1. Suggestions for planning potential RCTs

Ask if an RCT is the best way to learn what you want to learn

If you're a practitioner, monitoring your activities and evaluating their impact is important. But not every operation is conducive to an RCT (or even an impact evaluation). Quasi-experimental methods can be effective, although practitioners should be aware that some methods (including a badly implemented RCT) can yield less useful information than no impact measure at all (Evans and Wydick 2016). As Gugerty and Karlan write, "Despite the demonstrated value of high-quality impact evaluations, a great deal of money and time has been wasted on poorly designed, poorly implemented, and poorly conceived impact evaluations" (Gugerty and Karlan 2018). This does not mean fewer program evaluations should take place: on the contrary, many taxpayer dollars support programs with little empirical evidence of their benefit. An RCT is one of the available tools to provide that evidence.

If you are a researcher, draw on the full range of theoretical and empirical methods (Glennerster 2018). If a quasi-experimental method allows a better program allocation mechanism and also delivers high-quality estimates of impact, then it may be a strong alternative to an RCT. However, quasi-experimental methods do not always deliver high-quality estimates of impact, depending on whether the required assumptions are satisfied. RCTs require minimal assumptions to estimate average treatment effects (Deaton and Cartwright 2018).

Contrary to anecdotal narratives in economics, RCTs are not the only way to publish a development paper: in 2015, less than one-third of development papers in the top five economics journals were RCTs (McKenzie 2016). Even the profession's most high-profile proponents of RCTs identify impacts with quasi-experimental or descriptive analysis (Banerjee, Duflo, and Qian 2020).

A related question is whether what you want to learn is worth the resources spent on the RCT (or other evaluation). Resources spent on implementing an RCT may mean fewer resources are available for the intervention itself. This trade-off exists in all contexts, but it may be felt most acutely in fragile or humanitarian contexts, where urgent needs are great. At the same time, while it is important for researchers to question the value of the potential knowledge gain from the RCT, policymakers must recognize failing to evaluate the effectiveness of programs also presents ethical problems. Spending recurring resources on programs that are not proven to improve beneficiary welfare may mean that those resources are being wasted. Even if the program does improve beneficiary welfare, there may be alternative interventions which boost welfare more. For example, an RCT in Rwanda demonstrated that a workforce training program boosted various indicators of well-being for beneficiaries relative to a control group. However, the same RCT showed that

simply giving cash transfers equal to the value of the workforce training improved a wider range of indicators (McIntosh and Zeitlin 2022). Additionally, even if a program has proven beneficial in one setting, it may not deliver the same results in another setting. A vocational and life skills training program in improved life outcomes for adolescent girls in Uganda (Bandiera et al. 2020), but when the program was replicated in Tanzania, the benefits failed to materialize (Buehren et al. 2017). Failing to evaluate (with an RCT or in another rigorous way) requires ethical justification, just as implementing an RCT does.

Ask if an RCT is justified in this situation

Two conditions which may justify a policy RCT are policy equipoise—i.e., there is genuine doubt as to the merits of an intervention—and "just allocation of a scarce good" (MacKay 2020). As you consider an RCT, ask if there is genuine doubt about the merits of the intervention. First, in some cases, interventions that intuitively seem obviously beneficial have resulted in no average impacts, as with textbook distribution in Kenya and Sierra Leone, thus whether it is possible to implement a given program effectively in a given context may also be in question (Glewwe, Kremer, and Moulin 2007; Sabarwal, Evans, and Marshak 2014). Second, economists have made the case that there is value in understanding the size of an impact locally, to understand the best way to use local resources (McKenzie 2013). Third, there may be no doubt that an individual intervention has some positive impact, but there may be doubt that a collection of interventions has a broader impact. For example, the African Millennium Villages Project included a bundle of interventions like malaria control and improved access to safe water, which individually have demonstrated impacts on human well-being. However, there was no evidence that together, these interventions would dramatically reduce poverty—the objective of the intervention (Sanchez et al. 2007). Fourth, there may be evidence that an intervention works in a small, closely controlled trial but not at a large scale implemented through government systems. Increasingly, researchers are proving capable of implementing RCTs of large-scale, government implemented programs (Muralidharan and Niehaus 2017). In all these cases, an RCT may be justified even if there is some existing evidence on the efficacy of the intervention.

If the benefits of an intervention are well established in other contexts (e.g., distributing cash), then ask: are there more potential beneficiaries with equally good claims to the program than could be covered by the program? In those cases, an RCT could be the fairest way to distribute benefits (as well as learn whether the program works locally). In other cases, some potential beneficiaries may have a stronger claim than others and the knowledge to be gained from an RCT may not outweigh the benefits of an alternative allocation mechanism. In both cases, weigh the knowledge to be gained against any risks to participants, and err on the side of caution. As Cronin-Furman and Lake (2018) put it, "The social science community at large is obligated to relentlessly question whether the scientific contribution of the final product genuinely warranted sensitive firsthand research."

Take potential risks to participants and implementers seriously

Researchers conducting any RCT should carefully consider and document potential risks to participants. This applies beyond RCTs, to any research that involves interviewing or interacting with subjects, but RCTs require heightened scrutiny because they manipulate treatment (and the allocation thereof) in addition to eliciting information. For example, randomly assigning treatment runs the risk of excluding people who would likely benefit from a program or including people who have little need of it. Collecting data—whether as part of an RCT or another study—during a pandemic or a violent conflict could expose either enumerators or respondents to risk of illness or injury. Programs that encourage beneficiaries to take action—whether that is to invest in a new product or take a new job—entail the participants incurring risks (such as losing resources from a poor investment), and RCTs sometimes generate these new product or job opportunities.

Based on the risks, are there mitigation efforts your RCT can take? In Côte d'Ivoire, implementers were concerned with risks of a political education campaign during a contentious election year, so the researchers shifted the timing of the intervention (Davis 2020). In Rwanda, teachers were randomly assigned to two different types of contracts, so that some received contracts different from those they had initially applied for. However, a signing bonus was incorporated so that no recruit would be worse off as a result of the randomization (Leaver et al. 2021). Most importantly, is the expected benefit of the results worth the risks? Interventions in humanitarian settings, often targeting extremely vulnerable beneficiaries, are a salient case of potential risks of excluding participants. One design to learn from such interventions while mitigating such concerns is to omit the control group and rather test two alternative treatments; a second is to provide benefits to all of the most vulnerable and then only randomise treatment among less vulnerable households (Quattrochi et al. 2020).

A related point arises with any survey research: if RCTs include surveys that ask about sensitive topics (such as domestic violence), how will participants be protected and supported (Alderman, Das, and Rao 2016)? In medicine, the principle of essentiality is sometimes invoked: "The research being carried out should be essential for the advancement of knowledge that benefits patients, doctors and all others" (Sanmukhani and Tripathi 2011). While essentiality is not universally accepted as a condition for RCTs and views on what constitute essential research will vary, researchers can use this principle as a benchmark to weigh against any risks imposed (McKenzie 2016).

Engage effectively with local scholars and local populations

Local authors often have greater knowledge and understanding of institutions, as well as more of a vested interest in the well-being of the country. This is neither an encouragement of tokenism (i.e., including local authors in name only) nor a pretense that a social scientist in a given country has much in common with participants in a study on extreme poverty (Abimbola 2019). As Deaton writes, "Even in the US, nearly all RCTs on the welfare system are RCTs done by better-heeled,

better-educated and paler people on lower income, less-educated and darker people" (Deaton 2020). This critique is not unique to RCTs: by one recent count, economics papers about Africa were "78 times more likely to be written by authors without an institutional affiliation in Africa" (Panin 2020). But neither are RCTs exempt from the critique. Engaging local authors is one way to increase engagement with the research community in the country and improve the quality of the work.

For international researchers, involving local authors in a meaningful way can increase engagement with the research community in the country and improve the quality of the work. International researchers can work with local authors in several ways. First, they can search the literature—including international journals, local journals, and unpublished literature—to identify local authors who have relevant topical expertise, even if those local authors have not had access to funding to implement RCTs in the past. Second, they can begin consulting local researchers early in the process, at the stage of designing research questions and programs. Collaborating at the idea generation stage, both with local researchers and policymakers, can catalyse long-term partnerships and result in research questions with clear benefits for local communities (Herman et al. 2022). Third, they can encourage local researchers to lead dialogue with policymakers over the course of the research project, including the dissemination of results.

In years past, one concern with engaging local scholars as authors could be a fear of adding co-authors and thus diluting credit for the work, but economics has seen a study increase in the number of authors over time, with 80 percent of papers having multiple authors as of 2011 (Hamermesh 2013). Kuld and O'Hagan put it this way: "If present trends continue, the number of papers with four or more authors could soon exceed the number of solo-authored papers" (Kuld and O'Hagan 2017). Public health has a long history of many-authored papers.

Seek approval from institutional review boards (IRBs), including local boards

IRBs have the function of ensuring that research that involves "human subjects" (i.e., people) is carried out ethically. In high-income countries, most universities and many research institutes have an IRB. In many low- and middle-income countries, there is a local social science review board. Seeking approval for research locally—and if relevant, internationally—can be an important safeguard and can help researchers think through risks to research participants.

However, clearance from an IRB is not a substitute for engaging directly with ethical issues in RCTs and other research (O'Flynn, Barnett, and Camfield 2016). As Tony Watima wrote, "Being compliant does not mean it's ethical" (Watima 2020). Some IRBs in high-income countries may have little familiarity with conditions in low-income environments, and IRBs and social science review boards in both high and low income environments are heterogeneous in quality across countries and institutions (Cronin-Furman and Lake 2018). But engaging with IRBs still provides an important layer of oversight for research that studies people and their actions.

Take action from the outset to ensure that the RCT delivers accurate, replicable, transparent results

In addition to their interest in improving policy, researchers who conduct RCTs and other research have career incentives to publish their research, and it is well documented that statistically significant results are easier to publish—although this problem is decreasing in economics over time (Brodeur, Cook, and Heyes 2020). These incentives can lead researchers to test impacts on a large number of outcomes but then report only those outcomes that show a statistically significant impact, a practice known as p-hacking (a reference to the p-values used to measure statistical significance). This practice presents an ethical problem in that policymakers may adopt or expand programs based on spurious results. RCTs show less evidence of p-hacking than other empirical methods (Vivalt 2019), but researchers can still take action at the design stage to reduce the temptation to p-hack later. This can include registering trials in advance so that even if the results are not published, the fact that a trial was run on the topic is known. The American Economic Association has a free registry for RCTs (American Economic Association 2022). It can also include preparing a pre-analysis plan (Olken 2015) and making sure that appropriately anonymized data are publicly available—e.g., in the 3ie Dataverse or another data repository (3ie 2022)—after publication so that other researchers can re-analyse the data to see if the results are replicable.

Establish a plan for use of the RCT results from the beginning

Because RCTs require significant resources, investing in knowledge creation through RCTs but then not using those results to seek to inform policy would be a waste. While intuition might suggest that this would principally take place after the RCT results come in, a plan to influence policy from the design stage can boost the likelihood of policy influence. This can involve designing RCTs that begin with policy questions and the space in which policymakers have to act (Kaufman et al. 2022), establishing policy objectives (which will be dependent on the outcomes of the research), identifying stakeholders and key influencers in the policy decision process, and establishing a plan for engaging with those stakeholders throughout the RCT process (3ie and ODI 2016). An expert panel on policy influence highlighted that success in RCT and other research use in policy comes in part from holding "lots of meetings with the policy makers at each stage, especially the beginning, including a range of actors within the organization, so that everyone is aware of the project, everyone knows what value the researchers are offering that the organization doesn't already have, and everyone knows when results will be available" (Evans 2015). Researchers can also boost the influence of their research beyond the immediate context of the RCT (e.g., for policymakers in other countries seeking solutions to their own problems) by publishing the research itself and then using other outlets—such as blogs and other social media—to disseminate non-technical summaries (Evans 2018; McKenzie and Özler 2014).

Some RCTs seek primarily to test a theory of human behaviour and so do not have any intention to immediately influence policy. The policy influence of these RCTs would come indirectly, as a better

understanding of human behaviour may result in the design of better policies. In these cases, authors should still responsibly disseminate their findings so that those who would incorporate them into policy can do so correctly.

3.2 Suggestions for conducting RCTs

Ensure informed consent in data collection and engage the appropriate level of informed consent for the RCT

Almost all surveys incorporate the concept of informed consent, advising survey respondents that they are not obligated to respond to either the entire survey or to individual questions. Informed consent can be trickier with cluster randomised trials (Hoffmann 2020). For example, many education RCTs randomise at the community level, and while community leaders may be aware they are participating in a trial, individual community members may not. In some health trials, researchers have not sought consent from the control group if that might lead to observation and potential imitation of the treatment group (Glennerster and Powers 2016; Lignou 2018).
In government policy RCTs, informed consent may not be essential if two conditions are fulfilled. First, the government must have a "right to rule" in the sphere covered by the trial: a trial in a public school to which parents have willingly sent their children may meet this condition. (Obviously this cannot mean that governments can ethically do anything they please. This condition rests on a basic assumption of a broadly just government.) Second, data collection cannot infringe on individual rights. So data collection, particularly involving private information, would still require informed consent (MacKay and Chakrabarti 2019). When researchers are working to implement RCTs with non-government organizations, the "right to rule" is not present. Even in this case, some form of informed consent is often possible (McRae et al. 2011). This question should be reviewed by IRB and weighed against the risks to participants.

Consider compensating subjects in some way, including those in the control group

Many surveys compensate respondents for their time in some way, often with an in-kind payment of some sort. There is debate about the ethics of compensating research participants, as highly vulnerable participants may be in such need that the payment acts as a disproportionately powerful incentive to participate in an RCT (or any data collection exercise). Still, some research teams adopt a principle of "no survey without service" (Osrin et al. 2009). In practice, that has taken the form of various kinds of training workshops or health information campaigns (on topics not related to the area of study) for control group participants in examples in Bangladesh, Côte d'Ivoire, India, Malawi, and Nepal (Davis 2020).

3.3. Suggestions for writing up RCTs

Include an explicit discussion of ethical issues

Empirical economics and some other social sciences do not have a norm of explicitly referencing ethical issues or even human subjects review in their publications (Alderman, Das, and Rao 2016). Asiedu et al. (2021) recently proposed a full set of questions that could be addressed in the ethics appendix of a social science paper.

To give two examples, in an RCT in Malawi examining the impact of providing "information about the true risk of HIV infection, which is much lower than people's ex ante beliefs," a five-page appendix section discusses a wide array of ethical issues and why the author believes that they are addressed in the study (Kerwin 2018). In a study in the Philippines, which evaluated the randomly assigned removal of a religious values component from an ongoing skills training program, a shorter section is dedicated to discussion of ethical considerations (Bryan, Choi, and Karlan 2020). If you believe that your study poses no risks to participants and presents no ethical issues, then state that explicitly. This section can also explicitly reference the institutional review boards that cleared the project, which is common practice in medical journals.

Credit a wide range of contributions

Medical and public health journals often request that articles with multiple authors lay out the distinct contributions of each author. This is not the custom in social science studies, but it highlights the value of recognizing an array of contributions to the work. One practical way to implement this is providing an author statement using the Contributor Roles Taxonomy, or CRediT (Allen, O'Connell, and Kiermer 2019). As social science RCTs write up their results, they can incorporate recognition of a range of essential roles in the project. Researchers can make sure they credit collaborators and subjects in ways that the latter feel comfortable with (Cronin-Furman and Lake 2018). One coordinated program of RCTs on the topic of cash transfers (The Transfer Project) recommends including "on behalf of the Programme X Evaluation Team" in addition to the names of the principal paper authors, with the evaluation team members elaborated in the paper (The Transfer Project 2018).

4. Conclusion

None of these suggestions are intended to dissuade the research community from actively engaging in careful evaluation of program and policy impacts, nor from using RCTs. Many policies and programs that consume resources go largely unevaluated (or are evaluated badly). Instead, researchers should do rigorous research in an ethical way.

Some may be skeptical of the voluntary nature of these proposals. But a recent example in economics gives hope. Not many years ago, the practice of pre-registering RCTs in economics or preparing pre-analysis plans was virtually unheard of. Now, referees at top economics journals commonly inquire whether reports of RCTs follow their pre-analysis plans. Cultural changes in how RCTs are implemented are also possible.

None of these actions (or perhaps, even all of them taken together) will form an impermeable shield against ethical criticism for a given study. There will always be debate about the ethics of RCTs in general and of particular RCTs. As Kwame Owino, CEO of Kenya's Institute for Economic Affairs, has said, "Sometimes people study difficult things" (Green 2020), and studying difficult topics necessarily means engaging with thorny ethical issues.

However, I propose that—drawing on long experience in medicine and public health and more recent experience in the social sciences—it is possible to improve both on the ethics of RCTs and on transparency around those ethics, permitting more productive debate among both researchers and the potential beneficiaries of the research.

Works cited

3ie. 2022. "Transparent, Reproducible, and Ethical Evidence (TREE) Policy (Version 3)." https://www.3ieimpact.org/sites/default/files/2022-04/3ie-transparent-reproducible-ethical-evidence-policy-2022.pdf.

3ie, and ODI. 2016. "Policy Impact Toolkit—Tutorial Introduction." May 13, 2016. https://web.archive.org/web/20160513224527/http://policyimpacttoolkit.squarespace.com/pip/.

Abimbola, Seye. 2019. "The Foreign Gaze: Authorship in Academic Global Health." *BMJ Global Health* 4 (5): e002068. https://doi.org/10.1136/bmjgh-2019-002068.

Alderman, Harold, Jishnu Das, and Vijayendra Rao. 2016. "Conducting Ethical Economic Research: Complications from the Field." In *The Oxford Handbook of Professional Economic Ethics*, edited by George F. De Martino and Deirdre McCloskey. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199766635.001.0001/oxfordhb-9780199766635-e-018.

Allen, Liz, Alison O'Connell, and Veronique Kiermer. 2019. "How Can We Ensure Visibility and Diversity in Research Contributions? How the Contributor Role Taxonomy (CRediT) Is Helping the Shift from Authorship to Contributorship." *Learned Publishing* 32 (January): 71–74. https://doi.org/10.1002/leap.1210.

American Economic Association. 2022. "AER RCT Registry: The American Economic Association's Registry for Randomized Controlled Trials." 2022. https://www.socialscienceregistry.org/site/about.

Angell, Marcia. 1997. "The Ethics of Clinical Research in the Third World." *New England Journal of Medicine* 337 (12): 847–49. https://doi.org/10.1056/NEJM199709183371209.

Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry. 2021. "A Call for Structured Ethics Appendices in Social Science Papers." *Proceedings of the National Academy of Sciences* 118 (29): e2024570118. https://doi.org/10.1073/pnas.2024570118.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2020. "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa." *American Economic Journal: Applied Economics* 12 (1): 210–59. https://doi.org/10.1257/app.20170416.

Banerjee, Abhijit, and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1 (1): 151–78. https://doi.org/10.1146/annurev.economics.050708.143235.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2020. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." *Journal of Development Economics* 145 (June): 102442. https://doi.org/10.1016/j.jdeveco.2020.102442.

Bédécarrats, Florent, Isabelle Guérin, and François Roubaud. 2020. *Randomized Control Trials in the Field of Development: A Critical Perspective*. Oxford, New York: Oxford University Press.

Berg, Rob D. van den, Penny Hawkins, and Nicoletta Stame. 2021. *Ethics for Evaluation: Beyond "Doing No Harm" to "Tackling Bad" and "Doing Good."* New York: Routledge. https://doi.org/10.4324/9781003247234.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* 110 (11): 3634–60. https://doi.org/10.1257/aer.20190687.

Bryan, Gharad, James J. Choi, and Dean Karlan. 2020. "Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes." *The Quarterly Journal of Economics*, June. https://doi.org/10.1093/qje/qjaa023.

Buehren, Niklas, Markus P. Goldstein, Selim Gulesci, Munshi Sulaiman, and Venus Yam. 2017. "Evaluation of an Adolescent Development Program for Girls in Tanzania." WPS7961. The World Bank. http://documents.worldbank.org/curated/en/245071486474542369/Evaluation-of-an-adolescent-development-program-for-girls-in-Tanzania.

Cameron, Drew B., Anjini Mishra, and Annette N. Brown. 2016. "The Growth of Impact Evaluation for International Development: How Much Have We Learned?" *Journal of Development Effectiveness* 8 (1): 1–21. https://doi.org/10.1080/19439342.2015.1034156.

Chelwa, Grieve. 2020. "Pop Developmentalism in Africa." *CODESRIA Bulletin* 1. https://mcusercontent.com/cb1b4baf30d12f021ac20ecde/files/7ee66315-76d1-4b26-a377-1498f7b6d553/COD_Bul_No._1_2020_R.pdf.

Cohn, Emma, and Douglas MacKay. 2022. "Ethics of Field Experiments: A Bibliography." https://dmackay.web.unc.edu/ethics-of-field-experiments-a-bibliography/.

Concato, John, Nirav Shah, and Ralph I. Horwitz. 2000. "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *New England Journal of Medicine* 342 (25): 1887–92. https://doi.org/10.1056/NEJM200006223422507.

Cronin-Furman, Kate, and Milli Lake. 2018. "Ethics Abroad: Fieldwork in Fragile and Violent Contexts." *PS: Political Science & Politics* 51 (3): 607–14. https://doi.org/10.1017/S1049096518000379.

Davis, Justine M. 2020. "Manipulating Africa? Perspectives on the Experimental Method in the Study of African Politics." *African Affairs* 119 (476): 452–67. https://doi.org/10.1093/afraf/adaa009.

Deaton, Angus. 2020. "Randomization in the Tropics Revisited: A Theme and Eleven Variations." Working Paper 27600. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w27600.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine*, Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue, 210 (August): 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2006. "Using Randomization in Development Economics Research: A Toolkit." Working Paper 333. Technical Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/t0333.

Emanuel, Ezekiel J., David Wendler, Jack Killen, and Christine Grady. 2004. "What Makes Clinical Research in Developing Countries Ethical? The Benchmarks of Ethical Research." *The Journal of Infectious Diseases* 189 (5): 930–37. https://doi.org/10.1086/381709.

Evans, David K. 2015. "Bridge the Gap between Research and Policy, One Panel Discussion (and 145 Studies) at a Time." *World Bank Development Impact Blogs* (blog). 2015. https://blogs.worldbank.org/impactevaluations/bridge-gap-between-research-and-policy-one-panel-discussion-and-145-studies-time.

——. 2018. "Make Your Research Known – 10 Tools to Increase Consumption of Your Research." *World Bank Development Impact Blogs* (blog). 2018. https://blogs.worldbank.org/impactevaluations/make-your-research-known-10-tools-increase-consumption-your-research.

Evans, David K. 2021. "Florent Bédécarrats, Isabelle Guérin, François Roubaud (Eds.) Randomized Control Trials in the Field of Development: A Critical Perspective Oxford University Press, 2020, 448 p., $100.00." *Population and Development Review* 47 (2): 551–54. https://doi.org/10.1111/padr.12410.

Evans, David K, and Bruce Wydick. 2016. "Is My NGO Having a Positive Impact?" *World Bank Development Impact Blogs* (blog). February 2016. https://blogs.worldbank.org/impactevaluations/my-ngo-having-positive-impact.

Farmer, Paul. 2013. "Clinical Trials and Global Health Equity." *The Lance Global Health Blog* (blog). July 8, 2013. https://marlin-prod.literatumonline.com/pb-assets/Lancet/langlo/TLGH_Blogs_2013-2018.pdf.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. *Impact Evaluation in Practice, Second Edition*. World Bank Publications.

Glennerster, Rachel. 2018. "Rachel Glennerster on Twitter: 'Excited to Talk about the Why & How of Evidence in Policy…'" Twitter. March 17, 2018. https://twitter.com/rglenner/status/974962256445345792.

Glennerster, Rachel, and Shawn Powers. 2016. "Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations." In *The Oxford Handbook of Professional Economic Ethics*, edited by George F. De Martino and Deirdre McCloskey. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199766635.001.0001/oxfordhb-9780199766635-e-017.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2007. "Many Children Left Behind? Textbooks and Test Scores in Kenya." Working Paper 13300. National Bureau of Economic Research. https://doi.org/10.3386/w13300.

Goldberg, Jessica. 2014. "The R-Word Is Not Dirty." *Center For Global Development* (blog). March 5, 2014. https://www.cgdev.org/blog/r-word-not-dirty.

Green, Andrew. 2020. "A New Research Experiment in Kenya Raises Questions about Ethics." *Devex*, September 8, 2020. https://www.devex.com/news/sponsored/a-new-research-experiment-in-kenya-raises-questions-about-ethics-98039.

Gugerty, Mary Kay, and Dean Karlan. 2018. "Ten Reasons Not to Measure Impact—and What to Do Instead." *Standford Social Innovation Review* (blog). 2018. https://ssir.org/articles/entry/ten_reasons_not_to_measure_impact_and_what_to_do_instead.

Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51 (1): 162–72. https://doi.org/10.1257/jel.51.1.162.

Heckman, James J. 1991. "Randomization and Social Policy Evaluation Revisited." Working Paper 107. Technical Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/t0107.

Herman, Biz, Amma Panin, Elizabeth Iams Wellman, Graeme Blair, Lindsey D. Pruett, Ken Ochieng' Opalo, Hannah M. Alarian, et al. 2022. "Field Experiments in the Global South: Assessing Risks, Localizing Benefits, and Addressing Positionality." *PS: Political Science & Politics*, May, 1–4. https://doi.org/10.1017/S1049096522000063.

Hoffmann, Nimi. 2020. "Involuntary Experiments in Former Colonies: The Case for a Moratorium." *World Development* 127 (March): 104805. https://doi.org/10.1016/j.worlddev.2019.104805.

Hyder, Adnan A., Abbas Rattani, Carleigh Krubiner, Abdulgafoor M. Bachani, and Nhan T. Tran. 2014. "Ethical Review of Health Systems Research in Low- and Middle-Income Countries: A Conceptual Exploration." *The American Journal of Bioethics* 14 (2): 28–37. https://doi.org/10.1080/15265161.2013.868950.

Kaufman, Julia, Amanda Glassman, Ruth Levine, and Janine Madan Keller. 2022. "Breakthrough to Policy Use: Reinvigorating Impact Evaluation for Global Development (Brief)." https://www.cgdev.org/publication/breakthrough-policy-use-reinvigorating-impact-evaluation-global-development-brief.

Kerwin, Jason. 2018. "Scared Straight or Scared to Death? The Effect of Risk Beliefs on Risky Behaviors." SSRN Scholarly Paper ID 2797493. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.2797493.

Kuld, Lukas, and John O'Hagan. 2017. "The Trend of Increasing Co-Authorship in Economics: New Evidence." *VoxEU.Org* (blog). December 16, 2017. https://voxeu.org/article/growth-multi-authored-journal-articles-economics.

Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin. 2021. "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools." *American Economic Review* 111 (7): 2213–46. https://doi.org/10.1257/aer.20191972.

Lignou, Sapfo. 2018. "Informed Consent in Cluster Randomised Trials: New and Common Ethical Challenges." *Journal of Medical Ethics* 44 (2): 114–20. https://doi.org/10.1136/medethics-2017-104249.

MacKay, Douglas. 2020. "Government Policy Experiments and the Ethics of Randomization." *Philosophy & Public Affairs*, August, e12174. https://doi.org/10.1111/papa.12174.

MacKay, Douglas, and Averi Chakrabarti. 2019. "Government Policy Experiments and Informed Consent." *Public Health Ethics* 12 (2): 188–201. https://doi.org/10.1093/phe/phy015.

McIntosh, Craig, and Andrew Zeitlin. 2022. "Using Household Grants to Benchmark the Cost Effectiveness of a USAID Workforce Readiness Program." *Journal of Development Economics* 157 (June): 102875. https://doi.org/10.1016/j.jdeveco.2022.102875.

McKenzie, David. 2013. "How Should We Understand 'Clinical Equipoise' When Doing RCTs in Development." *World Bank Development Impact Blogs* (blog). September 2013. https://blogs.worldbank.org/impactevaluations/how-should-we-understand-clinical-equipoise-when-doing-rcts-development.

——. 2016. "Have RCTs Taken over Development Economics?" *World Bank Development Impact Blogs* (blog). June 13, 2016. https://blogs.worldbank.org/impactevaluations/have-rcts-taken-over-development-economics.

——. 2020. "If It Needs a Power Calculation, Does It Matter for Poverty Reduction?" *World Development* 127 (March): 104815. https://doi.org/10.1016/j.worlddev.2019.104815.

McKenzie, David, and Berk Özler. 2014. "Quantifying Some of the Impacts of Economics Blogs." *Economic Development and Cultural Change* 62 (3): 567–97. https://doi.org/10.1086/675383.

McRae, Andrew D., Charles Weijer, Ariella Binik, Jeremy M. Grimshaw, Robert Boruch, Jamie C. Brehaut, Allan Donner, et al. 2011. "When Is Informed Consent Required in Cluster Randomized Trials in Health Research?" *Trials* 12 (1): 202. https://doi.org/10.1186/1745-6215-12-202.

Mulligan, Casey B. 2014. "The Economics of Randomized Experiments." *Economix Blog* (blog). March 5, 2014. https://economix.blogs.nytimes.com/2014/03/05/the-economics-of-randomized-experiments/.

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31 (4): 103–24. https://doi.org/10.1257/jep.31.4.103.

O'Flynn, Peter, Chris Barnett, and Laura Camfield. 2016. "Assessing Contrasting Strategies for Ensuring Ethical Practice within Evaluation: Institutional Review Boards and Professionalisation." *Journal of Development Effectiveness* 8 (4): 561–68. https://doi.org/10.1080/19439342.2016.1242643.

Ogden, Timothy. 2017. *Experimental Conversations*. MIT Press. https://mitpress.mit.edu/books/experimental-conversations.

Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." *Journal of Economic Perspectives* 29 (3): 61–80. https://doi.org/10.1257/jep.29.3.61.

Osrin, David, Kishwar Azad, Armida Fernandez, Dharma S. Manandhar, Charles W. Mwansambo, Prasanta Tripathy, and Anthony M. Costello. 2009. "Ethical Challenges in Cluster Randomized Controlled Trials: Experiences from Public Health Interventions in Africa and Asia." *Bulletin of the World Health Organization* 87 (October): 772–79. https://doi.org/10.1590/S0042-96862009001000013.

Panin, Amma. 2020. "Economics Experiments in Africa: How Many and by Whom?" *CODESRIA Bulletin* 4 (1): 23–28.

Quattrochi, John, Jenny C. Aker, Peter van der Windt, and Maarten Voors. 2020. "Contributions of Experimental Approaches to Development and Poverty Alleviation: Field Experiments and Humanitarian Assistance." *World Development* 127 (March): 104830. https://doi.org/10.1016/j.worlddev.2019.104830.

Rodgers, Yana van der Meulen, Anthony Bebbington, Catherine Boone, Jampel Dell'Angelo, Jean-Philippe Platteau, and Arun Agrawal. 2020. "Experimental Approaches in Development and Poverty Alleviation." *World Development* 127 (March): 104807. https://doi.org/10.1016/j.worlddev.2019.104807.

Sabarwal, Shwetlena, David Evans, and Anastasia Marshak. 2014. "The Permanent Input Hypothesis: The Case of Textbooks and (No) Student Learning in Sierra Leone." WPS7021. The World Bank. http://documents.worldbank.org/curated/en/806291468299200683/The-permanent-input-hypothesis-the-case-of-textbooks-and-no-student-learning-in-Sierra-Leone.

Sabet, Shayda Mae, and Annette N. Brown. 2018. "Is Impact Evaluation Still on the Rise? The New Trends in 2010–2015." *Journal of Development Effectiveness* 10 (3): 291–304. https://doi.org/10.1080/19439342.2018.1483414.

Sanchez, Pedro, Cheryl Palm, Jeffrey Sachs, Glenn Denning, Rafael Flor, Rebbie Harawa, Bashir Jama, et al. 2007. "The African Millennium Villages." *Proceedings of the National Academy of Sciences* 104 (43): 16775–80. https://doi.org/10.1073/pnas.0700423104.

Sanmukhani, J., and C. B. Tripathi. 2011. "Ethics in Clinical Research: The Indian Perspective." *Indian Journal of Pharmaceutical Sciences* 73 (2): 125–30.

Sayers, Adrian. 2007. "Tips and Tricks in Performing a Systematic Review." *The British Journal of General Practice* 57 (542): 759.

Sturdy, Jennifer. 2022. "Book Review for Ethics for Evaluation: Beyond 'Doing No Harm' to 'Tackling Bad' and 'Doing Good.'" *Journal of Development Effectiveness*, August. https://doi.org/10.1080/19439342.2022.2102057.

The National Commission for the Protection of Human Subjects of and Biomedical and Behavioral Research. 1979. "Ethical Principles and Guidelines for the Protection of Human Subjects of Research." https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.

The Transfer Project. 2018. "Transfer Project Guidelines for Data Release, Analysis and Publication. Updated November 2018." https://data.cpc.unc.edu/docs/Transfer_Project_-_Guidelines_for_data_release_&_analysis_(Nov_2018).pdf.

Vivalt, Eva. 2019. "Specification Searching and Significance Inflation Across Time, Methods and Disciplines." *Oxford Bulletin of Economics and Statistics* 81 (4): 797–816. https://doi.org/10.1111/obes.12289.

Watima, Tony. 2020. "Big Shame of Researchers Breaking Ethics in Kenya's Low Income Areas." *Nation*, August 18, 2020. https://nation.africa/kenya/business/big-shame-of-researchers-breaking-ethics-in-kenya-s-low-income-areas-1920332.