

Why Randomised Controlled Trials need to include human agency

This year's Nobel Prize in Economics was won by Esther Dunflo, Abhijit Banerjee, and Michael Kremer for using Randomised Controlled Trials to assess poverty alleviation initiatives. Noted economist Naila Kabeer discusses some critiques of the method.

NAILA KABEER

There's a buzz abroad in the development community around a new way to tackle extreme poverty. For example BRAC's Targeting the Ultra Poor (TUP) programme combines asset transfers (usually livestock), cash stipends, and intensive mentoring to women and families in extreme poverty in order to help them "graduate" into more sustainable livelihoods within two years.

But how do we know if it works? One of the main ways is through a series of Randomised Controlled Trials (RCTs). Still seen by their fans as the "gold standard" of evidence, RCTs have come under sustained criticism in recent years, so we decided to take a closer look.

Between 2009 and 2010, a team of researchers including myself carried out qualitative evaluations of BRAC's TUP programmes in rural West Bengal and rural Sindh.

At the same time, close to our project in each location, RCTs were being carried out for other TUP pilots. This was not a coincidence as all four evaluations were part of the same graduation programme testing the TUP approach across the world. Ideally, the evaluations should have integrated the two methods, but resistance from the RCT practitioners ruled this out.

My recent article "Randomized Control Trials and Qualitative Evaluations of a Multifaceted Programme for Women in Extreme Poverty: Empirical Findings and Methodological Reflections", published in the *Journal of Human Development and Capabilities* reviews the two sets of studies in West Bengal and Sindh. It finds that inattention to the question of human agency is one of the key limitations of RCTs.

In keeping with established protocol, the RCT studies in West Bengal and Sindh identified households considered to be ultra-poor within selected villages and then randomly assigned some of them to participate in the pilot, i.e the "treatment" group, and the rest to a "control" group. This is intended to ensure that any improvement in the lives of the treatment group compared to the control group can be attributed to the pilot, because the two groups were identical in all other respects. Because of this assumption, RCTs generally do not consider it necessary to provide (or indeed collect) any information on how project support translates into impact.

Our qualitative evaluations took a different route. We worked closely with project staff to identify 20 women and their families in each location, for our RCT, half of whom the staff considered

to have benefited from the pilot and half who had not. We carried out in-depth and repeated interviews with these women and their families over a year, covering the final months of the project and after.

A published synthesis of the TUP RCTs was positive, but noted that those who had started out better off reported stronger impacts. The findings of our qualitative studies diverged to some extent. In West Bengal, we found evidence of positive impacts, but it was the least well off that did better. In Sindh, on the other hand, most participants failed to make progress, although here those who were better off reported stronger impacts.



PHOTO: BRAC

I tried to work out what lay behind the similarities and differences in the findings of the RCTs and qualitative studies in the two locations. This was not easy as the published version of the RCTs provided minimal information. Instead, I had to look for clues in the "grey" literature and draw my own conclusions.

A fairly detailed account of the West Bengal study reports that 50 percent of those selected to receive assets refused to participate. The majority of these were poorer Muslims who mistrusted project intentions. So one plausible reason why the treatment group reported sizeable positive impacts was that the poorest among them had dropped out while the poorest in the control group had remained. In other words, the better off group who remained in the treatment group were clearly driving the impacts, impacts large enough to prove positive whether they were averaged only

for those who participated or also for those who had dropped out.

Our qualitative study in West Bengal found it was households from the Scheduled Tribes, the poorest and most marginalised group, who did better than the rest. Our interviews with staff and participants told us these groups had been systematically bypassed by all previous development interventions. They were more determined than the others to make the most of what they saw as a once-in-a-life time opportunity. In addition, the implementing organisation had added a group component to project design

allowing women to save regularly and safely and provide support to each other.

The Sindh RCT encountered more serious implementation problems. There is only a brief report on the study, but these are discussed in detail in a separate evaluation commissioned by the Pakistan Poverty Alleviation Fund, which managed the graduation programme in Pakistan. The evaluation noted serious flaws in the randomisation process. It was correctly followed by some of the implementing organisations but misunderstood by others. As a result, there was no guarantee that the control and treatment households started out with similar characteristics or indeed that they were ultra-poor at all. In fact, around 80 percent of the treatment households in the Sindh pilots were found to be above the poverty line of USD 1.25 at the outset.

Most participants in the Sindh pilot

in the qualitative evaluation failed to make progress because the implementing organisation had no previous experience of working with rural women in extreme poverty and failed to understand local conditions. As a result, many of the animals and poultry they distributed died. The better off participants were still able to make some progress because they had started out with advantages that allowed them to make the most of TUP support. These are the kinds of causal mechanisms that RCTs don't pick up, or even look for.

In our examples, the refusal by Muslims to take part in the West Bengal pilot introduced precisely the biases that RCTs are meant to avoid. In Sindh, the failure of implementing organisations to follow randomisation procedures raised questions about the distribution of characteristics among treatment and control groups and resulted in a sample that was overwhelming above the poverty line. The lack of relevant experience on the part of the implementing organisation in the Sindh qualitative evaluation explains its abysmal results. Equally, it was the longer standing experience of the implementing organisation in West Bengal that not only led to positive changes in the lives of participants but also enabled some of the poorest participants to respond most actively to the opportunities it offered.

Conclusion? If evaluation studies are to provide an effective guide to address the persisting problem of poverty, they need to provide information that explains their findings: what works, what doesn't, for whom, why, and whether it matters. In particular, RCTs need to acknowledge the central role of human agency in enabling or thwarting project objectives at every stage of the processes they study. It is unlikely they will be able to do this by confining themselves to quantitative methods alone.

One of the problems with problematic RCTs of the kind described here is that we are, in the end, none the wiser as to how effective or ineffective an intervention was.

Naila Kabeer is professor of gender and development at the London School of Economics and Political Science (LSE). Her research interests include gender, poverty, social exclusion, labour markets and livelihoods, social protection and citizenship. Much of her research is focused on South and South East Asia.

A version of this article was published on oxfamblogs.org