DOCUMENT DE TRAVAIL DT/2020-15

Controversies around RCT in Development

Epistemology, Ethics, and Politics

Florent BEDECARRATS
Isabelle GUÉRIN
François ROUBAUD

UMR LEDa
Place du Maréchal de Lattre de Tassigny 75775 • Paris •Tél. (33) 01 44 05 45 42 • Fax (33) 01 44 05 45 45
DIAL • 4, rue d'Enghien • 75010 Paris • Tél. (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51
E-mail : dial@dial.prd.fr • Site : www.dial.ird

Controversies around RCT in Development

Epistemology, Ethics, and Politics[1]

Florent Bédécarrats, Isabelle Guérin and François Roubaud

Abstract

The article questions the scope of application of RCTs in the field of development, what they can achieve, why they sometimes fail and why other methods are both useful and necessary. It sets out the points of disagreement between supporters and critics of RCTs, which are both epistemological (scientism versus pragmatism), political (is development an aggregate of micro-interventions or a structural, institutional and political change? Is poverty understood a matter of privation or as a process leading to unequal power relationships?), and ethical (should we at all costs privilege the advancement of science or protect populations?). The paper draws the lessons, while updating them, from a recent collective book edited by the authors. Given the constraints of the method, RCTs are usually limited to micro-interventions and private goods and to behavioural tests in response to varied types of interventions. The inability of RCTs to respond to the poverty challenges posed by the current pandemic largely confirms our analyses. The article argues for the need of a real scientific controversy and proposes ways of improving RCTs and methodological alternatives.

Keyword: Covid, Ethics, Political Economy, Poverty, Randominzed Control Trials, Scientific Controversy

JEL Codes: C10, C83, C93, M31, O10, O43

Résumé

Cet article questionne la portée réelle des expériences randomisées (RCT) dans le champ du développement, ce qu'elles peuvent permettre de mesurer, pourquoi parfois elles échouent et pourquoi d'autres méthodes sont à la fois utiles et nécessaires. L'article expose les points de divergence entre les promoteurs des RCT et leurs critiques, qui sont à la fois épistémologiques (scientisme versus pragmatisme), politiques (est-ce que le développement est une constellation de micro-interventions ou synonyme de changements structurels, institutionnels et politiques ? La pauvreté est-elle une accumulation de "manques" ou un processus qui conduit à des relations de pouvoir inégales ?), et éthiques (doit-on privilégier à tout prix ce que l'on considère comme des progrès de la science ou protéger les populations ?). Cet article tire les enseignements, tout en les actualisant, d'un ouvrage collectif édité par les auteurs. Etant données les limites de la méthode, les RCT sont en général limitées à des micro-interventions et des biens privés, ou à tester des changements de comportements induits par différentes modalités d'interventions. L'incapacité des RCT à répondre au défi de l'explosion de la pauvreté liée à la pandémie confirme nos analyses. Enfin, l'article plaide en faveur d'une véritable controverse scientifique, jusqu'ici évitée, et propose des moyens d'améliorer les RCTs ainsi que des méthodologies alternatives.

Mots-clefs: Covid, Ethique, Economie politique, Expérience aléatoire, Pauvreté, Randominzed Control Trials

1 This paper elaborates on a preprint copy of our book introduction (Bédécarrats, Guérin and Roubaud 2020b), augmented with some insights on the potential contribution of RCTs to tackle poverty in time of COVID-19.

Introduction

In October 2019, Abhijit Banerjee, Esther Duflo and Michael Kremer jointly won the 51st Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. The three researchers were awarded "*for their experimental approach to alleviating global poverty*" and for having "*turned development economics — the field that studies what causes global poverty and how best to combat it — into a blossoming, largely experimental field*" (The Royal Swedish Academy of Sciences 2019: 2). The use of field experimentation, unlike laboratory experimentation, serves to conduct full-scale tests on interventions, behaviour and decision-making in the "real world'", and then to "make causal claims of impact" (*ibid*: 3). Consequently, stated the jury, "*We now have a large number of concrete results on specific mechanisms behind poverty and specific interventions to alleviate it*," (*ibid*). The cases of health, schooling, gender and politics, and credit are given as powerful illustrations of the laureates' achievements in their work. This award recognizes the success of a long-standing method inspired by the medical field – randomized control trials (hereinafter referred to as RCTs) – and now applied to poverty and development issues. The award did not really come as a surprise. RCTs were first launched in development in the early 2000s and have since become increasingly successful among academics, donors and development practitioners to the extent that RCTs are now considered the gold standard for the evaluation of anti-poverty policies and understanding the origins of poverty.

While there are reasons to welcome the prize (one of the three laureates is a young woman,[2] and the award brings to the fore the issue of poverty and the collection of primary data, which has long been passed over by development economics), there is also cause to raise questions about the validity and repercussions of the growing use of this method, which the prize may boost further. What scope do RCTs actually have? Have they really "dramatically improved our ability to fight poverty in practice," as suggested by the Sveriges Riksbank Prize jury? Which sorts of questions are RCTs able to address and which do they fail to answer? Is causal explanation the only way to understand poverty and do RCTs systematically manage to provide causal explanations? Last, but not least, is the supremacy of experimentation in development economics, as recognized and commended by the Nobel jury, scientifically legitimate and politically desirable?

2 The prize first awarded in 1969 has been won by a total of 84 laureates. Esther Duflo is only the second-ever female laureate. Over and above the prize itself, economics as a social science is the most marked by discrimination against women (Lundberg & Stearns 2019).

The volume we edited proposes precisely to answer these questions (Bédécarrats, Guérin and Roubaud 2020a). The initiative for our editorial project came from the EUDN[3] conference on *Malaise dans l'Evaluation* (Evaluation and its Discontents) held by AFD[4] in Paris in 2012 (AFD, 2012). At this event, we witnessed a real dialogue of the deaf. Whereas some critical voices set out the reasons for their doubts, those who we will call the *randomistas*,[5] in keeping with others and for the sake of expediency, confidently presented their convictions and their findings, sidestepping any substantive discussion of the matter.

We therefore decided to analyse the success of RCTs, taking three angles (Bédécarrats, Guérin and Roubaud 2013; 2019; Bédécarrats, Guérin, Morvant-Roux, et al. 2019a, 2019b): developing theoretical critiques based on the classic internal and external validity questions (RCTs in theory: *doing the maths*); focusing the critique empirically: how RCTs are conducted on the ground (RCTs in practice: *doing the cooking*); and analysing the political economy of RCTs in terms of both supply and demand (RCTs as a business: *doing the accounts, both financial and symbolic*). Whereas the first point had been largely explored and our contribution marginal, the other two matters were relatively uncharted territory.[6] Our own analyses from an in-depth observation of two RCTs (microcredit in Morocco (Morvant-Roux et al. 2014) and micro-insurance in Cambodia (Quentin and Guérin 2013)) were largely borne out by an analysis of three of the most emblematic RCTs.[7] These RCTs ultimately proved highly debatable, whereas they had been largely instrumental in elevating RCTs to the status of gold standard.

3 European Development Network

4 French Agency for Development

5 We mean by this term those researchers defending the superiority of the method over all others. On the non-pejorative term of '*randomista'*, see the Ravallion (2020) and Ogden (2020). See also Gibson (2019).

6 Two main conclusions emerged from our analyses. First, although RCTs represent a suitable way to estimate the causal impact of a certain number of bounded projects, this is only true in ideal conditions defined in theory and rarely observed on the ground. And in these ideal conditions, RCTs may be able to be used to statistically quantify the impacts (significance and magnitude), but they cannot identify the mechanisms through which these impacts channel (paradoxical for a method that makes the analysis of causality its fundamental principle). Second, three of the major claims made by randomistas are groundless: i.e. that RCTs are superior to any other method; that the proliferation of RCTs can solve the external validity issue, acknowledged by all as an intrinsic weakness (which we have termed a 'hegemonic plan'); and that RCTs can provide all the answers when it comes to "what works and doesn't work in development".

7 The famous RCT associated with the conditional cash transfer programme in Mexico (Progresa, renamed Oportunidades and then Prospera), which many see as the catalyst for the rush on RCTs, and CCTs accordingly, but whose implementation and hence internal validity are disputed (Faulkner 2014); the equally high-profile RCT on intestinal worms in Kenya by Miguel & Kremer (2004), whose findings have been challenged by a group of epidemiologists (Aiken et al. 2015; Davey et al. 2015; Humphreys 2015), which is paradoxical given that the *randomistas* have made RCTs in medicine the movement's flagship; and lastly, an RCT on the recruitment and supervision of teachers in Kenya (Duflo et al. 2015), wherein Bold et al. (2013) have shown that scale-up by a national government-implemented policy produced none of the expected results.

Following this preliminary research, we pressed on in two parallel directions. We took forward our work on rural microcredit in Morocco by conducting a replication. The results of this replication not only corroborated the hypothesis of a contradiction between RCTs in theory and RCTs in the field, but revealed new facets of this discrepancy (Bédécarrats, Guérin, Morvant-Roux, et al. 2019a, 2019b). Expanded to a set of other RCTs on microcredit, this contradiction is the subject of one of the chapters in the book (Bédécarrats, Guérin, Roubaud, 2020c). Keen to deliberate the issue and prompt a scientific controversy, or at least discussion, we then launched this project to produce a co-authored book to throw open the question to other disciplines, voices and opinions, including much more positive views of the method than ours. Some will argue that the debate is tiring and jaded (Dimova 2019; see also Ogden, 2020). We believe, however, that is it vital, both scientifically and democratically speaking, for reasons detailed later in the book.

Bringing together some of the leading specialists in the field from a range of backgrounds and disciplines (economics, econometrics, mathematics, statistics, political economy, socioeconomics, anthropology, philosophy, global health, epidemiology and medicine), this edited volume discusses the main weaknesses of RCTs in the field of development, but also some of their unexpected strengths. The book takes concrete examples to explain how RCTs work, what they can achieve, why they sometimes fail, how they can be improved and why other methods are both useful and necessary. It reviews issues of method, epistemology, ethics, theory and ideology. What stands it apart from other critical views is its emphasis (among others) on the implementation of RCTs *on the ground*, outside of their ideal laboratory conditions. This reveals some of their unsuspected uses and effects, their political uses and ends, but also their disruptive potential. The book explores the implicit worldview that many RCTs draw on and disseminate. It probes the gap between the method's narrow scope and its success worldwide. Yet it also proposes areas for improvement and alternative methods. Without disputing the contribution of RCTs to scientific knowledge, the book warns against their so-called superiority and the potential dangers of their misuse. It also argues that the best use for RCTs is not necessarily that which immediately springs to mind and which RCT proponents promote: understanding certain behaviour rather than evaluating interventions.

Although the principle of RCTs in science is over a century old – their use in international development is called the fourth wave (Jamison 2017) – their large-scale use in developing countries is unprecedented (Ravallion, 2020). RCTs represent an indisputable advance for development economics. They offer a solution (among others) to the thorny question of attribution (how to isolate the effect of an intervention from all the changes that occurred at the same time). They place centre stage the issue of aid evaluation and the need for aid accountability. They lend new momentum to first-hand survey data collection by development economists. Last but not least, economic research

in the past sidelined Southern countries due to their lack of quality data, especially longitudinal data. The spread of RCTs has elevated economic research on these countries to world-class level. The new wave of RCTs in development can also be interpreted as methodological progress initiated in the South and transferred to the North (Bédécarrats, Guérin and Roubaud 2019).

Yet despite their limited scope of application (detailed below and throughout the book), RCTs are still held up by many as the evaluation *gold standard* against which all other approaches are to be gauged, and the award of the Sveriges Riksbank prize is likely to reinforce this supremacy. Presented by their disciples as a true Copernican revolution in development economics,[8] RCTs are often the only approach to be proclaimed 'rigorous' and even 'scientific' (See Ravallion, 2020). Some media-celebrity RCT advocates are looking to take RCTs well beyond their methodological scope in a move to establish the full list of good and bad development policies (Labrousse, 2020). The motive advanced for this upscaling ambition is to build up an ever-growing number of impact studies from which scalable lessons can be drawn. Clearly, though, there are a certain number of drawbacks to the proclaimed supremacy of RCTs in evaluation. These include disqualification and crowding out of alternative methods, ever-growing use of allocated resources, rent position and the legitimization of a specific and narrow vision of 'development' (what Lant Pritchett, 2020, calls 'kinky development'). They also include the disqualification of development projects and policies that do not adhere to the constraints demanded by the randomization protocols (Ravallion, 2020; Garchitorena et al., 2020; see also Adams 2016).

We are obviously not the first to express criticism. Many voices have been raised.[9] James Heckman and Angus Deaton's critical voices (Deaton, 2010 and 2020; Deaton and Cartwright, 2018;

8 "Just as randomized evaluations revolutionized medicine in the 20th century, they have the potential to revolutionize social policy during the 21st," (Duflo, Glennerster, and Kremer 2004: 29).

9 See for instance (Barrett & Carter 2010; Deaton 2010; Deaton & Cartwright 2018; Harrison 2011; Heckman 1991; Pritchett & Sandefur 2015; Rodrik 2009). Several edited volumes have also contributed to this discussion. The first, a book edited by Jessica Cohen and William Easterly (2010), sparked the nascent controversy. The book contained just one chapter focusing specifically on the subject with a gripping, albeit brief, controversy between Banerjee, Rodrik, Mulathain and Ravallion. The other chapters discussed mainly how to learn whether and which development policies work, but the question of RCTs ran implicitly throughout. The book by Tim Ogden (2017) is the most recent and RCTs are its central focus. It is structured in the form of 20 interviews with prominent players in the field. Fourteen of these players are active figures in the RCT movement and four others are more moderately involved in RCTs. There are just two critical voices (Angus Deaton and Lant Pritchett) who, although icons, make quite short contributions whose content is copiously reinterpreted and criticized by the other contributors. Thirdly, the book edited by Dawn Teele (2014) is more detailed and balanced. It makes a major contribution to our understanding of the subject, in particular with a comparison of RCTs conducted in the North and the South by political scientists and economists. Nevertheless, it still centres on methodological and epistemological considerations. Lastly, many contributions are repeats of now-dated articles published elsewhere in the 2000s, before the RCT industry really took off. Ten years on, then, this co-authored volume brings the previous books up to date, drawing on the most recent literature and taking a broader view in terms of both disciplinary angles and issues

Heckman, 1991 and 2020) carry particular weight, especially given that both have also been awarded the Sveriges Riksbank Prize in Economics (Deaton in 2015 and Heckman in 2000). This criticism is now more frequently acknowledged by RCT movement members (Ogden, 2020), but there has been no actual scientific controversy over the issue. For want of a real controversy (the most eminent *randomistas* we invited declined to take part), the book creates a dialogue between approaches, disciplines, different intervention sectors, and ultimately different standpoints on the role and potential of RCTs.

Some of the book's authors consider that the RCT craze is "madness" (Pritchett, 2020), that their superiority is essentially "a narrative" (Labrousse, 2020), and that they are "ineffective as tools of organization accountability and learning," and are not strictly speaking evaluations (Picciotto, 2020). Others consider that they have their place in the toolkit of evaluation methods, but that their self-styled superiority is "more a matter of faith than science," and that, in certain situations and for certain issues, observational studies are much more appropriate (Ravallion, 2020). This is also shown by the sector analyses of healthcare (Garchitorena et al., 2020), rural sanitation (Spears et al., 2020) and microcredit (Bédécarrats, Guérin and Roubaud, 2020c).

A more optimistic view suggests that RCTs have taken on board the criticism and that, in their present version, they offer real answers to a large number of development questions (Ogden, 2020). Another vantage point is that RCTs are useful not so much to 'evaluate' as to 'explore' behaviour using manipulations of price structures, contracts, teaching methods, and so on: researchers can make use of the disruption created by randomized protocols to observe in situ changes to interventions and behaviour, study their repercussions and draw operational conclusions from them (Morduch, 2020).

Others call for them to be improved as much from an ethical point of view, which remains a blind spot for survey protocols in development economics (Abramowicz and Szafarz, 2020), as from the point of view of causal explanation, either with respect to making better use of priors (Vivalt, 2020) or the phenomena of non-compliance as indicative of the preferences of targeted populations (Heckman, 2020).

The purpose of this paper, which reflects solely our own point of view, is not to reconcile the authors of the book and find a compromise, but to give a clearer picture of the issues involved in the debate. Following this introduction, the first section details the epistemological, political and ethical arguments behind the debate. The second section endeavours to define the development policies and projects that might lend themselves to the particularities of RCTs. The third section comes back to the idea of a scientific controversy, which we call for in earnest and which unfortunately has not

yet taken place, looking into the reasons for this no-show. Section four proposes ways of improving RCTs and methodological alternatives, while the conclusion is dedicated at assessing the potential role of RCTs to identify the adequate policies to combat poverty in time of COVID-19.

1. The arguments behind the debate: epistemological, political and ethical

We will not go into all the criticisms made of RCTs here – they are already listed in different chapters (Ravallion, 2020; Ogden, 2020; see also Bédécarrats, Guérin and Roubaud 2019). We think it more useful here to look over the epistemological, political and ethical differences underlying – often implicitly – many of the disagreements surrounding RCTs.

Far from being purely technical debates, the debates surrounding RCTs make reference to different – and often hard to reconcile – concepts of knowledge and learning. Is social science research into human interactions perceived scientistically (Putnam, 2009),[10] as the search for the ultimate, universal answer to a given problem, or as an ongoing learning process to find reasonable responses limited in time and space, mindful of the diversity of knowledge, including the knowledge of the development target populations? Do we see figures, statistical and econometric methods applied to social sciences solely as instruments and techniques, as the fruit of linear scientific progress? Or do we consider them also as a social and political construct built by somewhat arbitrary conventions, inextricably linked with a certain conception of state and public policies, the market, power and collective action (Desrosières, 2013), which fashion in part the world they seek to represent, understand and advise (MacKenzie, Muniesa and Siu, 2007)? This second meaning of knowledge does not deny scientific evidence, but advocates its embeddedness in particular social and political contexts. And it clearly differentiates scientific knowledge from policy decision-making, which implies referring to values in order to choose between different options and assess their social, economic and political consequences (Drèze 2019).

The opposing views surrounding RCTs are also based on different notions of development, poverty and, more broadly, politics, seen as a conception of the world in which we live and which we endeavour to attain. Is the world an aggregate of individuals seeking to be independent or is it a complex system made up of dialectics, multiple interactions, retroactions and systemic effects between social beings who are interdependent and wish to remain so? Should we see the "causes of poverty as a lack or want of relevant variables or as an active process of impoverishment or

10 By scientism, we are referring to the idea that experimental science is the only reliable source of knowledge on the world and that it is the best means by which to organize humanity to solve all its more pressing problems. Experimentation allegedly does without the need for metaphysical, philosophical, ethical and aesthetic reasoning.

perpetuation of poverty" (Shaffer 2015: 154)? A "want-based" understanding of the causation of poverty calls for policies of "difference-making" wants (to cope with deficits in health, education, nutrition, water/sanitation, credit, and so forth); and understanding the impacts of such policies requires a counterfactual to be able to isolate the difference and attribute the impact to the policy in question. By contrast, a conception of the causation of poverty in terms of processes and social relations calls for macroeconomic and structural policies (exchange rate, capital control policies, social protection measures, and so forth); and understanding the impact of these measures requires a "mechanism-based approach" that explores the diversity and complexity of the causal processes that generate the impact (Shaffer 2015).

Finally, these divergent visions find expression in divergent versions of the economists' role. Is their role to 'fix' the world and concentrate on the practical details of policy implementation (Duflo 2017), like a plumber or engineer repairing cracked pipes? Or should economists keep a critical distance from the workings of the present system, even going so far as to radically challenge it?

These different epistemological positions (in the form of a continuum more than a binary opposition) permeate the debates on RCTs and can be seen in a string of opposites running through the chapters of the book: macro versus micro, public goods versus private goods, horizontal versus vertical health interventions, public action versus social marketing, structure versus behaviour, attribution versus processes, and so on.

1.1 The epistemology of RCTs in the field of development

In theory, *randomistas* see experimentation precisely as an antidote to preconceived ideas (see also Rodrik 2009). This pragmatism may well give the impression of being a rejection of scientism. Yet laying claim to the method's superiority clearly reflects a scientistic concept of science (Picciotto, 2020). This scientism can be seen at work in two ways. First of all, the *randomistas* purport to provide universal answers for a large number of development interventions. In response to the question of contextual particularities, some *randomistas* like Esther Duflo argue that they should be considered as 'global public goods' and an international body established to scale them up (Savedoff *et al.,* 2006; Glennerster, 2012). This body would then build a universal database and act as a 'clearing house", providing answers as to "what works and doesn't work" in development (Banerjee & He 2008; Duflo & Kremer 2005). Yet this hegemonic plan (Bédécarrats, Guérin and Roubaud 2019) does not solve the question of heterogeneity, whether of intervention practices or contexts (see, in particular, Spears et al., 2020).

Secondly, this scientism is seen at work in overconfidence in the technique, with something of an obsession with the *theoretical* protocol, supposed to guarantee sample balance and therefore settle the attribution question. The *implementation* of the protocol on the ground is secondary. As with all research – particularly RCTs considering the budgets concerned, the size of the samples, the constraints for comparison between control and treatment groups, and the risks of contamination – the implementation of the protocols necessarily deviates from what is planned in theory and calls for tweaking, accommodations and compromise.[11] In many cases, the collection of RCT data violates the assumptions of the statistical theorems used for inference. NGOs and governments working in development know only too well that interventions in the field never go according to plan (Mosse 2004; de Sardan 1995). Why should experiments be any different? As shown by the different chapters in the book, deviations between protocol and implementation can be observed all the way down the knowledge production line:

- In sample building with, here, three types of difficulties. The first difficulty is multiple biases between treatment and control groups (Ravallion, 2020). This results in a focus on highly specific populations, although this particularity is not made clear by the randomistas (see, for example, Bédécarrats et al. (2019a) and Wydick (2016) on microcredit; see also Barrett & Carter (2014: 75)). The second difficulty is insufficient take-up and consequently an insufficient difference in exposure to the intervention. This weakens the ability to draw conclusions due to a lack of statistical power, a problem that would require unrealistic sample sizes and therefore unrealistic budgets to resolve 2020). Insufficient take-up can also cause the intervention to be artificially transformed (see the following point). Lastly, the 'virginity' of the control zones, an often necessary condition for comparison, proves particularly complex and raises ethical and feasibility problems (Bédécarrats, Guérin and Roubaud 2019).
- In the type of intervention, whose implementation may turn out to be very different to the 'real world', as shown by Garchitorena et al. (2020) in health, or which may even be artificially transformed to encourage more take-up (Bédécarrats, Guérin and Roubaud, 2020c).
- In data collection, since the priority placed on econometric considerations can get in the way of statistical considerations. Statistics is not only the science of numbers: it is first and

11 Our replication experiment shows the difficulty some *randomistas* have acknowledging the practical difficulties of conducting an ideal RCT, the like of which does not actually exist (Bédécarrats, Guérin and Roubaud, 2020).

foremost a science of data collection, which requires multiple techniques to guarantee the collection of quality data (Bédécarrats, Guérin and Roubaud, 2020c).

- In the interpretation of the results which, far from being restricted to a comparison of averages, as claimed by the randomistas, actually implies a range of implicit hypotheses and an art of rhetoric, whose persuasive power is particularly manifest (Labrousse, 2020).

All in all, method implementation constraints can force researchers to concentrate on midpoint indicators, short timeframes, and specific populations or geographic areas and, in so doing, to restrict themselves to a very narrow set of questions or produce unusable results (Garchitorena et al. 2020; Spears et al. 2020; Bédécarrats et al. 2020 on different sectors give numerous examples of this). The disproportionate importance placed on the theoretical purity of the protocols and demonstration of causality at the expense of protocol *feasibility* and data *quality* is a major (albeit often implicit) sticking point in disagreements over the hierarchy of methods.

From our point of view, giving precedence to the method over the research questions is tantamount to "hunting for the lost keys under the streetlight". In a way, and to paraphrase the title of a book on development aid (Jean-David 1999), it is like finding problems (projects to evaluate) to the solution (RCTs).

1.2 RCTs and 'development'

As suggested by Lant Pritchett (2020), the success of RCTs is merely the symptom of a more serious disease: the abandonment by part of the international aid community of large-scale transformative development policies (national, international and even regional), including seeking to transform the socioeconomic systems[12]. Reviewing transformations in the field of aid is therefore useful to better understand the attraction of RCTs and their scope of application. The contrast between the narrow scope of RCTs and their scientific, media and political success is down to both supply and demand. On the supply side, we have shown elsewhere that the *randomistas* have produced an entirely new scientific business model, of which J-Pal is the most emblematic and accomplished example, and which combines the mutually reinforcing qualities of academic excellence (scientific credibility), public appeal (media visibility and public credibility), donor appeal (solvent demand), massive investment in training (skilled supply) and a high-performance business model (financial profitability) (Bédécarrats, Guérin and Roubaud 2019). As effective as these

12 The prize jury, in its press release, acknowledges this: "This year's Laureates have introduced a new approach to obtaining reliable answers about the best ways to fight global poverty. In brief, it involves dividing this issue into smaller, more manageable, questions".

strategies may be, they nevertheless assume that there is a *demand*. Some methods, theories and technologies succeed, not because of their scientific superiority, but because they manage to "sustainably galvanize and rally players and interests prepared to produce and use [the technologies in question]" (Callon 2006a: 155).

RCTs benefit here from a particularly RCT-friendly environment, which they nurture in return. They most probably would not have had the same success in a different age. The academic climate first of all, especially in economics, is conducive to the rise of RCTs: demise of the heterodox schools concentrating on social structures and domination processes, search for the micro-foundations of macroeconomics, and primacy of quantification and economics in the social sciences. The joint rise of behavioural and experimental economics, crowned by the 2002 award of the Sveriges Riksbank Prize in Economics to psychologist Daniel Kahneman and economist Vernon Smith, respective experts in the two fields, and then to economist Richard Thaler in 2017, shows just how far the discipline has come. RCTs draw extensively on the precepts of behavioural economics and are actually the vehicle that channelled behavioural economics into development economics to the extent that it now occupies a dominant position in the discipline (Fine et al. 2016).

It is also from transmutations in the aid field that demand has emerged for RCTs. With the end of the Cold War, the political sphere started to ease its grip on official development assistance (ODA). Cold War technical and financial cooperation was often merely another pawn in bloc rivalry. As the Berlin Wall fell, so too did cooperation's subordination to real politik. In the new post-modernist world, ODA promoters have found themselves under the spotlight as the aid crisis, MDGs and *New Public Management* have summoned them to the stand to prove their utility (Naudet 2006).

The new credo focuses development policy on poverty reduction and promotes results-based management. These guidelines were formulated in the 2005 Paris Declaration on Aid Effectiveness and thereafter systematically reiterated by the major international conferences on official development assistance in Accra in 2008, Busan in 2011 and Addis Ababa in 2015. The rise of the evidence-based policy paradigm, which consists of basing all public decisions on scientific evidence, has given scientists new credibility in these political arenas. RCTs in principle meet all the conditions required by this game change: agnostic empiricism, apparent simplicity (simple comparison of averages), elegant use of mathematical theory (guarantee of scientificity) and focus on the poor (compassionate mobilization and moral commitment; Labrousse, 2020). Their (apparent) simplicity makes them easy for policymakers to understand, lending them appeal as a vehicle for informing public decision-making. The evaluation of the *Progresa* programme in Mexico formed a

prototype for this method and a textbook example of its performance capabilities (Bédécarrats, Guérin and Roubaud 2019).[13]

The aid crisis is also a crisis of *official* development assistance. As ODA funding efforts lose speed, private investment and international remittances are taking up the slack (IFC 2017). Governments are now merely one body among others in a 'coalition of players' that includes businesses, NGOs and, more broadly, 'civil society', foundations and research institutes. Foundations taking up the philanthrocapitalism of the industrial period are playing a growing role, mainly in the health sector, but also in technological innovation, now cross-cutting most, if not all development sectors. These new players and funders are changing the aid *tools*. Not only does the withdrawal of the state as planner and developer lead to "thinking small" (Cohen and Easterly 2010), but when combined with the resurgence of philanthropy, it paves the way for development that juxtaposes privatization (of interventions and players), marketization (of the goods and services delivered) and also compassion.

By setting up the poor as barefoot entrepreneurs, microcredit with its promise of a double bottom line – poverty reduction with profitability or at least financial sustainability – was a pioneer in marketization. This marketization subsequently expanded under the name of BoP (bottom of the pyramid) in a low-cost repeat of trickle-down theory (with the idea that consumption by the poor will eventually form a factor for growth and redistribution (Elyachar 2012).

This economic reason combines with a 'humanitarian' reason (Fassin 2010). A moral duty to act is emerging in the face of public infrastructures seen as moribund, derelict or utopian and the resulting suffering and needs they create. Driven by a sense of compassion and urgency, financiers and practitioners – but also researchers – are joining forces to design and test an entire array of micro-scale interventions: these 'humanitarian goods', to use the expression coined by Redfield (2012), try as best they can to solve, ad hoc and temporarily, what are considered to be the most urgent and crying needs. These humanitarian goods aim to make good government failings, and they embrace this dual compassionate and economic line, even though the economic strand does not rule out redistributive measures (see below).

13 It is, however, enlightening to note that this programme was a powerful tool for social and political control, consumed by nepotism and corruption (Crucifix & Morvant-Roux 2018; Kidd 2019). Moreover, the abovementioned blind spots in its experimental evaluation, especially in terms of internal validity (Faulkner 2014), were precisely the arguments used by the new Mexican government to announce its withdrawal in early 2019 (Encisco 2019).

In this new configuration, and although public financing of large infrastructure continues to account for a large share of international aid, governments' decision-making and planning powers are gradually seeping away towards vertical funds,[14] foundations[15], private companies[16] and new financial mechanisms such as social impact bonds. The foundations, a fast-growing emerging player, are set to play an increasingly important role (see also Pritchett, 2020). Just as the Ford Foundation supported the rise of experiments in the United States in the 1960s, so too are numerous foundations today playing a driving role in the expansion of RCTs in development (starting with the establishment of J-Pal; Jatteau 2016: 230). The very principle of social impact bonds, in which repayment to investors is conditional upon specified social outcomes being achieved, favours a similar trend. Lastly, in this development privatization process (privatization of interventions and players alike), NGOs occupy a choice position as implementing partners.

Far from the reforming and sometimes idealistic aims of previous generations of development players, private, market and humanitarian goods have the merit of being realistic and concrete and offering a pragmatic solution for needs seen as urgent. Their implementation is not above criticism – probably the most well-known are the debates on therapeutic food as an unfair trade practice impacting on local agricultural systems. Yet from the point of view of their purpose – to solve a temporary, individual problem – they work (Redfield 2012). Now as a number of chapters point out, and we will come back to this later, it is precisely these types of goods, due to their individual targeting and short-term nature, that lend themselves the best to the constraints of randomized trials. Likewise, NGOs remain the choice implementing partners for *randomistas*, because they are more flexible, less bureaucratic, more open to innovation and more reliable than governments (Webber and Prouse 2018; Cohen and Easterly 2010). *Randomistas* express a will to work more with governments (Banerjee 2013), but are finding it hard to deliver on this will (Pritchett, 2020).

The transformation of the development field came well before RCTs, and it would be undue to say that they were responsible for it (Morduch, 2020), even though the crowding-out effects are to be taken seriously (see below). Yet should these changes be unreservedly condemned by an indictment of the abandonment of any true prospect for reform, the unsustainability of individual ad-hoc interventions, and the illegitimacy of private players who are not democratically accountable? Or

14 Such as the Global Fund and the GAVI Alliance (Global Alliance for Vaccines and Immunization) in health.
15 The Bill & Melinda Gates foundation remains the leader in many health sub-sectors, but also in everything involving new technologies. Banking foundations such as Citi and Mastercard are high-profile players in financial inclusion.
16 Such as Nutriset for therapeutic food to treat malnutrition and Vestergaard Frandsen for water filters, tsetse fly screens and insecticidal bednets.

should we learn to live with them rationally, considering that even though the pipes were poorly designed to begin with or are at breaking point, to use the plumber/engineer metaphor, it is still worth repairing the leaks? The answer to this (rarely spelled out) question explains many of the disagreements surrounding RCTs, as well as the different positions found in the book.

1.3 Ethics and RCTs

The ethical issue is a recurring one with RCTs, not only in the development field, but in general (especially in medicine). Although everyone agrees on the need to tackle this question head on, at least in principle, these caveats have not yet seen action (Abramowicz and Szafarz, 2020; and also Ravallion, 2020; Ogden, 2020; Picciotto, 2020). Among the *randomistas*, this acknowledgement remains marginal,[17] as if faith in the scientific advances that RCTs can bring - and their automatic policy and welfare improvement repercussions - were sufficient to exempt researchers from ethical consideration. Whereas all research entails ethical issues, RCTs are more concerned than observational studies by reason of their very principle (Teele 2014), since they typically feature a form of manipulation of the research environment (they "twist the lion's tail", to quote the expression used by Deaton and Cartwright (2018: 18)).

Neither are critical analyses free of this neglect of ethical considerations, since they often merely mention the issue with barely any details. The chapter by Michel Abramowicz and Ariane Szafarz (2020) is an exception in that it probes the implications of the principle of *equipoise*, i.e. the ethical requirement for an experiment involving human subjects to display "a state of genuine uncertainty on the part of the clinical investigator regarding the comparative therapeutic merits of each arm in a trial" (Freedman 1987, p. 141, quoted by Abramowicz and Szafarz, 2020). The authors ask why economics experimenters are virtually systematically ignorant of this principle, when it is an essential pillar in medical science. They provide a series of pointers to address the question. Ravallion (2020) also addresses this subject, insisting on the importance of properly assessing the risks and information already available, and showing that the principle of equipoise takes different forms depending on the different cases and types of randomization (inevitable treatment rationing, conditional randomization and equivalence trials). He also discusses the *adaptive experiment*

17 For example, none of the 22 pages on "Concerns about experiments" by Banerjee and Duflo (2014) addresses the ethical issue, except to say in response to randomization not being a fair way to allocate the programme (seen as a methodological problem, but not an ethical issue) that, "… implementers may find the easiest way to present it to the community is to say that the expansion of the programme is planned for the control areas in the future." (p.101)

proposition put forward by Narita (2018) to establish a Pareto balance between the possible positive and negative effects on participants based on available knowledge.

This virtual denial of ethical considerations by *randomistas* is all the more questionable in that various standards of best practices do exist, as much for medical RCTs as for most of the social science RCTs conducted in the North. The ethical principles designed to govern randomized trials on human subjects have been codified into recognized standards, in particular the Declaration of Helsinki in 1967 (WMA General Assembly 2014 (9th edition)); the Belmont Report in 1974[18] and the International Ethical Guidelines of Council for International Organizations and Medical Sciences (2002). These standards prescribe clear principles: informed consent, the do no harm principle, provision of specifically considered protection for vulnerable populations, risk analysis and responsive monitoring, to name but a few.

These fundamentals are rarely respected in the development field (Abramowicz and Szafarz, 2020). Like Barrett & Carter (2014), we detail four examples of RCTs that illustrate the ill effects of this ethical negligence. The first example was designed to demonstrate the mechanisms of corruption in the case of obtaining a driver's licence in India (Bertrand et al. 2010). One of the arms of the treatment was to offer a bonus to candidates for obtaining a licence. Barrett and Carter show that this RCT violated the ethical code of 'do no harm' (they even speak of "irresponsible research design") in two ways: not only did the treatment encourage corruption, but it also imperilled the lives of others by putting potentially reckless drivers on the roads, since the experiment showed that the treated group took fewer driving lessons. The second example concerns an RCT set up in Kenya to test the Rockefeller Effect (which states that too many resources do more harm than good) by means of a project providing assistance to groups of women (Gugerty & Kremer 2008). The project's effects proved to be negative (the poorer women were excluded from the positions of power), confirming the Rockefeller hypothesis. The problem is that the RCT harmed the experiment's subjects, when this harm might have been predicted, at least as a possibility, and the women should have been informed of that possibility for them to decide whether or not to take part (principle of informed consent). The third case is an RCT on secondary school pupils in the Dominican Republic to test whether information concerning higher labour market returns to education than would normally have been expected by the pupils could prompt them to stay longer in education (Jensen 2010). The ethical problem here is that the information given the secondary school pupils (estimated from observational data) – being both overestimated (given the endogeneity biases) and calculated on average without taking into account the characteristics of the pupils and schools – is likely to have

18 National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979).

led some of the pupils, probably the poorest, to "overinvest" in education on the basis of the expected return. And that is not to mention the effect of the increase in the supply of graduates likely to depress future returns (general equilibrium effect). Last but not least, the treatment in the fourth example consisted of granting credit to individuals rejected by a microcredit provider, since its scoring model predicted a high probability of default on payment (Karlan & Zinman 2009).[19] Quite aside from the fact that this strategy placed the treated group at risk of being incapable of reimbursing the loans (with the associated penalties) and in a potential situation of over-indebtedness, not having informed them of such is in breach of the principle of informed consent. This is a difficult dilemma to resolve since, if they had been informed of the risk, their behaviour would probably have changed and hence undermined the internal validity of the RCT. None of these flaws prevented these four RCTs from being published in leading academic journals, and this also raises questions about the role of economic journals in failing to respect ethical standards (Abramowicz and Szafarz, 2020).

Other examples are mentioned in the book (Ogden, 2020). The proliferation of RCTs, especially by less visible and hence even less ethically controlled institutions, could end up undermining the basic principles. Illustrating this point is the case of an RCT in progress. This donor-commissioned RCT was set up to test how information affects migratory behaviour in rural Mali. Participants were shown a short film chosen from among four randomly allocated films illustrating different outcomes of migration and non-migration (to Europe): successful migration; suffering, ill-treatment and ultimately a failed migration attempt; successful non-migration; and a comedy having nothing to do with migration serving as a placebo. Quite aside from the problems of informing participants of the implications of such a test and obtaining their informed consent, none of the treatment arms can be deemed beneficial to the participants (violation of the beneficence principle). Individuals' preferences can simply be changed based on what the RCT's commissioners consider to be good for them (or for themselves). Neither is the 'do no harm' principle respected as, following the films, some participants might decide to migrate and die in the Mediterranean or be tortured in Libyan jails. Lastly, the commissioner's political motive seems obvious (to curb African migration to Europe). It seems that this RCT was designed without any ethical consideration.

Considering the multitude of examples, it would appear that the creation of Institutional Review Boards in many academic institutions has done nothing to remedy the observed ethical lacunas. Two mutually reinforcing reasons can be put forward for this. The first is the difficulty of

19 A similar approach consisting of including subjects initially judged insolvent is also included in Augsburg et al. (2015) and discussed by Bédécarrats Guérin and Roubaud, 2020.

simultaneously guaranteeing the protection of the experiment's subjects and the internal validity of the protocol. The second is the *randomistas* ' flawed understanding of and manifest lack of interest in the subject. When faced with what could be called an ethical dilemma, they all too often come down on the side of the methodological imperative. Yet ethical safeguards are all the more necessary in the Southern countries. Firstly, not informing participants (informed consent principle), if not deliberately misinforming human subjects to ensure a clean identification strategy, is at odds with the principle of ownership promoted by the development policies. Secondly, participants are generally vulnerable individuals, both economically (poor) and politically (voiceless), on whom it is easier to impose the trial, if not deliberately mislead. This asymmetry is especially strong in that the surveys are more often than not tantamount to life-size laboratory games supervised by young students and research assistants from Northern universities. We also need to look into the choice of these populations, especially when testing a behavioural hypothesis or a theory put forward by certain RCT proponents (Banerjee & Duflo 2011; see Murdoch, 2020). Save advancing that the poor in Southern countries have specific rationality, the arguments of lower cost and less capacity to refuse to take part (a recurring problem with RCTs in Northern countries) due to a lack of knowledge of their rights and lopsided balances of power (including with respect to the experimenters) appear to be credible explanations (Teele 2014), as has already been observed in the "offshoring" of medical clinical trials (Petryna 2007).

The *randomistas'* ethical argument is the long-term improvement of the well-being of populations by means of scientific progress made possible by RCTs. Yet this is an assumption that is far from proven (Ravallion, 2020). All in all, then, in addition to the unassailable faith in the theory of the technique at the expense of its feasibility (as seen in section 1.1), it seems that all too often, a hardly acceptable hierarchy of values prioritizes scientific findings over the well-being of the populations.

2. What is the scope of application for RCTs?

After closely examining the many limitations of RCTs, in terms of both internal and external validity, Deaton and Cartwright (2018) suggest that RCTs nonetheless remain valid in two areas: 1) to test a theory, and 2) for a specific evaluation in a given context of a particular project or policy, provided that the potential internal validity problems have been solved and with the caveat that the explanation of the results obtained is often inadequate. The chapters in the book confirm and expand on this analysis. Randomized evaluations are only possible for a highly restricted field of interventions, more often than not concerning private, market and humanitarian goods. RCTs can also be used to test economic theory regarding behavioural responses to interventions, challenging

certain preconceived ideas. Ultimately, however, they answer neither the question of *impact*, such as it has long been defined in the development aid field, nor the question of the *explanation* for the measured effects.

2.1 Private, market and humanitarian goods

The conditions required by the randomized methods' protocols restrict them to a narrow spectrum that Bernard, Delarue and Naudet (2012) call "tunnel-type" programmes. These programmes are typified by short-term impacts, clearly identified, easily measurable inputs and outputs, and unidirectional (A causes B) linear causal links, and are not subject to the risks of low uptake by targeted populations. They tie in with the suggestions made by Woolcock (2013) that projects subjected to randomization need to exhibit "low causal density", require low implementation capability and feature predictable outcomes.

This type of method is therefore applicable only to simple or local short-term interventions targeting individuals. In concrete terms, these micro-interventions concern essentially private goods and services, i.e. rival and excludable (see Ravallion, 2020; Pritchett, 2020 and Picciotto, 2020).

In health, they concern actions to prevent and treat individual diseases. They also come in the form of water filters, mosquito nets, training and bonus systems for health professionals, free consultations, medical advice by text message, and micro-insurance. However, RCTs do not answer the question of the management of the health systems, which are necessarily complex and systemic, involving skilled, motivated manpower, an infrastructure, the provision of medicines, etc. (Garchitorena et al., 2020). In sanitation, these micro-interventions concern the distribution, construction and use of latrines. Here again, RCTs do not answer the question of the management of human waste flows using which type of sanitation or cleaning network, which type of infrastructure and which type of regulation (Spears et al., 2020). In poverty reduction, these micro-interventions are microcredit, savings, entrepreneurship training and financial education services. Once again, RCTs do not answer the question of regional or sectoral wealth creation processes (Bédécarrats, Guérin and Roubaud, 2020c) or the broader question of access to basic services (Pritchett, 2020).

Contrary to certain critical analyses (see, for example, Berndt 2015), RCT conclusions do not necessarily advocate the marketization of the private goods (which equates RCTs more with the abovementioned humanitarian camp). In the case of highly price-elastic insecticidal bednets and deworming treatment, RCTs have put the case precisely for their free distribution, considered to be more effective than billing and hence challenging popular belief in the health field. In the case of microcredit, RCTs have concluded that the poverty reduction impact remains marginal and that

poverty reduction therefore calls for other types of intervention (Banerjee, Karlan and Zinman 2015). Again in the case of microcredit, RCTs have shown that the poor are sensitive to interest rates, here too toppling the widely held idea that access is more important than cost, a popular belief among microfinance organizations and their funders held up to legitimize high interest rates (Morduch, 2020).

Although these findings can be useful, the subjects addressed remain limited compared with the host of development, poverty and inequalities issues. The conditions required to implement RCTs therefore rule out a huge number of development policies involving combinations of socioeconomic mechanisms and feedback loops (emulation effects, recipient learning effects, programme quality improvement effects, general equilibrium effects, etc.). This is precisely the case with public goods (Ravallion, 2020). Where interventions involve infrastructures and regulatory systems, experimental manipulation is impossible (Spears et al., 2020).

In the terms of reference for a study commissioned on the subject, a group of DFID managers estimated that less than 5% of development interventions are suitable for RCTs (DFID, 2012). Although this figure is not to be taken literally, there is no doubt that experimental methods are not suitable to evaluate the impacts of the vast majority of development policies. In their more formalized paper, Pritchett & Sandefur (2013) come to a similar conclusion.[20] In this volume, Garchitorena et al. (2020) point out that 97% of funding for health research worldwide is earmarked for the development of new technologies (mainly of the pharmaceutical variety), and that only the 3% left over goes into research on *implementation*, albeit essential to understanding and improving health system dysfunctions.

2.2 Evaluate impact or test behaviour?

As suggested by Jonathan Morduch (2020), RCTs actually pursue two aims: to measure impact and to explore, "the nature of economic contracts, behaviors, and institutions." He goes on to submit that it is ultimately this second less-debated type of "exploratory RCT" that is the most promising, representing a real gain over other methods and therefore greater potential in terms of expanding knowledge.

This second type of RCT shifts the focus from measuring the impact of interventions representative of public action or development aid to testing different modes of a given intervention

20 "The scope of application of the 'planning with rigorous evidence' approach to development is vanishingly small" (Pritchett & Sandefur 2013: 1).

and measuring the outcomes in terms of intervention take-up. This type of RCT, says Morduch, is a source of information, if not 'provocation', in challenging certain misconceptions in development economics (such as the abovementioned low price elasticity of demand for microcredit) and testing innovations and how behaviour reacts to those innovations. For example, it can test different crop insurance selling timeframes for a better understanding of the constraints of time and liquidity; or test the role of information and assistance in the use of mobile telephones by the ultra-poor for a better understanding of intra-household sharing mechanisms.

These purposes are useful and laudable (provided the ethical and internal validity criteria are met and the conclusions are valid), but the question could be asked as to why the *randomistas* persist in talking about impact when a large number of RCTs are actually more "exploratory" in nature and compare different modes of one and the same intervention, often merely measuring the take-up differentials. The sector analysis of sanitation comes to a similar conclusion: RCTs appear to be more suited to analysing behavioural changes than measuring impact per se (Spears et al., 2020).

In fact, the question of impact often remains unanswered. Since 1992, most development aid sector players have relied on five criteria defined by the OECD Development Assistance Committee (2002), among which is found an impact criterion: "Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended." Yet RCTs can only evaluate the *short-term impact of short causal chains*: this is not then strictly speaking an impact as defined above (see also Picciotto, 2020). Taking the example of insecticide-treated bednets, often seen as the jewel in the crown of RCTs (Ogden, 2020), the question usually asked by RCTs concerns *take-up* rather than impact, since insecticide-treated bednets are considered to be essentially 'good'. Yet their medium- and long-term effects are controversial due to genetic adaptation by mosquitoes and the destruction of local production systems (Beisel 2015). Omitting long-term and collateral effects is just as problematic in the microcredit sector (Bédécarrats, Guérin and Roubaud, 2020c).

This type of RCT is ultimately redolent of the notion of 'social marketing', a term very much in vogue in development circles, which quite naturally complements the above-described circulation of private, market and humanitarian goods and behaviourist trend. Social marketing is the application of commercial marketing tools and principles to the design, implementation and evaluation of behaviour change programs in pursuit of individual benefits and the public interest (French et al. 2010). Modelled on behavioural science, social marketing techniques include nudges,[21]

21 A nudge refers in behavioural economics to a small and cost-effective device that does not imply a formal obligation or prohibition intended to influence people's behaviour in a predictable way.

but also more classic marketing methods (packaging, price, identification of the most suitable distribution channels and places, etc.). Social marketing originated in the 1970s in the health and social fields, including in the South, and in areas such as reproductive health, AIDS prevention, rehydration therapy for diarrhoea, and sanitation), before expanding to target behavioural change in a large number of sectors (environment, agriculture, education, financial management, consumption, etc.).

2.3 Measuring versus explaining

RCTs might be able to measure and test some intervention impacts and aspects, but they cannot analyse either their *mechanisms* or their underlying *processes*. In a 'want-based' analysis of the causation of poverty, as found in randomized approaches, the question of processes and mechanisms is set aside (Shaffer 2015). Overcoming this limitation of the probabilistic theory of causality would call for a "causal model" (Cartwright 2010), a coherent theory of change (Woolcock, 2013), a structural approach (Acemoglu 2010) and evaluation of the intervention in context (Ravallion, 2009, 2020; Pritchett & Sandefur 2015).

In the face of this criticism, *randomistas* are now grounding their results in explicit theories of change (Ogden, 2020), based largely on behavioural economics. Behavioural economics is useful to disentangle the complexity of the psychological and cognitive processes, individuals' internal struggles, and the multitude of their 'mental accounting' practices (Thaler 2015), and to explore and test how behaviour reacts to such or such an intervention (see above). However, behavioural economics cannot capture the complexity of atypical, unexpected and 'sub-optimal' behaviour, wherein this latter term moreover presumes what is a debatable normative dimension. Two levels need to be differentiated here: the level of individual behaviour, which does not always fit in the behavioural boxes (Servet 2018), and the level of the interventions, which rarely go according to plan.

With respect to individual behaviour, people are social, plural beings who cannot be reduced to mere target populations. People's agency is not limited to their refusal or take-up. Neither is it limited to their cognitive or social 'biases'. Local rationalities and motivations are constructed; they develop from and reflect social and political norms and realities. They tie in with pre-existing forms of interdependences, balances of power, and social and political structures, but also with desires and aspirations. The fact that the social is so unpredictable does not mean that it should be seen solely as an obstacle and a constraint, that it should be eliminated by dint of nudges. Local populations sometimes have good reasons to act the way they do, especially when the global environment does

not change. People have their own conceptions and representations of the world (and their own theories of change) and their own knowledge and know-how regarding care, illness and well-being, cleanliness and dirtiness, finance, poverty and wealth, and so on. Although some of these representations are sources of discrimination, the fact remains that they shape behaviour. Some of these representations also reflect specific worldviews, which are not necessarily less 'optimal' than the researchers' own (see the example of microcredit with Bédécarrats, Guérin and Roubaud, 2020c).

Interventions are also complex, combining multiple levels and players. Local realities shape, frame, constrain and influence the intervention (Mosse 2004; Olivier de Sardan 1995). Such is the case with the three sectors represented in the book. In global health, for example, "One of the most important questions […] is why known technologies – those that are proven to work in certain settings – systematically fail to reach the people for whom they are intended," (Garchitorena et al., 2020). Answering these questions necessarily calls for a focus on the workings of 'systems': local health systems, organizations' systems, the particularity of interactions between 'target' populations and healthcare providers, etc. Likewise, sanitation and microcredit do not mean the same thing to different people and cover countless realities, methods and forms of implementation. This diversity narrowly restricts the potential of RCTs in their generalization endeavours (Spears et al., 2020; Bédécarrats, Roubaud and Guérin, 2020c). And this complexity and diversity are probably not limited to these three sectors.

3. Why is a scientific controversy needed and why has it not taken place?

As we have seen, far from being a unanimously accepted gold standard, RCTs are a subject of debate and much criticism. This should have prompted a scientific controversy. But there has not (yet) been any such controversy, when it is vital for scientific progress and democratic debate. What has gone wrong? Without claiming to be exhaustive on the subject, and in view of its importance, we felt it useful to outline a few avenues for analysis borrowing from science studies.

Bear in mind, first of all, that contrary to a naïve view of science, scientific progress is not always a rational, linear process wherein the most effective and useful methods and findings systematically prevail over the others and a consensus surrounds certified knowledge. Scientific knowledge is also a historical, social and political product forged by advances and setbacks, cycles, debates and disagreements, which sometimes turn into *controversies*, defined as differences between two parties brought before and debated on a public stage.

A scientific controversy is therefore not to be understood in a negative light, as the symptom of errors of reasoning (where the 'true' will ultimately prevail over the 'false') or of untoward

interference by politics or interests other than the advancement of knowledge (an area supposed to be free of all subjectivity). The controversy is inherent in the collective production of knowledge. It is often the controversy that enables the emergence of key scientific progress. All scientific fields are marked by major controversies that are sometimes violent (water memory, GMOs, the 'Eldorado scandal' and gravitational waves), but are also sometimes nipped in the bud (Callon 2006a).

A controversy may be defined as a difference between two conflicting positions taking as witness an audience made up of scientific peers or a broader public (Lemieux 2017). The conflicting positions taken are sometimes virulent, but participants are bound to respect the conventions of the academic world such as the principle of equality between participants, the importance of logical reasoning, control of aggression, and respect for the principle of the dignity of the protagonists. However, these conventions remain vague, and accusing an opponent of abusing a dominant position or overstepping the limits of civility is often a way of shifting the balance of power or disqualifying the rival.

As in many areas of sociology, schools differ as to how to approach controversies depending on whether they give precedence to logic and evidence (Raynaud 2018) or whether they concentrate on the beliefs, social conventions and balances of power that affect the content of the arguments and arbitration between rival rationales (Akrich et al. 2013). In any case, controversies are seen by *science studies* as the reflection of a social and historical reality. Disputing processes reveal balances of power, institutional positions and social networks. They drive this social world forward by altering the balances of power, redistributing prestige and resources, and producing new conventions that will constrain future actions and positions (Lemieux 2007).

Coming back to our question – why the controversy has not taken place – the conceptual interpretations developed by Callon provide some insight (Callon 2006b, 2006a). First, the line between what warrants and what does not warrant a controversy is always the subject of agreements negotiated in the disputing processes. In our case, the professional community of development economists gives precedence to what it considers to be fundamental research, in particular the statistical purity of randomized trials and command of causal identification biases. This aspect takes the priority here over considerations, seen as secondary by this professional community, that have more to do with an applied dimension, implying acknowledgement of the different 'tricks of the trade', tactics and tweaking required to put the method into practice, and also acknowledgement of the agency of the experimenters and the trial's subjects and the sets of players it produces (Bédécarrats, Guérin, Morvant-Roux, et al. 2019b; Kabeer 2019). This brings us back to the epistemological differences discussed above. The advent of a controversy then implies setting up

sufficiently structured forums for sustained discussions to take place. In the absence of such arenas, confusion reigns among the protagonists as to who is speaking and in what context: the same players can uphold one-sided narratives in some forums and, without ever withdrawing them, make much more balanced and cautious statements in expert arenas.

3.1 Avoiding the controversy, but listening and adapting

The absence of public dialogue does not prevent the *randomistas* from adapting their methods and practices (Ogden, 2020), even though responses vary by groups of researchers. Some make their data available, thereby encouraging replications. Some acknowledge the legitimacy of methodological pluralism and combine RCTs with other methods. Some focus in detail on the impact mechanisms and processes and use specific theories (based mainly on behavioural economics). Others take the question of external validity seriously and ramp up the number of case studies in different settings (the special issue on microcredit edited by Banerjee, Karlan and Zinman (2015) is a typical example of this; Bédécarrats, Guérin and Roubaud, 2020), or reanalyse ex-post a number of RCTs (Meager 2019). Still others take the question of 'thinking small' seriously and focus on large-scale programmes and national policies. On the question of little bearing on public policies (Pritchett, 2020), some *randomistas* create dedicated bodies if not become decision-makers themselves.

What remains to be seen is the extent to which the *implementation* of this new generation of RCTs in development economics can withstand contingencies on the ground and really evaluate more complex interventions. At the risk of repeating ourselves, we must emphasize the fact that one of the cruxes of the debate is this obsession with the protocol, seen as the priority over its feasibility and its ethical issues. Yet the more complicated the programmes and policies studied, the more likely it is to find tweaking, compromises made and also risks of compromise with the initial protocol. The point is not just to adjust the technique, but to relinquish a scientistic epistemological position in the sense defined above.

3.2 Can we really afford not to have a controversy considering the crowding-out effects?

If a controversy is vital, it is also because the claimed hierarchy of methods has crowding-out effects, in terms of both method (the other methods are discredited), funding and types of interventions, with consequently a performative dimension: the success of RCTs is transforming the development field.

On the question of funding, consider two examples by way of illustration. In the Indian setting, a study truly capable of evaluating the impact of sanitation on infant mortality (the most

appropriate indicator, but one that RCTs do not have the statistical power to capture) would cost around $90 million (subject to certain conditions; Spears et al., 2020). The cost of a classic RCT is between $500,000 and $1,500,000,[22] and each RCT often generates just one published research paper. Is this cost effective when a poor country's statistical household surveys system could be funded for the same amount, with a host of possible studies drawn from these observational data?

On the performative effects of RCTS, the case of health is particularly illustrative. Although they may not have been the primary cause, RCTs did contribute to the rise in vertical health approaches (projects in silos) focused on the individual treatment of specific diseases at the expense of horizontal approaches designed to develop complex, integrated health systems (Garchitorena et al., 2020). Other studies point up the performative (and problematic) effects of the growing use of RCTs (Adams 2016; Biehl et al. 2014): neglecting non-randomizable programmes, altering programmes to make them more easily randomizable, prioritizing evaluation at the expense of the intervention itself (in particular by changing the field staff's work (Adams 2016)). The disruption caused by RCTs and affecting the quality of interventions has been documented in other areas such as microcredit (Bédécarrats, Guérin and Roubaud, 2020c) and micro-insurance (Quentin and Guérin 2013).

4. What are the research alternatives?

Our purpose is not to reject RCTs, since they constitute a promising method ... among others. However, they should still be conducted by the book, take their feasibility and ethical implications seriously by aligning with best practices established in the medical world, and interface with other methods. Although RCTs remain fit and proper for certain precisely defined policies, other methods can and should be used, and the methods combined for the projects that RCTs can address (in part).

An alternative position to the gold standard is to take a pragmatic approach, defining the research questions and methodological tools required on a case-by-case basis according to the prior knowledge available, the intervention design and the particularities of the settings, in liaison with the different stakeholders, whether field operators, donors, governments or the largely overlooked local populations.

These alternative methods also draw on a range of methodologies based on interdisciplinarity and acknowledging the different ways of producing evidence, both quantitative and qualitative. These approaches do not set out to lay down universal laws, but to explain causal

22 No precise estimate of the cost of RCTs exists to our knowledge, but Pamiès-Sumner (2015) provides approximations. See also Ravallion, 2020.

links specific to a particular time and place. Note here the disconnect between the repeated advocacy for mixed methods, whether from researchers[23] or institutions,[24] and their low level of application in practice. On the side of the *randomistas*, although some publicly acknowledge the legitimacy of alternative methods (Ogden, 2020), the fact that they frequently ignore the results of non-randomized methods appears to contradict this apparent open-mindedness (Bédécarrats, Guérin and Roubaud, 2020a).

In the field of global health, the complexity of the interventions is such that randomization is often impracticable and observational and quasi-experimental methods are more appropriate. As shown by Garchitorena et al. (2020), there are numerous examples of alternative and complementary methods to RCTs, even if RCTs remain useful for certain specific interventions. These alternative methods have the particularity of being based on complexity theory (a health system as a whole, rather than fragmented components), combining methods and scales of analysis, drawing where possible on national statistics systems, and addressing not only impact, but also effectiveness (by introducing outputs and outcomes, but also inputs and processes into the analysis).

In addition to the examples mentioned in the book, we would also point up the need to conduct meta-analyses and replications, which are starting to emerge in development economics, but are still too thin on the ground (Camfield and Duvendack 2014). These replications can also be qualitative and revisit a field study, as has been done in Morocco and Bangladesh (Kabeer 2019; Morvant-Roux et al. 2014). Qualitative methods (semi-structured interviews, focus groups, participant observation, ethnography, case studies, life stories, etc.) can serve a number of purposes: to contextualize interventions, develop original hypotheses, identify new and unexpected phenomena, and analyse interventions as a whole, studying the complexity of the causal links and the many, dynamic and contradictory interactions between different entities in a location-specific way. When faced with complex causal chains, which is the case with many interventions, qualitative methods are often the only way to really address the thorny question of causality (White and Masset 2018). Often (unduly) criticized for their inability to 'prove' findings, qualitative methods are also the victims of superficial and non-rigorous uses. Agnès Labrousse (2020) illustrates this misuse by discussing storytelling, a type of narrative designed to illustrate an argument, but which has no power of demonstration, which some *randomistas* misuse in their interpretations of quantitative

23 See, for example, the two books mentioned in the introduction (Cohen and Easterly 2010a; Teele 2014), wherein most of the chapters and introductory statements insist on the need for mixed methods. See also Camfield and Duvendack (2014).

24 See Picciotto (2020), for the evaluation world in development. See also. See also (Pamiès-Sumner 2015) for AFD, and CEDIL's work (White and Masset 2018) for DFID.

results. At the end of the day, the only standard that holds is “good use of good evidence” (Spears et al., 2020).

To sum up and wrap up the book, we believe that some key principles should guide development research, not as alternatives to RCTs, but with RCTS playing a commensurate part. These principles are probably not revolutionary … perhaps one small step for an experienced researcher, but one giant leap for humanities. First, and to make the transition from the general to the specific, research should be guided by important questions to be addressed rather than by methods for which applications need to be found. To paraphrase a famous quote: *Ask not what you can do for an RCT, ask what an RCT can do for your research*! Second, we need to get over the obsession with causal impact, which has dominated the community of development economists ever since the so-called credibility revolution (Angrist & Pischke 2010). Other research questions and approaches are at least as important to advance knowledge such as analyses of observational data, thick description, analytical narratives, especially if we consider that poverty is not only a problem of deprivation but also and sometimes above all the result of social and power relations. Third, on the subject of quantitative approaches, it is essential to rebalance research efforts to take in other components of the analytic chain: what might have been gained in terms of causal attribution (in theory, since different chapters of the book show that nothing is guaranteed in practice in this area), and overinvestment in this area, has left other equally important aspects by the wayside. First and foremost, there is the question of data quality, all too often sacrificed out of a lack of interest and competence. There is then the rise in the number of replications that tackle head on a meticulous diagnostic on the data, and its inclusion in academic journals’ peer review criteria. At the same time, closer attention should be paid to the question of sample designs. All too often, the implications of the use of complex sample designs are overlooked. These oversights result in the underestimation of estimator variance and the consideration of impacts as statistically significant when they are not (Gibson 2019) and when others are, but the repeatedly underpowered RCTs cannot identify them. Fourth, it is time to really put into practice two recommendations on which everyone agrees, but which remain empty talk for now without any tangible effects in practice: real consideration of the ethical issues and the combination of qualitative and quantitative methods. Last but not least, it is time to recognize once and for all that randomized control trials are not the gold standard for evaluation. The hubris that has gripped part of the pro-RCT movement is steering research up against a brick wall (into an impasse). Restoring a sense of moderation to this immoderation is an imperative that can only do good. If not, all aforesaid attempts to listen to the critical voices and adapt will amount to changing everything so that everything can stay the same.

Will the consecration of the Sveriges Riksbank Prize lead the *randomistas* to be more

balanced in their appreciation of the benefits of the different methods or, on the contrary, to take advantage of this consecration to consolidate their already virtually hegemonic position? Only time will tell, but let us insist on the fact that putting an end to the "gold standard" and the quest for the "indisputable" that is characteristic of the *randomistas'* claim to superiority calls for an epistemological break, but also the advent of this controversy which we call for in earnest. Drawing on an examination of the controversies surrounding climate change, Bruno Latour (2012) advocates building debating spaces and methods to discuss and debate the different forms of scientific knowledge (in all their plurality), and non-scientific knowledge, ensuring that the ideological and political bases of these multiple forms of knowledge are neither repudiated nor eclipsed, but are spelled out and debated (Egil 2015). We believe that this project, as ambitious as it may be, is a scientific and democratic necessity if we really hope to improve development policies.

Conclusion: What responses to the pandemic?

Finally, what to expect from social experiments to fight poverty in the context of COVID-19? The answer is unfortunately pessimistic. In addition to all the limitations identified in Bédécarrats, Guérin and Roubaud (2020a) and summarized here succinctly, there are new ones specific to this unique context. The magnitude of the shock is a game changer. The problematic transposition of local results obtained in normal times has no chance to hold in times of pandemic, where behaviors are upset.

The transmission of the disease does not depend only on individual but also collective behavior, synonymous with generalized externalities the method is unable to capture. RCTs serve a knowledge generation strategy where, rather than tackling complex questions as a whole, they are segmented into a series of small questions, the only ones the method is able to address. This micro, "small step" approach is not suited to the need for macro responses, imperative scaling up, national or regional levels, with systematic general equilibrium effects. These are massive emergency measures of partial unemployment, cash transfers, subsidies at decentralized administration, guarantees and lines of credit ..., facing fiscal, monetary and budgetary policies to finance them; mobilizing all the resources (human and financial) of the administration. These kinds of interventions cannot be evaluated through RCTs. In addition, the urgency of implementing measures compared to the long time required to implement the method, as well as the ethical questions posed by the exclusion of populations drawn at random for the purposes of randomization are all additional reasons for ruling out experiments.

Governments around the world have made no mistake about it. While for the development of a vaccine, clinical trials have multiplied, to our knowledge no public policy in the field of non-

pharmaceutical interventions and even less in the economic and social field has been based on the results of RCTs conducted so far or in progress. Even in the medical and epidemiological field, as soon as we look at interventions involving human behavior, the clinical method shows its limits. A systematic review used by WHO of randomized experiments concluded that masks were ineffective for non-caregivers, even pointing to their negative effects (Jefferson et al., 2020). At the same time, another systematic review published by The Lancet, which relied mainly on observational studies, concluded that masks were highly effective for the general population (Chu et al., 2020). Does this mean that we should have given up on masks, or more judiciously applied the precautionary principle, while waiting for scientific studies to settle the question?

The ineffectiveness of social experiments does not mean that the randomistas have given up, on the contrary. Reiterating old studies, RCT proponents offer tips for improving insurance enrollment or online learning. They embark on massive campaigns (25 million people targeted in West Bengal) to test different incentive techniques to respect social distancing measures (Banerjee et al., 2020). One of these techniques is to stage a "celebrity" (in this case Abijit Banerjee himself). They suggest that governments invest massively in two areas, that of cash transfers and digital infrastructure to effectively distribute these transfers to target populations. This is not unnecessary. But on the one hand, many others also argue for this type of interventions and the method of randomized trials does not give them any particular legitimacy. On the other hand, here again, the structural dimension of these measures, both in their large-scale implementation and in their short, medium and long-term effects, remains intact. Yet it is essential. As indicated previously, what are the fiscal, monetary and budgetary policies making it possible to finance these monetary transfer policies? What is the technical and above all the regulatory infrastructure guaranteeing a democratic use of digital finance? While it can deliver money faster to vulnerable populations, it can also lead to control and invasion of privacy.

More than ever, RCTs must be put in their rightful place: one method among others without exceptional status, possibly to be combined with others, and ill-suited to respond to the challenge caused by the pandemic. Even though critical voices are heard, many academic and other circles have yet to realize their limits. RCTs, boosted by the Nobel Prize award and the redoubled efforts of their promoters to extol its merits, continue to drain millions of euros, often to the detriment of other approaches that are at least as rigorous and useful.

List of References

Abramowicz, M. & Szafarz A. (2020). 'Ethics of RCTs: Should Economists Care about Equipoise?', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 280-292.

Acemoglu, D. (2010). 'Theory, general equilibrium, and political economy in development economics', *Journal of Economic Perspectives*, 24/3: 17–32.

Adams, V. (Ed.). (2016). *Metrics: What counts in global health*. Durham and London: Duke University Press.

AFD (ed) (2012). Unease in Evaluation: What Are the Lessons to be Drawn from the Development Experience. *Revue d'Economie du Développement*, 20(4), Special Issue.

Aiken, A. M., Davey, C., Hargreaves, J. R., & Hayes, R. J. (2015). 'Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication', *International journal of epidemiology*, 44/5: 1572–1580.

Akrich, M., Strum, S., Callon, M., & Latour, B. (2013). *Sociologie de la traduction: textes fondateurs*. Presses des Mines via OpenEdition.

Angrist, J. D., & Pischke, J.-S. (2010). 'The credibility revolution in empirical economics: How better research design is taking the con out of econometrics', *Journal of economic perspectives*, 24/2: 3–30.

Augsburg, B., De Haas, R., Harmgart, H., & Meghir, C. (2015). 'The impacts of microcredit: Evidence from Bosnia and Herzegovina', *American Economic Journal: Applied Economics*, 7/1: 183–203.

Banerjee, A. (2013). 'The J-PAL story: A Decade of Partnerships'. *https://www.youtube.com/watch?v=AkC9tVUptM4&list=PL5Dr5MK6NSso3iEqn6BDu8OzyMFyLwiNE&index=5*.

Banerjee, Abhijit V., Marcella Alsan, Emily Breza, Arun G. Chandrasekhar, Abhijit Chowdhury, Esther Duflo, Paul S. Goldsmith-Pinkham, & Olken, B. (2020). 'Messages on Covid-19 Prevention in India Increased Symptoms Reporting and Adherence to Preventive Behaviors among 25 Million Recipients with Similar Effects on Non-Recipient Members of Their Communities'. *NBER Working Paper* (w27496).

Banerjee, A., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Banerjee, A., & He, R. (2008). 'Making aid work'. Easterly W. (ed.) *Reinventing foreign aid*, pp. 47–92. MIT Press: Cambridge.

Banerjee, A., Karlan, D., & Zinman, J. (2015). 'Six randomized evaluations of microcredit: Introduction and further steps', *American Economic Journal: Applied Economics*, 7/1: 1–21.

Barrett, C. B., & Carter, M. R. (2010). 'The power and pitfalls of experiments in development economics: Some non-random reflections', *Applied economic perspectives and policy*, 32/4: 515–548.

——. (2014). 'A retreat from radical skepticism: rebalancing theory, observational data, and randomization in development economics', *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences*, 58–77.

Bédécarrats, F., Guérin, I., Morvant-Roux, S., & Roubaud, F. (2019a). 'Estimating microcredit impact with low take-up, contamination and inconsistent data. A review of Crépon, Devoto, Duflo and Pariente (American Economic Journal: Applied Economics, 2015)', *International Journal for Re-Views in Empirical Economics*, 3.

——. (2019b). 'Lies, damned lies, and RCT : une expérience de J-PAL sur le microcrédit rural au Maroc'. *Working Paper 2019-04*. DIAL: Paris.

Bédécarrats, F., Guérin, I., & Roubaud, F. (2013). 'L'étalon-or des évaluations randomisées: du discours de la méthode à l'économie politique', *Sociologies pratiques*, 2: 107–122.

——. (2019). 'All that Glitters is not Gold. The Political Economy of Randomized Evaluations in Development', *Development and Change*, 50/3: 735–62. DOI: 10.1111/dech.12378

——. (eds) (2020a), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press.

——. (2020b). 'Editors Introduction. Controversies around RCTs in Development: Epistemology, Ethics and Politics', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 1-28.

——. (2020c). 'Microfinance RCTs in Development: Miracle or Mirage?' in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 186-226.

Beisel, U. (2015). 'Markets and Mutations: mosquito nets and the politics of disentanglement in global health', *Geoforum*, 66: 146–55. DOI: 10.1016/j.geoforum.2015.06.013

Bernard, T., Delarue, J., & Naudet, J.-D. (2012). 'Impact evaluations: a tool for accountability? Lessons from experience at Agence Française de Développement', *Journal of Development Effectiveness*, 4/2: 314–327.

Berndt, C. (2015). 'Behavioural economics, experimentalism and the marketization of development', *Economy and Society*, 44/4: 567–591.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). 'What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment', *The Quarterly Journal of Economics*, 125/1: 263–306. DOI: 10.1162/qjec.2010.125.1.263

Biehl, J., Petryna, A., Biehl, J., & Petryna, A. (2014). 'Peopling Global Health', *Saúde e Sociedade*, 23/2: 376–89. DOI: 10.1590/S0104-12902014000200003

Bold, T., Kimenyi, M., Mwabu, G., Nganga, A., Sandefur, J., DiClemente, R. J., Swartzendruber, A. L., et al. (2013). 'Scaling up what works: Experimental evidence on external validity in Kenyan education.', *Sexually Transmitted Diseases*, 40/2: 111–2.

Callon, M. (2006a). 'Pour une sociologie des controverses technologiques'. Akrich M. & Callon M. (eds) *Sociologie de la traduction. Textes fondateurs*, pp. 135–157. Presses des Mines: Paris.

——. (2006b). 'Quatre modèles pour décrire la dynamique de la science'. Akrich M. & Latour B. (eds) *Sociologie de la traduction. Textes fondateurs*, pp. 201–251. Presses des Mines: Paris.

Camfield, L., & Duvendack, M. (2014). 'Impact Evaluation – Are We "Off the Gold Standard"?', *The European Journal of Development Research*, 26/1: 1–11. DOI: 10.1057/ejdr.2013.42

Cartwright, N. (2010). 'What are randomised controlled trials good for?', *Philosophical studies*, 147/1: 59.

Chu, Derek K., Elie A. Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J. Schünemann, Derek K. Chu, Elie A. Akl, Amena El-harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, Anisa Hajizadeh, Anna Bak, Ariel Izcovich, Carlos A. Cuello-Garcia, Chen Chen, David J. Harris, Ewa Borowiack, Fatimah Chamseddine, Finn Schünemann, Gian Paolo Morgano, Giovanna E. U. Muti Schünemann, Guang Chen, Hong Zhao, Ignacio Neumann, Jeffrey Chan, Joanne Khabsa, Layal Hneiny, Leila Harrison, Maureen Smith, Nesrine Rizk, Paolo Giorgi Rossi, Pierre AbiHanna, Rayane El-khoury, Rosa Stalteri, Tejan Baldeh, Thomas Piggott, Yuan Zhang, Zahra Saad, Assem Khamis, Marge Reinap, Stephanie Duda, Karla Solo, Sally Yaacoub, and Holger J. Schünemann. 2020. 'Physical Distancing, Face Masks, and Eye Protection to Prevent Person-to-Person Transmission of SARS-CoV-2 and COVID-19: A Systematic Review and Meta-Analysis'. *The Lancet* 395(10242):1973–87. doi: 10.1016/S0140-6736(20)31142-9.

Cohen, J., & Easterly, W. (2010a). *What works in development?: Thinking big and thinking small*. Washington, DC: Brookings Institution Press.

——. (2010b). *What Works in Development?: Thinking Big and Thinking Small*. Brookings Institution Press.

Council for International Organizations and Medical Sciences. (2002). *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. Geneva: World Health Organization.

Crucifix, C., & Morvant-Roux, S. (2018). 'Fragmented rural communities: The faenas of Prospera at the interface of community cooperation and state dependency'. Balen M. E. & Fotta M. (eds) *Money from the Government in Latin America: Conditional Cash Transfer Programs and Rural Lives*, pp. 123–48. Oxon, Routledge: Abingdon, New-York.

Davey, C., Aiken, A. M., Hayes, R. J., & Hargreaves, J. R. (2015). 'Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a

cluster quasi-randomized stepped-wedge trial', *International Journal of Epidemiology*, 44/5: 1581–1592.

Deaton, A. (2020). 'Randomization in the Tropics Revisited: A Theme and Eleven Variations', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 29-46.

Deaton, A. (2010). 'Instruments, randomization, and learning about development', *Journal of economic literature*, 48/2: 424–55.

Deaton, A., & Cartwright, N. (2018). 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine*, 210: 2–21.

Desrosières, A. (2013). *Pour une sociologie historique de la quantification: l'argument statistique I*. Presses des Mines via OpenEdition.

Development Assistance Committee. (2002). *Glossary of Key Terms in Evaluation and Results Based Management'. , Paris.* Paris: Organisation for Economic Co-Operation and Development.

Dimova, R. (2019). 'A Debate that Fatigues…: To Randomise or Not to Randomise; What's the Real Question?', *The European Journal of Development Research*, 31/2: 163–8. DOI: 10.1057/s41287-019-00198-3

Drèze, J. (2019). 'Evidence, policy, and politics'. *Ideas For India*. Retrieved October 25, 2019, from <http://www.ideasforindia.in/topics/miscellany/evidence-policy-and-politics.html>

Duflo, E. (2017). 'Richard T. Ely Lecture: The Economist as Plumber', *American Economic Review*, 107/5: 1–26.

Duflo, E., Dupas, P., & Kremer, M. (2015). 'School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools', *Journal of Public Economics*, 123: 92–110.

Duflo, E., Glennerster, R., & Kremer, M. (2004). 'Randomized evaluations of interventions in social service delivery', *Development Outreach*, 6/1: 26–29.

Duflo, E., & Kremer, M. (2005). 'Use of randomization in the evaluation of development effectiveness', *Evaluating development effectiveness*, 7: 205–231.

Egil, F. (2015). 'Les Objectifs de développement durable, nouveau «palais de cristal»?', *Politique africaine*, 4: 99–120.

Elyachar, J. (2012). 'Next practices: Knowledge, infrastructure, and public goods at the bottom of the pyramid', *Public Culture*, 24/1 (66): 109–129.

Encisco, A. L. (2019). 'Acaba el clientelar Prospera; surge el programa Becas Benito Juárez'. *La Jornada*.

Fassin, D. (2010). *La raison humanitaire. Une histoire morale du temps présent*. Paris: Gallimard/Seuil.

Faulkner, W. N. (2014). 'A critical analysis of a randomized controlled trial evaluation in Mexico: Norm, mistake or exemplar?', *Evaluation*, 20/2: 230–243.

Fine, B., Johnston, D., Santos, A. C., & Van Waeyenberge, E. (2016). 'Nudging or fudging: The world development report 2015', *Development and Change*, 47/4: 640–663.

French, J., Blair-Stevens, C., McVey, D., & Merritt, R. (2010). *Social Marketing and Public Health: Theory and Practice*. Oxford: Oxford University Press.

Garchitorena, A., Murray, M., Hedt-Gauthier, B., Farmer, P., & Bonds, M. (2020). 'Reducing the Knowledge Gap in Global Health Delivery: Contributions and Limitations of Randomized Controlled trials', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 152-165.

Gibson, J. (2019). 'Are You Estimating the Right Thing? An Editor Reflects', *Applied Economic Perspectives and Policy*, 41/3: 329–350.

Gugerty, M. K., & Kremer, M. (2008). 'Outside funding and the dynamics of participation in community associations', *American Journal of Political Science*, 52/3: 585–602.

Harrison, G. W. (2011). 'Randomisation and its discontents', *Journal of African Economies*, 20/4: 626–652.

Heckman, J. J. (2020). 'Epilogue: Randomization and Social Policy Evaluation Revisited', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 304-330.

Heckman, J. J. (1991). *Randomization and social policy evaluation*. National Bureau of Economic Research Cambridge, Mass.

Humphreys, M. (2015). 'What Has Been Learned from the Deworming Replications: A Nonpartisan View'. *Columbia University*. Retrieved November 11, 2019, from <http://www.columbia.edu/~mh2245/w/worms.html>

IFC. (2017). *Strategy and business outlook FY18-FY20. Creating markets and mobilising private capital*. Washington, DC: International Finance Corporation.

Jamison, J. C. (2017). *The entry of randomized assignment into the social sciences*. Washington DC: The World Bank.

Jatteau, A. (2016). *Faire preuve par le chiffre? Le cas des expérimentations aléatoires en économie* (PhD Thesis). Paris Saclay.

Jean-David, N. (1999). *Trouver des problèmes aux solutions Vingt ans d'aide au Sahel: Vingt ans d'aide au Sahel*. Washington; Biggleswade: OECD Publishing.

Jefferson, T., M. A. Jones, L. Al-Ansary, G. A. Bawazeer, E. M. Beller, J. Clark, J. M. Conly, C. Del Mar, E. Dooley, E. Ferroni, P. Glasziou, T. Hoffmann, S. Thorning, and ML van Driel (2020). 'Physical

Interventions to Interrupt or Reduce the Spread of Respiratory Viruses. Part 1 - Face Masks, Eye Protection and Person Distancing: Systematic Review and Meta-Analysis'. *MedRxiv* 2020.03.30.20047217. doi: 10.1101/2020.03.30.20047217.

Jensen, R. (2010). 'The (perceived) returns to education and the demand for schooling', *The Quarterly Journal of Economics*, 125/2: 515–548.

Kabeer, N. (2019). 'Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections', *Journal of Human Development and Capabilities*, 20/2: 197–217.

Karlan, D., & Zinman, J. (2009). 'Expanding credit access: Using randomized supply decisions to estimate the impacts', *The Review of Financial Studies*, 23/1: 433–464.

Kidd, S. (2019). 'The demise of Mexico's Prospera programme: a tragedy foretold'. *Development Pathways*. Retrieved November 5, 2019, from <https://www.developmentpathways.co.uk/blog/the-demise-of-mexicos-prospera-programme-a-tragedy-foretold/>

Labrousse, A. (2020). ' The Rhetorical Superiority of Poor Economics', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 227-255.

Latour, B. (2012). *Enquête sur les modes d'existence. Une anthropologie des Modernes*. Paris: La Découverte.

Lemieux, C. (2007). 'À quoi sert l'analyse des controverses?', *Mil neuf cent. Revue d'histoire intellectuelle*, 1: 191–212.

Lundberg, S., & Stearns, J. (2019). 'Women in economics: Stalled progress', *Journal of Economic Perspectives*, 33/1: 3–22.

MacKenzie, D. A., Muniesa, F., & Siu, L. (2007). *Do economists make markets?: on the performativity of economics*. Princeton: Princeton University Press.

McKenzie, D. (2012). 'Beyond baseline and follow-up: The case for more T in experiments', *Journal of development Economics*, 99/2: 210–221.

Meager, R. (2019). 'Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments', *American Economic Journal: Applied Economics*, 11/1: 57–91.

Miguel, E., & Kremer, M. (2004). 'Worms: identifying impacts on education and health in the presence of treatment externalities', *Econometrica*, 72/1: 159–217.

Morduch, J. (2020). 'The Disruptive Power of RCTs', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 108-125.

Morvant-Roux, S., Guérin, I., Roesch, M., & Moisseron, J.-Y. (2014). 'Adding value to randomization with qualitative analysis: the case of microcredit in rural Morocco', *World Development*, 56: 302–312.

Mosse, D. (2004). *Cultivating Development: An Ethnography of Aid Policy and Practice.* London: Pluto Press.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington DC: US Department of Health, Education and Welfare.

Naudet, J.-D. (2006). 'Les OMD et l'aide de cinquième génération', *Afrique contemporaine*, 2: 141–174.

Ogden, T. N. (2020). 'RCTs in Development Economics, Their Critics, and Their Evolution', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 126-151.

Ogden, T. N. (2017). *Experimental conversations: Perspectives on randomized trials in development economics*. Cambridge, Massachusetts: MIT Press.

Olivier de Sardan, J.-P. (1995). *Anthropologie et développement: essai en socio-anthropologie du changement social*. Paris: Karthala.

Pamiès-Sumner, S. (2015). *Development Impact Evaluation, State of Play and New Challenges* (A Savoir). Paris: AFD.

Petryna, A. (2007). 'Clinical trials offshored: on private sector science and public health', *BioSocieties*, 2/1: 21–40.

Picciotto, R. (2020). 'Are the 'Randomistas' Evaluators?', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 256-279.

Pritchett, L. (2020). 'Randomizing Development: Method or Madness?', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 79-107.

Pritchett, L., & Sandefur, J. (2013). 'Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix', *Center for Global Development Working Paper*, 336.

——. (2015). 'Learning from experiments when context matters', *American Economic Review*, 105/5: 471–5.

Putnam, H. (2009). *Renewing philosophy*. Harvard: Harvard University Press.

Quentin, A., & Guérin, I. (2013). 'La randomisation à l'épreuve du terrain', *Revue tiers monde*, 1: 179–200.

Ravalllion, M. (2020). 'Should the Randomistas (Continue to) Rule?', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 47-78.

Raynaud, D. (2018). *Sociologie des controverses scientifiques*. Paris: Éditions Matériologiques.

Redfield, P. (2012). 'Bioexpectations: Life technologies as humanitarian goods', *Public Culture*, 24/1 (66): 157–184.

Rodrik, D. (2009). 'The new development economics: we shall experiment, but how shall we learn?'. Cohen J. & Easterly W. (eds) *What works in development?: Thinking big and thinking small*, pp. 24–47. The Brookings Institution: Washington, DC.

Servet, J.-M. (2018). *L'Economie comportementale en question*. Paris: Fondation pour le Progrès de l'Homme.

Shaffer, P. (2015). 'Two concepts of causation: Implications for poverty', *Development and Change*, 46/1: 148–166

Spears, D., Ban, R., & Cumming, O. (2020).'Trials and Tribulations: The Rise and Fall of the RCT in the WASH Sector', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 166-185.

Teele, D. L. (Ed.). (2014). *Field experiments and their critics. Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven & London: Yale University Press.

Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. New York: Penguin.

The Royal Swedish Academy of Sciences. (2019). 'Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019. Understanding Development and Poverty Alleviation'.

Vivalt, E. (2020). 'Using Priors in Experimental Design: How Much Are We Leaving on the Table?', in Bédécarrats, F., Guérin, I., & Roubaud, F. (eds), *RCT in Development: A Critical Perspective*, Oxford: Oxford University Press, 293-303.

Webber, S., & Prouse, C. (2018). 'The New Gold Standard: The Rise of Randomized Control Trials and Experimental Development', *Economic Geography*, 94/2: 166–87. DOI: 10.1080/00130095.2017.1392235

White, H., & Masset, E. (2018). 'The rise of impact evaluations and challenges which CEDIL is to address', *Journal of Development Effectiveness*, 10/4: 393–9. DOI: 10.1080/19439342.2018.1539387

WMA General Assembly. (2014). *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects.*, 9th revision (1st version: 1964). Fortaleza: World Medical Association.

Woolcock, M. (2013). 'Using case studies to explore the external validity of 'complex'development interventions', *Evaluation*, 19/3: 229–248.

Wydick, B. (2016). 'Microfinance on the margin: why recent impact studies may understate average treatment effects', *Journal of Development Effectiveness*, 8/2: 257–265.