

RCTs for policy in India - Ethical Considerations, Methodological Concerns and Alternative Approaches

Sneha P*¹

Abstract

Randomised controlled trials (RCTs) have become a popular methodological choice for policy analysis in the developing world. This paper describes the various ethical and methodological considerations when choosing to adopt RCTs for policy decisions through a review of literature in multiple disciplines. Unlike previous critical analysis of RCTs, this paper contextualises its critique to India, a country that has been the site of well over a hundred RCTs. Through illustrations of recent Indian policy RCTs on corruption, livelihoods, Public Distribution System, conflict and others, the paper raises concern about violations of ethical requirements like equipoise, informed consent, data harms, human costs to research participants and research staff. The paper discusses methodological limitations of RCTs for Indian policy making including heterogeneity, researcher effects, generalisability, policy-relevant unobserved mechanisms and other socio-political considerations. The paper ends with a description of alternative approaches and a simple checklist for practitioners, specifically policy makers, to assess the feasibility of RCTs for informing decision making in their context.

Keywords: Randomised Controlled Trials, Research ethics, Methodology, Evaluations, Observational studies, Quasi-experimental Research

Publication Date: 07 May 2021

* Sneha is a senior associate at the IDFC Institute

I. Motivation

The announcement of the 2019 Nobel Prize in Economic Sciences² re-sparked debates about the methodological and ethical foundations of Randomised Controlled Trials (RCTs). This debate brought to the forefront a growing worry that RCTs are crowding out all other approaches (Bédécarrats et al., 2019). While the push for quantitative empiricism and the enthusiastic adoption of methods from natural sciences is not a new trend, it has led to the creation of an implied ‘hierarchy of methodologies’ in development economics. Here, RCTs are seen as an improvement on quasi-experimental methods like discontinuity designs, which are preferred to more observational methods. ‘Structural RCTS’ are seen as a further improvement of standard RCTs. This implied hierarchy has also driven publication norms with implications on the academic discipline of development economics both from a recruitment and publication perspective (Jatteau 2017). As a result, academics studying developing countries have begun to use RCTs for answering policy questions, including socio-politically sensitive subjects like political candidate selection (Casey et al. 2019), HIV testing, sexual safety in partner selection (Angelucci et al. 2016) and even religious education (Bryan et al. 2020).

However, this trend has not gone unchallenged in academic and policy circles. Multiple international aid organizations have begun to rethink their evaluation strategies (Stern et al. 2012). Economists, too, have repeatedly cautioned against adopting a ‘hierarchy of research designs’ (Ravallion 2018) (Deaton 2009) especially in the context of the global south, as such a methodological ordering is not consistently applicable across policy questions and contexts. As Markus Goldstein, a World Bank economist, said in an interview (Shah et al. 2015) ‘It is not wrong that academics want to answer fundamental questions for theory. But let’s not pretend that the policy relevance is always high on those’.

Policy cycles and their research needs have varying constraints on time, cost, data and other resources. The research questions themselves may be general or specific with respect to the geographic context or policy intervention or community of interest. Further, each context and its socio-political dynamics pose varying ethical concerns for research. Therefore, each of these considerations demands methodological variance to be policy relevant as opposed to research or RCTs that are purely used for ‘knowledge creation’ (ibid). Yet, the use of RCTs for policy related decision making or hereafter ‘policy RCTs’ in India continues to grow rapidly and seemingly indiscriminately. So far, India has been the site of 139 evaluations by just one RCT-focussed think tank, J-PAL³, making it one of the two global ‘RCT epicentres’ as Chelwa and Muller (2019) describe, with Kenya being the latter.

While there has been considerable research critically analysing the viability of RCTs from a methodological standpoint, few do so from an ethical perspective. Further, there has been little meta-analysis assessing these considerations in the context of countries that host these trials. This paper aims to review the multidisciplinary literature analysing RCTs and situates this analysis in the Indian context through illustrations of Indian policy RCTs, particularly those that study innovations in public service delivery.

The rest of the paper is organised as follows: Section 2 discusses ethical considerations and describes the various social costs associated with the studies and whether they could ever claim to have a sufficiently big enough policy impact to justify the costs. Section 3 on methodological concerns weighs the technical limitations of RCTs to answer certain kinds of important policy questions. It is followed by Section 4, which presents alternative designs. Section 5 on re-imagining policy RCTs discusses proposed improvements to RCTs, how policy makers (and other development practitioners) can assess the viability of RCTs for their research question and finally, a short discussion on the role of RCTs in development.

II. Ethical Considerations

Most often, RCT work is grounded by ethical regulatory guidelines such as Belmont Principles (respect for persons, beneficence and justice etc.) for conducting research in addition to their Institutional Review Boards (IRBs) (NCPHBBR 1974)⁴. Baele (2013) highlights six ethical problems with RCTs: the randomisation problem (equipoise), the consent problem (informed consent and full disclosure), the instrumentalisation problem (using people as instruments), the accountability problem (being accountable to survey respondents) and finally, foreign intervention (social distance between researcher and respondent).

Consider the study by (Bertrand et al. 2007) as mentioned in (Carter and Barrett 2010), where study participants in Delhi were offered multiple financial incentives to obtain a driving licence as quickly as procedurally possible. This was an effort to test if participants made extra-legal payments to obtain licenses without knowing how to drive. The study found that such incentives encourage corruption. Arguably, any driver or aspiring driver in India could have predicted this outcome. Moreover, note that the study itself directly increased the number of untrained drivers in the city. The following section delves deeper into the ethical concerns of such studies in the context of Indian public policy, focusing on the implications of policy RCTs on survey participants and/or policy beneficiaries.

2.1. Equipoise

The principle of equipoise is the requirement of genuine uncertainty about the merits of a treatment before an experiment to justify running it. While medical trials have been held accountable to the principle of equipoise, the same standards do not apply to social science RCTs. This is more difficult to achieve in the social science RCT context, where there is no experimental blindness (Abramowicz and Szafarz 2019). If anything, there may be pressure in the direction opposite to equipoise to satisfy a state/funder's preference by demonstrating impact (Ravallion 2018).

Consider the driving license example cited earlier. Was there genuine uncertainty about the effect of an intervention to merit such a study? Often, researchers may not experience uncertainty regarding the direction of the effect of an intervention but rather the magnitude of it (especially to make cost-effect calculations). Methodological approaches to incorporate such prior beliefs in experimental research are discussed in Section 3. However, implementing a social science experiment where there is prior knowledge about resultant harm regardless of uncertainty about the magnitude of the harm is ethically questionable.

2.2. Informed Consent, Disclosure and Right to Decline

Researchers have also pointed out that populations are easy to exploit in developing country contexts and are often unaware of their rights to full disclosure about experimental design and subsequent informed consent or denial (Abramowicz and Szafarz 2019). Unlike medical trials, social science RCTs are not often mandated to have full disclosure. This is because if respondents knew they were part of a trial, they would automatically know that they are in the 'treatment' or 'control' group, and they may take steps to change the assignment, taking away from statistical independence of the assignment from treatment. In a systematic review of RCTs in economics journals, Hoffmann (2020) shows that only 10% of papers actually discuss informed consent, and in fact, 12% of studies intentionally left participants ignorant. None of the studies indicated whether participants were explicitly informed about being experimented upon, nor did they discuss whether participation took place because of a financial incentive.

In state-commissioned RCTs, such as those evaluating a variable in a public policy or public entitlement — for instance, a Direct Benefit Transfer (DBT) or a subsidy — respondents do not even have the right to deny participation.

Interestingly, RCT researchers have previously highlighted the risk of ‘over-regulation’ (Glennister and Powers 2013), citing a Kenyan study that was asked to procure written consent from parents before giving children deworming tablets. The authors’ reasoning was that given the low risk of deworming tablets, imposing a written consent may have made it less likely that children would have received the intervention. However, having any a-priori belief about the treatment outcome (that the tablets are low-risk) goes against the principle advantage of RCTs, as argued by its proponents, of being objective and unbiased. Second, arguing that written consent prevents parents from acting in the best possible interest of their child, especially in the realm of medical interventions, places the researcher’s assessment over parental agency in the decision about their children.

Finally, Alderman et al. (2013) also discuss the importance of communicating the *study results* to participants and how that should be viewed as a minimum compensation for study participation (Chambers 2002).

2.3. Non-intrusive and Data Minimalism

More recently, RCT research in India has taken the approach of becoming a ‘public good’ of data, in that data from one RCT spawns multiple research papers because of the comprehensive nature of data collected. However, the downside of this approach is very large quantities and types of data collected from the same households and participants.

For example, a National Science Foundation grant (Award Abstract #1123899) investigating the impact of microfinance in India proposed to collect data on ‘nutrition, food security, health expenditures, physiological indicators of stress through cortisol measurements in hair samples, and psychological stress measures.’ from households in the treatment and control groups of a microcredit program. These form one part of four studies on the impact of microfinance, which is arguably a very well-researched subject. In fact, 3ie’s evidence hub identifies 45 existing impact evaluations of microfinance in India and 20 systematic reviews relevant to the Indian context.

From a researcher’s perspective, there is an incentive to maximise the data collected and outcomes measured in order to study ‘impact’ multi-dimensionally. However, from the respondents’ perspective, collecting more data than required for stated, specific objectives amount to a form of data harm. In doing so, researchers expend participants’ time, effort and resources without having to provide adequate justification. Moreover, given the social distance or power difference between the enumerator and respondent, respondents may also feel uncomfortable refusing participation or seeking more information. This is why researchers need to be respectful, non-intrusive and collect data minimally. This also requires that research questions are themselves narrowly defined with specific outcomes of interest.

Another factor to consider, apart from the cost of respondents’ time and the effect of the intervention on their daily lives, is participant fatigue, which has been commonly observed in parts of India subject to regular RCTs. Not only does this increase the risk of misreporting, but respondents may also learn over time to strategically interact with the study design, thus biasing the results.

2.4. Human Cost

RCTs are often carried out to assess the cost-effectiveness of a programme and calculate the bang for policy buck. Unfortunately, the same principles are not applied to the research methods themselves. Perhaps this is also because the programme funding and evaluation funding are often delineated. (Shah et al. 2015) chart out the time and financial costs of different evaluation methodologies, estimating that RCTs of the highest rigour on average take four years to complete and cost USD 1 million.

However, in the case of welfare policies, long term experiments can also have human costs. Consider the evaluation of the Jharkhand state government's move to mandate biometric authentication for identifying public distribution system (PDS) beneficiaries. Aadhaar-based biometric authentication (ABBA) was made compulsory in August 2016, and an RCT was commissioned in 2017 (Muralidharan et al. 2017). The study period spanned three years and a working paper was published only in February 2020 (Muralidharan et al. 2020). The authors estimated that up to 25,000 beneficiaries from the 132 study blocks had been excluded from their entitlements throughout this trial (ibid) (Misra 2020).

During the study, there had been multiple starvation deaths reported in Jharkhand due to failure of the authentication process (IndiaSpend and Saha 2018). Several exclusionary errors discussed in this working paper were pointed out by journalistic accounts (Scroll.in n.d.) and observational studies very early in implementing this policy (Drèze et al. 2017). It is therefore worth considering the human cost of a policy that is locked into an RCT. While Jharkhand state itself revoked the mandatory status of the Aadhaar card in October 2017 (Indian Express 2017)⁵, as per the study, both the intervention, i.e., ABBA, and the study's endline surveys continued up to December 2017. Therefore, for well over a year, participants could not opt-out of the state policy even if they lost out on their legal entitlements under the National Food Security Act of 2013. The study continued despite early feedback from journalists, activists, other research, and presumably its own preliminary results. Ultimately in 2020, the findings of the working paper were in agreement with the early-stage researchers and 'consistent with the critique that ABBA per se caused at least some 'pain without gain' (ibid)' (Muralidharan et al. 2020).

Policies like ABBA need pilots, continual monitoring and iteration and not a one-time, long-term evaluation. Even in long-term trials, when the intervention has human costs, the real-time feedback from study participants into study design is essential in order to adapt or end trials as soon as there is a realisation that the intervention is causing harm. This has always been a standard practice in medical trials including, as we observed recently, the various treatment trials for COVID-19.

2.5. Ethical Review

One concern about Internal Review Boards (IRB) and RCTs is that there are exemptions given to 'evaluations' which are distinguished from the research of a more exploratory nature because they are project or implementer specific despite having human study subjects and interventions that are comparable to exploratory studies (Glennister and Powers, 2013).

Second, a large number of these IRBs are external to the context of the study. In a systematic review by (Hoffmann 2020), 84% of experiments conducted in former colonies had authors based in institutions in the United States or Western Europe. (Alderman et al. 2013) cite multiple studies with difficult ethical interactions. The author points to the 'Eurocentric bio-medical' model of institutional review as a primary cause that may miss out on both contexts relevant and socially salient ethical questions.

For example, some Indian RCTs include lab-in-field studies where subjects or groups of subjects play behavioural games. Often hypotheses revolve around religious and caste identities in such games requiring

that groups be visibly classified. For example, the RCT by (Bhalotra et al. 2018) reviewed by the IRB at the University of Notre Dame, examines the role of religious identity and political leadership. Because of the nature of the research question, the study chooses Uttar Pradesh precisely because of its history of religious conflict and its ‘salience to politics’. The research question also requires researchers to record participants’ religious identity and, in some games, communicate participants’ religious identity to one another. While that might not have been the case in this study, it is easy to imagine certain contexts where such social identification exists, classification may be socially sensitive. As an example of this, Sarin (2019) cites (Nair and Sambanis 2019), reviewed by the Yale IRB, which randomised Kashmiri respondents into being exposed to violent politically charged media content and its effect on ethnic and national identification.

As Alderman et al. (2013) observe, while the no-harm principle is a norm in development research, they remark that researchers have to be ‘acutely aware’ and ‘go beyond existing protocols’ in order to sufficiently judge welfare considerations. This is why they advocate for more decentralised ethical review institutions. On the other hand, Hoffman (2020) and Sarin (2019) make a case for a moratorium on studies on vulnerable populations until effective regulatory institutions are in place and until the voluntary nature of their participation in such studies can be truly established.

2.6. Research Staff

While previous sections address ethical implications with regard to policy beneficiaries, another group to consider is field⁶ researchers viz. research assistants, enumerators and other staff. Kaplan et al. (2020) discuss these issues in detail in their paper about field research in the global south, focusing on RCTs. The authors highlight five areas of concern: safety and risk of harassment, poor working conditions, emotional burden, role conflict, and inadequate acknowledgement for contributions. Indeed, criticism of unfair terms of employment for local researchers compared to their international counterparts in field research in the global south is common on web portals⁷. It merits much more formal documentation and academic consideration.

It is important to clarify that much of the discussion in this paper applies to other forms of field research, but often conditions are exacerbated when it comes to RCTs, which are characterised by tight timelines, budgets, and ‘limited opportunities to adequately address the complexities in the field’ (ibid).

III. Methodological concerns

Even though there is a consensus that holding RCTs to be a ‘gold standard’ undermines other methods, the narrative persists in several policy-oriented communications. For instance, in a toolkit of evaluations for policy makers, Innovations for Poverty Action advises that quasi-experimental methods should only be used when an ‘RCT is not possible’ (Cowman et al. 2016). The following section discusses the various methodological shortcomings of RCTs from a policy perspective.

3.1. Heterogeneity and stratification

Participant selection and the receipt of treatment before a trial are randomised in order to ensure that groups of individuals are observably similar for each arm of the study. This way, any changes occurring ex-post can be attributed to the treatment alone. However, while it is possible to check whether the randomised samples are balanced on observable characteristics, it is impossible to do so on unobservable

characteristics (like preferences and behaviour). This is because a random sample does not distinguish between the various types of participants within each arm of the trial. There may be very different outcomes for each participant type. (Deaton & Cartwright 2016) argue that randomisation increases noise and skewness because of the asymmetric distribution of treatment outcome indicators.

In the case of a policy implementation process with a no-harm principle, it is important to avoid detrimental effects on sub-groups. Therefore, an RCT must be required to report heterogeneity in treatment effects through subgroup analysis (Baldassarri & Abascal 2017) and out of sample analysis. One way to do this is to stratify samples based on prior information and knowledge, and later, limit the study to only sub-samples that are positively impacted. However, statistical power requirements will have an implication on sample sizes which will have to be larger to accommodate sub-group analysis.

A second consideration related to heterogeneous effects is that often treatment effects in RCTs are calculated on an intent-to-treat basis, i.e., effects are averaged over all those who are eligible for the treatment, not just those who opted into treatment. This is in order to overcome selection biases. However, how participants choose to select into a program may itself be policy relevant. Kabeer (2020) illustrates this with an example of a West Bengal-based RCT where most participants that refused treatment belonged to a religious minority.

3.2. RCT - An intervention in itself

Another important point to consider whilst generalising RCT results is whether programme implementation in an RCT resembles programme implementation by the state. As Baldassarri and Abascal (2017) put it, ‘the people who carry out RCTs (NGO personnel, volunteers, etc.) are an exceptionally competent and motivated group, unlike some public officials who may implement interventions in the long term’. Drèze (2016) in (Deaton & Cartwright 2018) cautions similarly about foreign agencies who implement RCTs and about the power dynamics that accompany the treatment. In her essay about ‘mis-behaving RCTs’, Kabeer (2020) discusses the various ‘tweaks’ implemented in a microcredit RCT research design in Morocco in order to respond to unexpected initial results. Social scientists, in an effort to prove impact, present the best possible version of the intervention during the trial showcasing effect sizes that are not otherwise scalable and therefore, it is important to ‘evaluate the effects of public policy as opposed to its intentions’ (Dubner et al. 2020).

In the Indian context, this is exemplified by the various RCTs on public service delivery. Similarly, in a policy brief about the evaluation of smart cards for governance in Andhra Pradesh, Muralidharan et al. (2012) discuss how despite implementation challenges in the initial launch of the smartcards initiative, researchers worked with the state government “to relaunch the program in eight districts and test its effectiveness through a large-scale randomised evaluation reaching nineteen million people”.

Often, researchers are present at the site of implementation providing inputs into the design of the technological/administrative innovation. Not only does this introduce ‘Hawthorne effects’⁸, but in the process, the researcher may also be eliminating technology design and implementation flaws as a part of the research design in an effort to isolate outcomes arising out of reform from its design and adoption. However, technological/administrative failures and adoption failures are very much part of the political economy of a policy reform. Therefore, RCT results may not hold when the policy is scaled up.

3.3. Generalisability

The second assessment of scalability is the external validity of the RCT results. Deaton and Cartwright (2018) argue that ‘Establishing causality does nothing in and of itself to guarantee generalizability’.

Pritchett and Sandefur (2013), in their paper about the importance of context, discuss how parameter heterogeneity is driven by economy-or institution-wide factors rather than personal characteristics. They argue that estimation of heterogeneous treatment effects in a single localised sample is not enough to claim external validity. Amongst other things, they mention a few requisites for external validity, including the satisfaction of certain invariance laws, enough heterogeneity in models and random placement of the RCTs themselves.

Secondly, researchers have shown that internally valid estimates are also time-variant in the context of indicators such as returns to investments in agriculture, small and medium non-farm enterprises and schooling (Rosenzweig and Udry 2019). This is an important consideration in the Indian context, given that transportability of results even across states would require experiments to have high external validity. Glennerster and Powers (2013) respond to this criticism by suggesting that RCTs are run on large representative samples and acknowledge that these are expensive endeavours.

Another more feasible suggestion proposed by (Gisselquist 2020) is the adoption of a ‘case study approach’ for comparing results across contexts. As Nancy Cartwright describes in a speech about how philosophy can help policy effectiveness, there is a fundamental flaw in the ‘simple induction’ that goes into transporting programs across contexts⁹. Cartwright exemplifies this with the relatively unsuccessful case of the Bangladeshi Integrated Nutrition Program that was modelled on the Tamil Nadu Integrated Nutrition Program without identifying the context-specific dynamics in the causal chain.

3.4. Failure to replicate

The famous worm wars (Evans 2015) were an animated debate within public health and economics about the effectiveness of deworming tablets in promoting school participation. It brought to the forefront several issues of replicability of economics research. Young (2019), in his paper, uses multiple tests of replications to find that only 25 to 50 percent of experimental papers are able to reject the null of no treatment effect anywhere. Part of the reason why replication is difficult terrain for social science experiments compared to natural sciences is that there is a higher risk of bias in the former (Eble et al. 2017).

It is important to note here that replication and reproduction studies are not as highly valued in the economics discipline and find little incentive from universities and journals. In their paper about the failure of programmes to replicate, List et al. (2019) discuss two reasons why a cost-benefit analysis at a trial level does not scale up. This includes ‘researcher competition’, which drives up false positives and also strategic selection of sample populations. They argue that funding for replication studies needs to be higher and that higher benchmarks for reported treatment effects would incentivise funding replications.

3.5. Unobserved individual dynamics

As Cartwright (2010) puts it, “Causes is one word but many things”. This is why Shaffer (2018) calls for an approach of ‘Causal pluralism’ where poverty is not seen as a stock concept but a dynamic flow concept. Because RCTs are often conducted at a household level, one does not observe intra-household individual variation in decision making (Akram-Lodhi 2020), which is important in contexts with gender and age-related power dynamics.

Second, respondents may adapt their preferences and choices to the intervention of an RCT. For example, Das et al. (2013) show that households substitute educational expenses if they anticipate an educational grant which might introduce a secondary change in educational outcomes (where the primary change is through the grant itself). Barrett and Carter (2010) call this a ‘faux exogeneity’ problem, where

a treatment is seemingly exogenous in implementation, but actually, agents heterogeneously receive it. Similarly, Bulte et al. (2020) conduct and compare the results of a double-blinded RCT with a standard (unblinded) RCT of an improved seed intervention and show that treatment effects in a standard RCT are driven by farmers reallocating their best plots to the new seeds.

This is also a warning in (ibid) where the authors warn that subjects' interactions with the intervention may vary and depend on their self-perception or perception of the intervention. This is why the authors advise that theory be taken seriously in 'identifying structural heterogeneity ex-ante of empirical research design'.

Another way of calibrating econometric research and informing its priors is through qualitative research. For example, qualitative research was able to shed light on the intangible outcomes in the success of the programme for women, like the co-operation and contribution of male family members and mentorship of project staff (Kabeer & Datta 2020), which would not be typically measured in an evaluation survey. Similarly, in an evaluation of a citizen participation intervention that simultaneously conducted an RCT and an ethnographic study, researchers were able to discern why the programme failed to achieve desired outcomes because of context-specific politics and qualitative differences in implementation (Rao et al. 2017).

3.6. Political Economy

The historical and political context, which is often more powerful in determining policy results than economic considerations, is another important intangible (Morvant-Roux et al. 2014). Akram-Lodhi (2020) refers to this as 'social-property relations', which may constrain or otherwise influence household choice and behaviour. In the piece, '(Don't) leave politics out of it' Das (2020) illustrates this interplay with examples of how politics changes the impact of policy in India, such as in West Bengal, where the state exhibited preferential treatment toward villages with aligned incumbents. Stevano (2020) describes food security-related RCTs which discuss inter-temporal food decisions of households while ignoring the agri-food industry, intra-household power dynamics and social and cultural food relations.

Vivalt (2019) finds that government-implemented programmes have smaller effect sizes than academic or NGO-implemented programmes, even controlling for sample size. This is because government programme implementation is faced with unique socio-political challenges, especially when interest groups are opposing/promoting a reform process.

In conclusion, understanding more structural dynamics necessitates more general equilibrium work (Acemoglu 2010), more qualitative research (Kabeer and Datta 2020) and, finally, participatory and deliberative dialogue (Rao 2020).

IV. Alternative Designs

"It is almost never the case that an RCT can be judged superior to a well-conducted observational study simply by virtue of being an RCT" (Deaton & Cartwright 2018). This is because often, 'naïve' Ordinary Least Squares (OLS) estimates regressions can be more informative in its own context than an imported RCT result (Pritchett & Sandefur 2013). Therefore, observational research must continue to be an integral first option for policy research in India. Following this, if time and resources permit, other methodological options take into consideration the multi-dimensionality of a policy space and prior information as listed below.

4.1. Observational Research

In his book, *Sense and Solidarity*, Drèze (2017) prescribes the ‘Keep it Simple, Sweetie’ principle for action-oriented research and provides numerous examples of how simple but powerful observational studies have been able to uncover nuanced insights about public policy in the realm of livelihoods, food security etc. in India. This is because, as Drèze (2020) says, ‘good policy requires understanding – not just evidence’. This includes, among other things, ‘observation, reasoning, theory, tradition and debate’.

These observational household studies, often conducted with minimal resources, have been able to inform reforms in policies and legislation like the National Rural Employment Guarantee Act (NREGA), Public Distribution System and Social Security Pensions Scheme etc.¹⁰. One example cited in (ibid) is how household interviews and interactions with workers have been more informative about serious issues like NREGA payment delays than RCTs or other ‘rigorous’ impersonal data analysis. The same holds true for Pratham’s Annual Status of Education Report surveys that have informed several educational programs and reforms in the country¹¹.

Mckenzie (2020) provides a helpful illustration of when it is methodologically enough to simply observe, using effect size and power calculations to show that when treatment effect on the treated group is expected to be large or time effects are expected to be small, observational studies are sufficient.

4.2. Programme-driven iterative adaptation

Pritchett et al. (2013) recommend the use of an evaluation technique called ‘crawling the design space’. This was one of the early papers motivating the toolkit of ‘Program-driven iterative adaptation (PDIA)’. PDIA acknowledges that policy design spaces are not hyper-dimensional and non-linear and therefore encourages policymakers to experiment their way through a policy design process. An important feature of this methodology is having tight feedback loops to learn and iterate quickly (Samji et al. 2018), which are missing in RCTs, as discussed earlier.

PDIA is also in its early stages of adoption for policy purposes and so far has been adopted in multiple West African countries in the realm of public financial management¹² with mixed success. PDIA is also a very time and resource-intensive methodology and may not always be suitable for short decision cycles.

4.3. Experiment-As-Markets

One of the oft-mentioned value propositions of an RCT is that it is free of ‘expert bias’ and that experimental results are not influenced by prior beliefs or judgment. This belief has come under significant criticism not only because social science RCTs cannot prove themselves to be free of bias (especially since they are not double-blinded) but also because it may not be ethical to ignore prior beliefs (equipoise argument) or because one should not undermine the experience, judgement and local knowledge of those in the sector (Leão and Eyal 2020).

Yusuke Narita (2019) recommends overcoming this by including welfare and ethics measures while carrying out stratified experiments, dubbed Experiment-As-Markets. Based on a Food and Drug Administration’s adaptive designs methodology (FDA et al. 2018), (ibid) proposes creating a ‘design market’ using participant preferences and predicted effects in the design, maximising the use of existing prior knowledge. The author finds that such a model improves RCTs when it comes to participant welfare without much loss of information and incentives.

In other words, the preferences of beneficiaries and prior information about how the programme might affect them are used to guide experiment design. This is particularly viable in situations where there is a lot of beneficiary awareness and existing research on the subject, such as microfinance and livelihood programs.

Experiment-as-markets, to the best of the author's knowledge, remains to be implemented in the context of development/policy research.

4.4. Quasi-Experimental Methods

Researchers could also use quasi-experimental methods such as exploiting natural experiments or discontinuities designs that are less intrusive but empirically credible. In 2006, The World Bank published a report on Quality Evaluation under Constraints of Time and Resources (Independent Evaluation Group 2006) providing helpful methods of reducing costs by using quasi-experimental methods to construct comparison groups, reconstructing baseline data, leveraging secondary data and participatory rapid assessments. This includes regression discontinuity and propensity score matching etc. 'Rapid Cycle Research' as a term has been gaining popularity within development research with origins in public health (Johnson et al. 2015).

V. Re-imagining policy RCTs

5.1. Improving RCT design

In his defence of RCTs, Imbens (2018) describes how this area of research has become interdisciplinary, including researchers from computer science and statistics etc. However, perhaps it will also benefit from the participation of sociologists, anthropologists and gender studies experts etc. In his response to (Deaton and Cartwright 2018), the author also highlights several methodological improvements within the RCT literature to address concerns about validity and unobservables. For instance, to address the issue of long term unobservables, Athey et al. (2016) suggest analysing 'multiple statistical surrogates', i.e. immediate observables that can predict long-term treatment effects. The paper described the various independence conditions to be met for this and the information to be gained.

Several economists have suggested block randomisation models and regression tree models to tease out heterogeneity within treatment effects, which allow information about differential treatment effects to feed into research design. In their paper, Green and Kern (2012) discuss Bayesian Additive Regression Trees to test for systematically varying treatment effects in experimental data. Imbens and Athey (2016) suggest the use of recursive partitioning by selecting different samples to estimate heterogeneity and subsequently test hypotheses about differential treatment effects across subgroups.

Muralidharan and Neihaus (2017) recommend the adoption of RCTs on a large scale to observe significant variation to measure heterogeneity and observe the administrative and technological failures of 'scaling up' as discussed earlier. They illustrate the benefits of this approach with their state-wide RCT of biometric smart cards in Andhra Pradesh. However, as discussed in previous sections, such studies must be approached cautiously with sufficient piloting, checks and balances, and participants' option to opt-out.

In terms of improving the accountability of RCTs, Burlig (2016) discusses the various pre-analysis plan measures that can be taken by researchers who collect their own data, conduct prospective studies, and research using restricted-access data. Policy RCTs in India have begun to register their pre-analysis plans

to prevent (publication and specification) bias and open up to ethical scrutiny. However, while several such plans are available online on platforms such as Social Science Registry¹³, much less has been done on communicating these plans and creating transparency around them.

While several universities conducting research in India and Indian think tanks like IFMR have IRBs for social science RCTs, these RCTs are not subject to standardised norms, unlike, for instance, medical RCTs which report to ethical standards such as CONSORT. Non-medical RCTs also need to be subject to stringent ethical evaluation locally, just like medical RCTs (Srinivasan 2009).

5.2. For Policy Makers

From the perspective of governments in India, particularly state governments who partner with various universities and organisations in the social sector to run RCTs, identifying certain prerequisites for suggesting RCT as a suitable method would be useful. While RCTs for creating knowledge and progression of social understanding need not be concerned with the same factors, policy decision-focused RCTs should be demand-driven, tailored, embedded and cost-effective (Shah et al. 2015). Listed below are a few fundamental preconditions that could be used to assess the viability of an RCT for answering a policy question. This is not a comprehensive list of considerations, but a starting point for policy makers and researchers.

1. Is there enough uncertainty about the nature of the impact of the treatment on participants to merit an experiment?
2. If government/researchers/other stakeholders have prior beliefs on impact, can the research design be informed by these priors? For example,
 - a. If their prior on impact is positive, can treatment be randomly phased in so that the control group is not deprived of the intervention?
 - b. If their prior is negative, will the study be adapted upon first observation of negative effects?
 - c. If certain groups are expected to be impacted differently, can the study be stratified to minimise negative effects and maximise positive effects?
3. Would participants have the opportunity to opt-in and opt-out in an informed manner?
4. Is data collected non-intrusive, minimal and restricted to the objective of the study under consideration?
5. Does the policy question that the study seeks to answer require the investment of a large-scale trial (cost-effectiveness)?
6. Is the study duration able to meet the urgency of the policy question (time-effectiveness)?
7. Does the program or policy require continual monitoring or a one-time evaluation?
8. What are the limits of the context within which the potential findings of the trial will be held valid?
9. Will the study design be ethically reviewed by those familiar with the context of the study?
10. Will the administration of the intervention within the study resemble the administration of the intervention in an eventual scale-up effort?
11. Are there sufficient checkpoints during the study, and does design adapt to findings at these checkpoints?
12. What other methodologies can be simultaneously adopted to complement and inform the trial?

Consider environmental conservation, a sector requiring urgent policy action and multiple stakeholder consultation, and scientific study. Such a policy area necessitates ‘methodological hybridity’ (Ali 2020).

RCTs could be one component of this, but certainly not the primary one. This is also true with regard to other policy areas such as political participation, livelihood development, food security and gender empowerment etc. As Haan et al. (2020) remark in the context of RCTs in gender-related work, methodologies need not substitute one another but should adopt complementary approaches with different disciplines, aiming at answering ‘What’, ‘Why’ and ‘How’.

One set of policy RCTs that meet most of the above criteria are small RCTs that evaluate informational ‘nudge’ interventions implemented in limited contexts of improving citizen behaviour — for example, tax payment, energy use, garbage disposal, public urination etc. In most of these interventions, respondent participation has little scope to harm the participant or society. From the policymaker’s perspective, there is often uncertainty about the exact design features which would elicit the best civic response. The interventions, often informational or communicational, tend to be cheap, leading to the studies being cost-effective and easily replicable in multiple contexts.

However, it should not be forgotten in the Indian context that even such communication interventions can adopt strong communal and patriarchal overtones. The nudge campaign ‘*Beti Aapki Dhan Lakshmi Aur Vijay Lakshmi (BADLAV)*’ (Your daughter is the goddess of wealth and success), described in the Indian Economic Survey 2018-19 is a perfect example of this.

5.3. Role in Development

It is worth considering what kind of research and evidence have successfully motivated policy innovations and reform historically. In their paper on analysing the historical role of RCTs for public policy, Leão and Eyal (2020) show that during the ‘first wave’ of RCTs in public health and education (from the 1960s to 1980s), researchers realised that administrators did not prioritise ‘experimental control’ which is often ‘politically inappropriate’.

Instead, RCTs were adapted into quasi-experimental designs and their purpose understood to be more research than evaluation. As Humphreys and Scacco (2020) put it, there is a micro-macro disconnect as findings of an RCT do not sufficiently aggregate to solve a macro problem. Indeed, the focus on public service delivery in RCT literature is justified because this is the one aspect of public policy that is particularly amenable to empirical experimentation (Kapur 2020).

As this paper has hopefully illustrated, RCTs’ claim to ‘neutrality’ cannot hold from an *economic* perspective since the analysis is largely rooted in neoclassical microeconomic theory (Kvangraven 2020) and from a *political* perspective because these interventions are designed for and implemented in partnership with the government. Withholding information from participants has implications for a representative democracy where participants have a right to express their policy preferences. Instead, the experiments should embrace contextual complexities and allow them to inform design and analysis.

VI. Concluding Remarks

Since this paper focuses on policy RCTs and consequently the state as an actor, it does not discuss other stakeholders in detail. Other development practitioners, particularly donors, are crucial players in the design and implementation of RCTs and the adoption of results. Indeed, one explanation for the proliferation of RCTs has been donor organisations’ focus on ‘empiricism’ and ‘results-driven programming’.

Second, this paper tries to list ethical and methodological considerations distinctly for the sake of coherence, but in actuality, most of these concerns are interlinked. For instance, inadequate heterogeneity analysis can lead to failures of external validity and replication. Similarly, being inconsiderate toward socio-political dynamics at a stakeholder or population level can lead to ethically insensitive design choices or as Barrett and Carter (2010) phrase it, treating humans as ‘subjects’ rather than ‘agents’. As India sees a greater adoption of evidence-based policy research, how such evidence is generated and used to implement policy at scale will require careful examination over the coming years.

References

- Abramowicz, Michel, and Ariane Szafarz. 2019. "Ethics of Randomized Controlled Trials: Should Economists Care about Equipoise?" *SSRN Electronic Journal*, October. <https://doi.org/10.2139/ssrn.3465762>.
- Acemoglu, Daron. 2010. "Theory, General Equilibrium, and Political Economy in Development Economics." *Journal of Economic Perspectives* 24 (3). <https://doi.org/10.1257/jep.24.3.17>.
- Akram-Lodhi, A. Haroon. 2020. "Follow the Yellow Brick Road?: Structural Shortcomings in Randomized Control Trials." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104803>.
- Alderman, Harold, Jishnu Das, and Vijayendra Rao. 2013. "Conducting Ethical Economic Research: Complications from the Field." *World Bank Policy Research Working Paper* 6446 (May). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2265757
- Ali, Saleem H. 2020. "Environmental Urgency versus the Allure of RCT Empiricism." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104844>.
- Angelucci, Manuela, Daniel Bennett, Jenny Trinitapoli, Dan Black, Alessandra Voena, Sara Yeatman Hannah Furnas, Ashley Larsen Gibby, and Ngoc Nguyen. 2016. "Lemons: HIV Testing and Marriage in Rural Malawi". National Institute of Child Health and Human Development. https://economics.yale.edu/sites/default/files/marriage_v6.pdf
- Athey, Susan, Raj Chetty, Guido Imbens, and Hyunseung Kang. 2016. "Estimating Treatment Effects Using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index." Cornell University. <https://arxiv.org/abs/1603.09326>
- Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences of the United States of America* 113 (27). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4941430/>
- Baele, Stéphane J. 2013. "The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?" *Journal of Philosophical Economics* VII (1): 2–42. <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/17048/EthicsRCT.pdf?sequence=1&isAllowed=y>
- Baldassarri, Delia, and Maria Abascal. 2017. "Field Experiments Across the Social Sciences" *The Annual Review of Sociology* <https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-073014-112445>
- Barrett, Christopher B., and Michael R. Carter. 2010. "The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections." *Applied Economic Perspectives and Policy* 32 (4) http://barrett.dyson.cornell.edu/Papers/Power_and_Pitfalls_of_Experiments_in_Development_Economics_Oct_2010_Final.pdf
- Barrett, Christopher B., and Michael R. Carter. 2020. "Finding Our Balance? Revisiting the Randomization Revolution in Development Economics Ten Years Further On." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104789>.

- Bédécarrats, Florent, Isabelle Guérin, and François Roubaud. 2019. "All That Glitters Is Not Gold. The Political Economy of Randomized Evaluations in Development." *Development and Change* 50 (3): 735–6. <https://onlinelibrary.wiley.com/doi/abs/10.1111/dech.12378>
- Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2007. "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption." *Quarterly Journal of Economics*. https://scholar.harvard.edu/files/remahanna/files/1_qje_driving_license.pdf
- Bhalotra, Sonia, Irma Clots-Figueras, Lakshmi Iyer, and Joseph Vecci. 2018. "Leader Identity and Coordination." IZA Discussion Paper, no. 11803. <http://ftp.iza.org/dp11803.pdf>
- Bryan, Gharad, James J. Choi, and Dean Karlan. 2021. "Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes." *Quarterly Journal of Economics* 136 (1). <https://doi.org/10.1093/qje/qjaa023> .
- Bulte, Erwin, Salvatore Di Falco, and Robert Lensink. 2020. "Randomized Interventions and 'Real' Treatment Effects: A Cautionary Tale and an Example." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104790> .
- Burlig, Fiona. 2016. "Improving Transparency in Observational Social Science Research: A Pre-Analysis Plan Approach." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2731920> .
- Cartwright, Nancy. 2010. "Hunting Causes and Using Them: Approaches in Philosophy and Economics." *Structural Equation Modeling: A Multidisciplinary Journal* 17 (3). [http://eprints.lse.ac.uk/32087/1/Hunting%20causes%20and%20using%20them%20\(LSERO\).pdf](http://eprints.lse.ac.uk/32087/1/Hunting%20causes%20and%20using%20them%20(LSERO).pdf)
- Casey, Katherine, Abou Bakarr Kamara, and Niccoló Meriggi. 2019. "An Experiment in Candidate Selection." *National Bureau of Economic Research Working Pa.* <https://www.nber.org/papers/w26160.pdf> .
- Chambers, Robert. 2002. "Qualitative Approaches: Self-Criticism and What Can Be Gained from Quantitative Approaches" in Ravi Kanbur (Ed) *Q-Squared: Qualitative and Quantitative Methods in Poverty Appraisal.* <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/148>
- Chelwa, Grieve, and Seán Muller. 2019. "The Poverty of Poor Economics." *Africa Is a Country*, 2019. <https://africasacountry.com/2019/10/the-poverty-of-poor-economics#:~:text=Countries%20that%20are%20now%20developed,be%20achieved%20without%20randomised%20trials>.
- Cowman, Jennifer, Heidi Mcannally-linz, and Megan Mcguire. 2016. "Impact Measurement with the CART Principles," https://www.poverty-action.org/sites/default/files/publications/Goldilocks-Toolkit-Impact-Measurement-with-CART-Principles_0.pdf
- Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School Inputs, Household Substitution, and Test Score." *American Economic Journal: Applied Economics* 5 (2). <https://doi.org/10.1257/app.5.2.29>.
- Das, Sabyasachi. 2020. "(Don't) Leave Politics out of It: Reflections on Public Policies, Experiments, and Interventions." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104792>.

- Deaton, Angus. 2009. "Instruments of Development." National Bureau of Economic Research. <https://doi.org/10.1017/CBO9781107415324.004> .
- Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine*, no. 667526. <https://www.sciencedirect.com/science/article/pii/S0277953617307359>
- DFID. 2012. "Broadening the Range of Designs and Benefits of Trade Methods for Impact Evaluations," <https://www.oecd.org/derec/50399683.pdf>
- Drèze, Jean. 2017. *Sense and Solidarity - Jholawala Economics for Everyone*. Oxford University Press. <https://global.oup.com/academic/product/sense-and-solidarity-9780198833468?cc=ph&lang=en&>
- Drèze, Jean. 2020. "Policy beyond Evidence." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104797> .
- Drèze, Jean, Nazar Khalid, Reetika Khera, and Anmol Somanchi. 2017. "Aadhaar and Food Security in Jharkhand." *Economic and Political Weekly* 52 (50): 50–59. <https://www.epw.in/journal/2017/50/special-articles/aadhaar-and-food-security-jharkhand.html>
- Dubner, Stephen J., John A. List, and Dana Suskind. 2020. "Policymaking Is Not a Science (Yet) (Ep. 405)." *Freakonomics*. 2020. <https://freakonomics.com/podcast/scalability/> .
- Duvendack, Maren, Richard Palmer-Jones, James G Copestake, Lee Hooper, Yoon Loke, and Nitya Rao. 2011. *What Is the Evidence of the Impact of Microfinance on the Well-Being of Poor People?* EPPI Centre. http://www.givedirectly.org/wp-content/uploads/2019/06/DFID_microfinance_evidence_review.pdf
- Eble, Alex, Peter Boone, and Diana Elbourne. 2017. "On Minimizing the Risk of Bias in Randomized Controlled Trials in Economics." *World Bank Economic Review* 31 (3): 687–707. <https://doi.org/10.1093/wber/lhw034> .
- Evans, David. 2015. "Worm Wars: The Anthology." 2015. <https://blogs.worldbank.org/impactevaluations/worm-wars-anthology> .
- FDA, CDER, and CBER. 2018. "Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry Draft Guidance." <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry>
- Gisselquist, Rachel M. 2020. "How the Cases You Choose Affect the Answers You Get, Revisited." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104800> .
- Glennerster, Rachel, and Shawn Powers. 2016. "Balancing Risk and Benefit: Ethical Tradeoffs in Running Randomized Evaluations." In *The Oxford Handbook of Professional Economic Ethics*. <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199766635.001.0001/oxfordhb-9780199766635-e-017>
- Gugerty, Mary Kay, Dean Karlan, and Delia Welsh. 2016. "Goldilocks Toolkit: Impact Measurement with the CART Principles." *Innovations for Poverty Action*. <https://www.poverty-action.org/publication/goldilocks-toolkit-impact-measurement-with-cart-principles>

- Haan, Arjan de, Gillian Dowie, and Jane Mariara. 2020. "To RCT or Not, Is Not the Question: Methods for Policy-Relevant Research on Gender Equality." *World Development*. <https://www.sciencedirect.com/science/article/abs/pii/S0305750X19304437>
- Hoffmann, Nimi. 2020. "Involuntary Experiments in Former Colonies: The Case for a Moratorium." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104805> .
- Humphreys, Macartan, and Alexandra Scacco. 2020. "The Aggregation Challenge." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104806> .
- Imbens, Guido. 2018. "Comments On: Understanding and Misunderstanding Randomized Controlled Trials." *Stanford Working Paper*. <https://www.gsb.stanford.edu/faculty-research/publications/comments-understanding-misunderstanding-randomized-controlled-trials>
- Independent Evaluation Group. 2006. *Conducting Quality Impact Evaluations Under Budget Time and Data Constraints*. World Bank. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/255551468178129539/conducting-quality-impact-evaluations-under-budget-time-and-data-constraints>
- Indian Express. 2017. "Aadhar Card Not Mandatory to Get Ration in Jharkhand: Minister." PTI, 2017. <https://indianexpress.com/article/india/aadhar-card-not-mandatory-to-get-ration-in-jharkhand-minister-4900064/> .
- IndiaSpend, and Devanik Saha. 2018. "Are 14 Deaths Due to 'starvation' in Jharkhand Linked to Aadhaar Glitches." *Business Standard*. https://www.business-standard.com/article/current-affairs/are-14-deaths-due-to-starvation-in-jharkhand-linked-to-aadhaar-glitches-118081100216_1.html .
- Jatteau, Arthur. 2017. "Arthur Jatteau, Faire Preuve Par Le Chiffre ? Le Cas Des Expérimentations Aléatoires En Économie." *Revue de La Régulation*, no. 21. <https://doi.org/10.4000/regulation.12323>. <https://journals.openedition.org/regulation/12323>
- Johnson, Kim, Bernard Ewigman, Lloyd Provost, and Rebecca Roper. 2015. "Using Rapid-Cycle Research to Reach Goals: Awareness, Assessment, Adaptation, Acceleration." *Agency for Healthcare Research and Quality*. https://pbrn.ahrq.gov/sites/default/files/docs/AHRQ_PBRN_Webinar_Rapid_Cycle_Research_9.9.15.pdf
- Kabeer, Naila, and Sanchari Datta. 2020. "Randomized Control Trials and Qualitative Impacts: What Do They Tell Us about the Immediate and Long-Term Assessments of Productive Safety Nets for Women in Extreme Poverty in West Bengal?." *London School of Economics and Political Science (LSE), Department of International Development, London Working Paper Series (No. 19-199)*. <https://www.econstor.eu/bitstream/10419/224825/1/wp199.pdf>.
- Kaplan, Lennart, Jana Kuhnt, and Janina I. Steinert. 2020. "Do No Harm? Field Research in the Global South: Ethical Challenges Faced by Research Staff." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104810>.
- Kapur, Devesh. 2020. "Poverty, Power and RCTs." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104811>.

- Kvangraven, Ingrid Harvold. 2020. "Impoverished Economics? A Critical Assessment of the New Gold Standard." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104813>.
- Leão, Luciana de Souza, and Gil Eyal. 2020. "Searching under the Streetlight: A Historical Perspective on the Rise of Randomistas." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104781>.
- List, John A., Dana Suskind, and Omar Al-Ubaydli. 2019. "The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments." *SSRN Electronic Journal*, no. May: 0–33. <https://bfi.uchicago.edu/working-paper/the-science-of-using-science-towards-an-understanding-of-the-threats-to-scaling-experiments/>
- Low, Daniel C.M., Helen Roberts, and Rosalind H. Whiting. 2015. "Board Gender Diversity and Firm Performance: Empirical Evidence from Hong Kong, South Korea, Malaysia and Singapore." *Pacific Basin Finance Journal* 35: 381–401. <https://doi.org/10.1016/j.pacfin.2015.02.008>.
- McKenzie, David. 2020. "If It Needs a Power Calculation, Does It Matter for Poverty Reduction?" *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104815>.
- Mckenzie, David. 2016. "Have RCTs Taken over Development Economics." *World Bank Blogs*. 2016. <https://blogs.worldbank.org/impactevaluations/have-rcts-taken-over-development-economics>.
- Ministry of Finance. "Economic Survey 2018-19." India. <https://www.indiabudget.gov.in/budget2019-20/economicsurvey/>
- Misra, Udit. 2020. "Karthik Muralidharan: 'To an Extent, Both Supporters and Critics of Aadhaar for Service Delivery Are Correct.'" *The Indian Express*. 2020. <https://indianexpress.com/article/governance/karthik-muralidharan-to-an-extent-both-supporters-and-critics-of-aadhaar-for-service-delivery-are-correct-6283226/>.
- Morvant-Roux, Solène, Isabelle Guérin, Marc Roesch, and Jean Yves Moissoner. 2014. "Adding Value to Randomization with Qualitative Analysis: The Case of Microcredit in Rural Morocco." *World Development* 56 (201): 302–12. <https://doi.org/10.1016/j.worlddev.2013.03.002>.
- Muller, Seán M. 2020. "The Implications of a Fundamental Contradiction in Advocating Randomized Trials for Policy." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104831>.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2017. "Pre-Analysis Plan: EPDS Impact Evaluation in Jharkhand." *Social Science Registry*. <https://www.socialscienceregistry.org/versions/39275/docs/version/document>
- Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." In *Journal of Economic Perspectives*. Vol. 31. <https://doi.org/10.1257/jep.31.4.103>.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2020. "Identity Verification Standards in Welfare Programs: Experimental Evidence from India." <https://www.socialscienceregistry.org/trials/1620>.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2012. "Improving Governance Through Biometric Authentication and Secure Payments in India." *JPAL*

- <https://www.povertyactionlab.org/evaluation/improving-governance-through-biometric-authentication-and-secure-payments-india>
- Nair, Gautam, and Nicholas Sambanis. 2019. "Violence Exposure and Ethnic Identification: Evidence from Kashmir." International Organization. <https://doi.org/10.1017/S0020818318000498>.
- Ould Mohamedou, Mohammad-Mahmoud. 2020. "Burning the Field." *Le Temps*, 2020. <https://blogs.letemps.ch/mohammad-mahmoud-ould-mohamedou/2020/06/21/burning-the-field/>.
- Pritchett, Lant, Salimah Samji, and Jeffrey Hammer. 2013. "It's All About MeE: Using Structured Experiential Learning to Crawl the Design Space," no. April 2013. <https://bsc.cid.harvard.edu/publications/its-all-about-mee-using-structured-experiential-learning-e-crawl-design-space>
- Pritchett, Lant, and Justin Sandefur. 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix." *SSRN Electronic Journal*, no. August 2013. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2364580
- Rao, Vijayendra. 2020. "Evidence-Based Development Needs a Diversity of Tools, with a Bottom-up Process of 'Embedded' Dialogue." *World Development*. <https://doi.org/10.1016/j.worlddev.2019.104823>.
- Rao, Vijayendra, Kripa Ananthpur, and Kabir Malik. 2017. "The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India." *World Development* 99. <https://doi.org/10.1016/j.worlddev.2017.05.037>.
- Ravallion, Martin. 2018. "Should the Randomistas (Continue to) Rule? Martin Ravallion Should the Randomistas (Continue to) Rule?" <https://www.nber.org/papers/w27554>
- Rosenzweig, Mark R, and Christopher Udry. 2020. "External Validity in a Stochastic World: Evidence from Low-Income Countries." *The Review of Economic Studies* 87 (1). <https://doi.org/10.1093/restud/rdz021>.
- Samji, Salimah, Matt Andrews, Lant Pritchett, and Michael Woolcock. 2018. "PDIA Toolkit - A DIY Approach to Solving Complex Problems." *Building State Capability*. https://bsc.cid.harvard.edu/files/bsc/files/pdiatoolkit_ver_1_oct_2018.pdf
- Sarin, Ankur. 2019. "Indecent Proposals in Economics." *India Forum*, 2019. <https://www.theindiaforum.in/article/indecnt-proposals-economics>
- Scroll. n.d. "Identity Project." <https://scroll.in/topic/38792/identity-project>.
- The Guardian. Secret Aid Worker. 2016. "The Field' Is Not a Lab Where You Can Experiment without Consequence." <https://www.theguardian.com/global-development-professionals-network/2016/jul/12/secret-aid-worker-field-fieldwork-neocolonial-vocabulary>
- Shaffer, Paul. 2018. "Causal Pluralism and Mixed Methods in the Analysis of Poverty Dynamics," United Nations University - UNU Wider. <https://www.econstor.eu/handle/10419/190162>

- Shah, Neil Buddy, Paul Wang, Andrew Fraker, and Daniel Gastfriend. 2015. "Evaluations with Impact Decision-Focused Impact Evaluation as a Practical Policymaking Tool." 3ie Working Paper 25. <https://www.3ieimpact.org/evidence-hub/publications/working-papers/evaluations-impact-decision-focused-impact-evaluation>
- Srinivasan, Sandhya. 2009. "Ethical Concerns in Clinical Trials in India". Centre for Studies in Ethics and Rights. https://www.researchgate.net/publication/265071503_Ethical_concerns_in_clinical_trials_in_India_a_n_investigation
- Stevano, Sara. 2020. "Small Development Questions Are Important, but They Require Big Answers." World Development. <https://doi.org/10.1016/j.worlddev.2019.104826>.
- U.S. Department of Health and Human Services (HHS). 1978. "Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research." <https://pubmed.ncbi.nlm.nih.gov/25951677/>
- Vivalt, Eva. 2019. "How Much Can We Generalize from Impact Evaluations ?" Working Paper, Australian National University. <http://evavivalt.com/wp-content/uploads/How-Much-Can-We-Generalize.pdf>
- Vivalt, Eva. 2016. "How Much Can Impact Evaluations Inform Policy Decisions?" Australian National University. http://devpolicy.org/2017-Australasian-Aid-Conference/Presentations/Panel3d_Vivalt.pdf
- Young, Alwyn. 2016. "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." Quarterly Journal of Economics. <http://economics.mit.edu/files/11362>
- Yusuke Narita. 2019. "Experiment-as-Market - Incorporating Welfare into Randomized Controlled Trials." Human Capital and Economic Opportunity Global Working Group Working Papers. <https://ideas.repec.org/p/hka/wpaper/2019-025.html>

Notes

¹ The author is grateful to Niranjana Rajadhyaksha and Vikram Sinha for their inputs and editorial oversight. The author would also like to thank Jean Drèze, Vijayendra Rao, Vaidehi Tandel, Chinmaya Kumar, Anirudh Tagat, Ashwin Nair, Ashwin MB, Tanvi Ravel Mehta, Abdul W. A. Mohammed and Anmol Somanchi, for their comments and suggestions. All errors remain the author's own.

² Officially known as The Sveriges Riksbank Prize in Economic Sciences

³ Source: <https://www.povertyactionlab.org/evaluations> accessed 03 March 2020

⁴ See <https://www.povertyactionlab.org/ethics>

⁵ However, by September 2018, a court ruling restored the state's ability to mandate Aadhaar for social programs including PDS.

⁶ The term 'field' too has also been oft criticised for its subtext of being a reference to the 'Global South' and consequent implication of skewed power dynamics (Ould Mohamedou 2020) (The Guardian 2016).

⁷ For example see The Bukavu Series <https://www.gicnetwork.be/silent-voices-blog-bukavu-series-eng/>

⁸ 'Hawthorne effects' describe a bias introduced in empirical research as a result of participants being aware that they are being studied and consequently modifying their behaviour, based on the work of Mayo E. and Roethlisberger F. at the Hawthorne plant of the Western Electric Company in the 1920s.

⁹ See 'Will this policy work for you? - Predicting effectiveness better: How philosophy helps', Nancy Cartwright, LSE and UCSD, Presidential Address PSA 2010

https://philosophy.ucsd.edu/_files/ncartwright/phil152/PSA-2-Nov-0900.pdf

¹⁰ See Public Evaluation of Entitlement Programmes (PEEP) Survey 2013 and Public Distribution System (PDS) Survey 2011 available at <http://web.iitd.ac.in/~reetika/projects.html>

¹¹ See Arvind Subramanian (former Chief Economic Advisor, India) in an interview with Devesh Kapur <https://casi.sas.upenn.edu/iit/deveshkapur2018>

¹² See Collaborative Africa Budget Reform Initiative (CABRI) <https://www.cabri-sbo.org/en/search?q=My+PDIA+Journey>

¹³ See <https://www.socialscienceregistry.org/>