

Some More Reflections on RCTs

Maria Perrotta Berlin, SITE

October, 2013

In preparation of next year's elections, the Swedish government chose recently to replace the Minister for International Development Cooperation. During her long mandate, former Minister Gunilla Carlsson championed the importance of aid evaluation and result focus, and managed to move aid from a quiet consensus to become a hotly debated topic. She also closed down the aid evaluation agency SADEV, following the publication of critical reviews about the work of the agency. Now, an expert group is in charge of rethinking and redesigning development policy evaluation and planning. One of the tools under consideration is randomized control trials (RCTs). This is an area in which Swedish development cooperation has no previous experience. Here are some reflections on RCTs.

In recent years, the methods of development economics have been crucially altered by the introduction of randomized control trials (RCTs). The idea behind RCTs is that development policies can be evaluated similarly to clinical trials in medicine, where subjects are randomly assigned to receive a treatment or to function as a reference or control group. The main benefit of this approach is that the random assignment allows for an estimation of the effect of the treatment (that is, the policy in question), while avoiding unobservable confounding factors or selection issues (see more about the advantages of the method in Banerjee et al. (2008)).

The diffusion of experimental methods in development economics has undoubtedly been a revolution in the academic and, if not yet fully, in the policy world. In the blogosphere there has even been talk of awarding *Sveriges Riksbank's Prize in Economic Sciences in Memory of Alfred Nobel*, informally called the Nobel Prize of Economics, to the MIT couple Banerjee – Duflo. Due to their young age and the closeness in time of their contribution, this would be a "shock" prize meant to give a strong signal. Their creation, the Abdul Latif Jameel Poverty Action Lab (J-PAL), stands for a new approach to both scientific and policy work in development that is a fantastic

contribution, and definitely has the connotation of seminal.

However, it might be too early for the profession to sanction a method that has much good to show for, but also potentially undesired consequences. In the camp of critics there are heavy weights such as Angus Deaton and Dani Rodrik of Princeton, and the World Bank's Philip Keefer and Martin Ravallion. The core of their position is of course not to deny the merits of RCTs, but to advocate their use in the *right way* and, in particular, as *one* tool among many others, with important complementarities to the others.

Some points in this context are often made, well understood and widely accepted: the limits of the approach per se, in particular the problem of *external validity* (the question of how generally applicable are the findings from such studies); the conflict between short-run and long-run implications, especially with respect to some policy areas (support to institution-building among others), and the incentives of policy actors. Another [brief](#) in this series by Anders Olofsgård spells out these points very clearly and references to further readings for those interested.

One aspect I find to be missing in the debate is a reflection on what impact this new method has on the three main actors involved, namely the researchers and practitioners in development and their way of working, and the people living in the countries and regions where these studies take place. This will therefore be the focus of this brief.

The Impact on the Scholarly Profession

The creation of experimental infrastructures and the popularity of the RCT methodology have rubbed off on the rest of the empirical practice in development economics and beyond, with ever-increasing demands and expectations on the econometric identification of new studies. However, when it comes to what is possibly the main weakness of RCTs as compared to most observational studies, namely external validity, the corresponding demands and expectations on how this is dealt with seem to fall behind. As pointed out in Rodrik (2008), it is enough to compare the number of pages spent on describing the identification in an average observational study to that on external validity in an average RCT-based paper. If the purpose is to learn “what works in development”, as opposed to “what worked once for a set of 25 primary schools in Uttar Pradesh faced with high drop out rates”¹, it is natural to expect the researcher that really wants to serve this purpose to provide for a desired generality of her findings. With no generality, the findings may be of limited practical use to politicians and practitioners who need to choose a policy tool or make a decision in conditions, which are likely to differ from the exact setting of the study.

During a recent presentation by one of the most active and prominent RCT researchers, the researcher clearly stated at some point that: “[t]his intervention was never thought for

scaling up as a policy.” That made me pause. But what is the purpose, then? In my meaning, these studies should fit into a “bigger-picture” understanding, or at least hypothesizing on how development works, what the binding constraints and open challenges are, what might contribute to overcoming them, and how do we proceed from there. Once some candidates are identified, RCTs might, depending on the setting, be used to evaluate and compare before and after the preferred policy is implemented. Unfortunately, this attitude is far from common, beyond what has become the standard of the ‘Introduction paragraphs’.

Quite often RCT studies are extremely precise and accurate on “the impact of X on Y”, even in cases of very small effects, and can be perhaps a bit vague or face bigger uncertainties on the ‘bigger’ question. This means that many, more general (and very relevant) questions are not addressed by development economists just because a RCT is not feasible. An example mentioned in a recent keynote lecture by David Laitin is the BetterBirth Project. This is a WHO program that seems to be making a big difference for infant and maternal health in India’s poorest states through a list of 29 easy, low-cost, low-technology and well-known practices. The main lesson drawn by observers at the Harvard School of Public Health is that people follow the list more accordingly when it is spread through “human contact”. No mass media advertisement campaign, no punishment or incentive schemes, just “nice” people visiting, explaining, and demonstrating the list, while – in the words of an interviewed nurse – “smiling a lot”. At first sight, this seems like something that could be randomized. However, the treatment is so diffuse and fuzzy that the practical implementation would be very challenging. If it is the case that the person meeting the clinics’ personnel and spreading the information has to be somewhat of a mentor in order for the transition to happen, to be kind and pedagogic, repeat the visits indefinitely to make sure that the practices have been adopted, and do whatever

¹ The example is fictitious. Any resemblance to real studies is unintended and purely coincidental.

else it takes to make them learn, this is very hard to observe with precision. To simply define X as "presentation of the list in person", to be compared to, for example, the "diffusion of the list through an information campaign" would probably run the risk of severely underestimating the impact. This would be because it would bundle together different types of informers and different levels of human interaction. This means that there would be a high risk of zero or insignificant results from such a study. A RCT would need to be complemented by other investigations, for example surveys, in order to find out if there really was an effect and how it came about. All of the above is likely to undermine the publication chances for an academic paper on the issue, thereby discouraging development scholars to study this program.

There are two main ways of augmenting the RCT methodology in the direction of generalizability and external validity: the elbow-grease approach of replication and the resuscitation of the concern for theoretical mechanisms. Replication studies are not very appealing in the perspective of a scholar that aspires academic publications. Besides completely new clever designs that establish a link of causation in a specific case - and possibly for each of these corresponding studies that establishes the absence of such a link in different settings - journals have little interest in publishing more variations on the same theme. Replications with small variations should instead be highly attractive for development institutions and practitioners, precisely for the reason, mentioned above, that they want to learn about effectiveness of alternative strategies in as many different specific contexts as possible.² In an ideal world, development institutions and aid bureaucracies would work in close cooperation with universities and academic institutions, involving young researchers before their career-concern-stress phase (perhaps Ph. D. students?) in the design and evaluation of as

many of their planned interventions as possible. Moreover, in an ideal world this would be enough reward for the young researchers. This wealth of replications would then favor the possibility of "taking stock" and really learning about some general truth. I do not, however, have a good recipe for making this happen.

Luckily, some scholars are in the meanwhile working on making the pendulum swing back from the purest empiricism to the involvement with theory. Here is a list of possibilities that are important to reflect about, starting from a given RCT:

- **The macro problem.** How does the found effect compare to the "bigger issue", the one that most likely set the scene in the 'Introduction paragraph' of the study? Few studies go back to this point, after presenting their results. Numerical simulations or structural estimation of theoretical models might help answering this question. (See some examples in Buera et al. (2011) and Kaboski et al. (2011)).
- **The alternative hypothesis.** What is the particular intervention compared against? If the set of circumstances or policy-relevant parameters that might be varied are too big or too dense for replications, maybe a theoretical model can help to vary them in a smooth and continuous way?
- **The strategic reaction.** How are the involved economic agents likely to respond in case of an expansion in space, time or both, of the intervention? How would they have responded in the absence of the intervention?

The Impact on Development Practices

As stated above, RCTs may be a powerful tool for the learning and decision-making in

² At least in theory – this point is discussed more in the next section.

development institutions, public or private. However, this assumes a seldom-questioned willingness to learn and change practices on their part. Brigham et al. (2013) show, through a RCT, that these organizations might be subject to confirmation bias. Brigham et al. sent out an invitation to microfinance institutions, offering partnership to evaluate their programs, randomly accompanying it with a survey of previous studies finding positive impact of microcredit, or a survey of studies finding no impact. The second treatment elicited barely half as many responses as the first one, which suggests that at least this type of organizations might not be so interested in learning whether what they do is effective or can be improved. Coupled with the mentioned publication bias, this might skew the distribution of reported, published and established findings even further.

The Impact on the Local Context

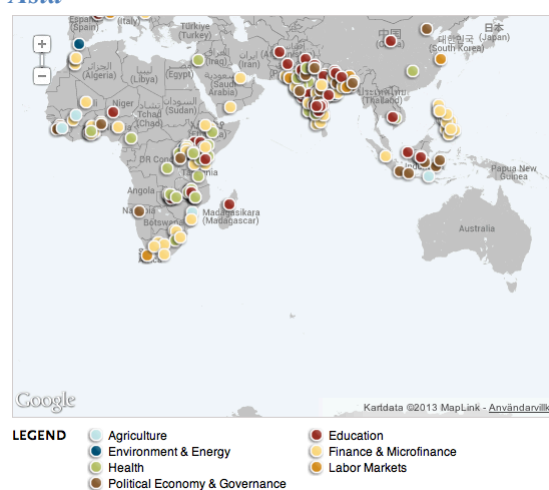
Individual studies can of course be affected by the so-called Hawthorne effect or experimenter effect. The phenomenon, by which the act of being experimented upon changes a subject's behavior, was first observed and got its name in the 1920s in industrial psychology. Although it is clearly hard to establish, it has for decades been a central criticism of the "participant observation" methodology in anthropology and ethnography. Also behavioral economists, that more recently started using experiments both in labs and in the field, are explicitly careful about it.

Depending on the definition of causality that the researcher has in mind, the fact that having knowledge about being treated impacts outcomes, might not be an issue at all for the measurement of the overall effect of an intervention. The overall effect *should* include also the (optimal) reaction of the agents (for example a change in behavior, the adoption of other complementary inputs, etc.) and this is

actually considered one of the advantages of the method. However, this raises problems for the interpretation of the size of the effect and the analysis of the channels that bring it about. This point is made very clearly by Bulte et al. (2012), who compare a double-blind RCT with a regular one. If all or most of the effect simply comes from the participants knowing to be "treated" and reacting to it, is the effect still going to be there when the intervention becomes a regular policy? The majority of both authors and critics mostly ignore this important question.

Beyond the perspective of a single study, a different concern comes to mind when considering how a substantial number of RCT studies are clustered geographically. The map below shows a snapshot of the J-PAL interventions in Africa and Asia, which are only a fraction, albeit substantial, of the total.

Figure 1. J-PAL Interventions in Africa and Asia



Reading study after study set in Kenya, or some Indian state, I wonder if people there are starting to get used to private organizations going around giving away assets, or used to temporary local government programs with funky benefit schemes. To my knowledge, no study has yet reflected upon the aggregate impact of experiments and randomized interventions in an area that has many. Might it be the case that exposure to many conditions eventually results in "experimental fatigue", or

practice effects, which may influence the results of the studies and make the interpretation of the findings difficult?

Even more worrisome, given the frequency of and the resources involved in these interventions, perhaps we should expect an impact on the local political economy. As a parallel, I think about the agrarian reform and the later establishment of the welfare state in post-war Italy, and how they gave major local actors the ability to uphold their clientelistic systems. The newly established rights and entitlements, the various benefits and redistribution programs, were "filtered" by the local elites and channeled through the traditional ties of family, kinship, friendship and neighborhood. According to comparative analyses of European welfare regimes, clientelism exists, in different forms and intensities, in all Mediterranean welfare states, and it appears to be linked to the process of political mobilization and the establishment of welfare state institutions in these nations.

A recent study by Ravallion et al. (2013) finds that unemployed fail to act on information about the National Employment Guarantee Scheme (NEGS) in India. They hypothesize that the bottleneck lies with the local government institutions (*Gram Panchayats*). The GP are supposed to receive the applications and apply for central government resources for planning and implementation of projects, so as to guarantee 100 days of work per year to all adults from rural households who are willing to do unskilled manual labor at the statutory minimum wage. But perhaps - argue the authors - given the strict controls on corruption, the GP officials do not find anything in it for themselves, and hence do not proceed. Of course this is just one of the possible explanations, and moreover the NEGS is not a RCT. But in general the involvement of local official or unofficial power structures in contexts where this type of interventions are increasingly common could be interestingly related to the hypothesis on the "Mediterranean welfare state" outlined

above. The idea definitely deserves investigation.

Conclusions

The popularity of RCTs among development scholars is finally spreading to practitioners. This is mostly good news, there is much to gain and learn from this approach, especially in contexts where it is grossly underexploited, as has been the case in Sweden. However, a near-monopoly of this approach is though not granted, given its non-negligible limitations, often belittled in light of its numerous strengths. Spurring development "one experiment at a time" might take unnecessary extra time and efforts, and bring about other undesirable consequences. Both development scholars and practitioners should not forget the other arrows in their quiver.

■

References

- Bannerjee, A. and E. Duflo (2008), "The Experimental Approach to Development Economics", NBER Working Paper 14467.
- Brigham, Matthew, Michael Findley, William Matthias, Chase Petrey, and Daniel Nelson. "Aversion to Learning in Development? A Global Field Experiment on Microfinance Institutions". Technical Report, Brigham Young University March 2013.
- Buera, F. J., J. P. Kaboski, and Y. Shin (2011). "The macroeconomics of microfinance."
- BREAD working paper.
- Bulte, E., Pan, L., Hella, J., Beekman, G. and S. di Falco (2012). "Pseudo-Placebo Effects in Randomized Controlled Trials for Development: Evidence from a Double-Blind Field Experiment in Tanzania." Working Paper.
- Kaboski, J. P. and R. M. Townsend (2011, July). "A structural evaluation of a large-scale quasi-experimental microfinance initiative." *Econometrica* 79, 1357–1406.
- Olofsgård, A. "What Do Recent Insights From Development Economics Tell Us About Foreign Aid Policy?" FREE Policy Brief Series, October 3, 2011.

Ravallion, M., et al. "Try Telling People their Rights? On Making India's Largest Antipoverty Program work in India's Poorest State." *Department of Economics, Georgetown University, Washington DC* (2013).

Rodrik, D. (2008). 'The New Development Economics: We Shall Experiment, but How Shall We Learn?'. Harvard Kennedy School Working Paper No. RWP08-055.▪

Maria Perrotta Berlin

Stockholm Institute of
Transition Economics (SITE)

Maria.Perrotta@hhs.se

<http://www.hhs.se/site>



Maria Perrotta Berlin is Assistant Professor at the Stockholm Institute of Transition Economics. Perrotta holds a Ph.D. in Economics from the International Institute of Economics Studies (IIES), Stockholm University. Her main research areas are political economics and development.