

## RCTs AND RURAL DEVELOPMENT – AN ABUNDANCE OF OPPORTUNITIES

No, it is not a Root Canal Treatment, although some sceptics view them similarly. RCT, or Randomised Controlled Trial, is an evaluation design which when applied to the right type of questions, done in the right way, supplemented by alternative methods and interpreted correctly yields probably the most robust results any evaluation design can achieve, and this with less pain (dental or otherwise) than some of the alternatives. An overview on challenges and opportunities.

By Marie Gaarder\* and Sven Harten

Randomised controlled trial (RCT) is one of a range of designs known as impact evaluations whose explicit purpose is an analysis of attribution. The name is actually misleading as it could be understood as designs that exclusively measure impact indicators. Impact evaluations should more appropriately have been titled *counterfactual evaluations* or *attribution analysis*. Be that as it may, attributing an effect (be it an output, outcome or impact) to an intervention means that all other factors outside of the intervention that could also influence the outcome are held constant (or in the jargon “are controlled for”). Typical factors in the rural environment that need to be controlled for are season, weather, access or distance to markets and market prices.

### A ‘WITH VERSUS WITHOUT’ ANALYSIS

So how do impact evaluations, and RCTs more specifically, control

for these factors? Impact evaluations look at what difference a programme or intervention made: did it lead to measurable improvements on some outcome of interest, be it latrine-use in rural India, higher farm incomes through weather index insurance in Tanzania, or increased women empowerment through mobile money (digital financial services) in Northern Uganda? Impact evaluation is a ‘with versus without’ analysis: what happened with the programme (which is a factual record) compared to what would have happened in the absence of the programme (which requires a counterfactual). Most development agencies produce reports about implementation and results at the time of project closure, if not earlier. Why are these usually misleading when it comes to results? These reports typically rely solely on information and monitoring data provided by the programme, and thus

quite frequently fall victim to the before-after fallacy. Consider measuring an outcome both before the programme starts and after it has been implemented for a while.

Typically, if there is an improvement, the programme manager considers the intervention a success. But over the period of any programme, many other factors come into play, not least of which all the other programmes that are being implemented in the same country. Without a valid counterfactual, there is no way of knowing whether the improvement can be attributed to the programme’s activities or may have happened in spite of these. Moreover, spending money on anything or conducting any kind of activity produces some positive effect in many cases: when farmers are trained on a new, semi-automatic irrigation technique, at least some of them will change their behaviour and get some better yields. But was this the most effective and efficient way to increase yield or could a training on improving traditional irrigation have produced much



better results? Then again, if there is no measurable improvement, it could well be that the programme acted as a safety net if for example the same outcome worsened in the rest of the country.

## A CHALLENGING TASK

In designing an impact evaluation, it is important to carefully consider first what is already known (no need to reinvent the wheel), what the important questions are that the programme implementers and wider development community want answered (are they interested in effectiveness? – compare DEval’s Evaluation Programming and Reference Group Model), and how much time and resources are available. RCTs are data-intensive and hence relatively expensive (but not necessarily more so than alternative designs). Designing an under-powered RCT, which has a too little sample size to detect statistically significant effects, is therefore not an effective use of resources. If indeed, as it unfortunately mostly continues to be the case in development programmes, we still do not know whether the type of programme/intervention or some sub-activity of it ‘works’ or has important adverse side-effects, an impact evaluation may be called for. But bear in mind that depending on the effects of interest, these aspects may take time to emerge and to be discernible in the data.

To give an example, agricultural productivity effects will at least take a year to detect (next harvesting season of similar type), while to find out the effect on employability of early childhood development interventions will take one generation (15–20 years). Once we have established that a counterfactual analysis is desired, the next issue to consider is how to establish a counterfactual that best mimics the population that was targeted by the intervention, while taking into consideration what is ethical and feasible in the particular context.

## SPECIFICS OF THE RURAL CONTEXT

There are a number of variants to RCTs that distinguish themselves through the unit of randomisation, the rule applied to assign the population to treatment or control, and the ways in which the treatment is allocated or spaced in time. Each of these will be introduced through an example from the rural development field. It is worth noting that in the databases of the International Initiative for Impact Evaluation (3ie), a large number of RCTs can be found in the rural space.

Based on our experience, there are three issues that distinguish rural RCTs from urban ones. One the one hand, they are easier to implement in the rural context as threats of contamination are relatively low due to limited transmission of information, which is typically contained within the villages. On the other, two issues may be complicating factors in the rural context: first, responses to survey-questions may be more prone to various types of response biases (e.g. social desirability bias), second, depending on the type of intervention, sampling may be more complicated since villages or individual farms are often located in particular micro-climates, soil quality and access to ground water which are hard to detect and measure. In Rwanda, for instance, two apparently identical villages may be in separate valleys within just five kilometres distance, but subject to very different climate conditions. In the rural region around Cochabamba (Bolivia), soil quality at one side of the road may be very different than at the other side.

## RCTs IN PRACTICE

An ongoing 3ie funded study uses individual randomisation to examine the impact of a hybrid risk mitigation financial product that combines credit and insurance, called Risk Contingent Credit (RCC), in rural Kenya. 1,500 households were randomly assigned to receive one of three treatments – (1) the RCC, (2) traditional credit and (3) no credit. The randomisation was done through a public lottery at village level, and the villagers thus knew the treatment status of every participating individual. Individual randomisations are relatively low-cost as sample size requirements are lower. However, such a randomisation within a village faces threats to internal validity. The first risk, namely that of contamination by having individuals ‘switching’ their treatment status, was addressed by making the insurance/credit contracts non-transferable. The other main threat is known as the John Henry effect, when the control group changes its behaviour due to knowledge of what is happening in the treatment group, and in this case, for example, is triggered to seek traditional credit from other banks operating in the region. While the research team may not be able to control this, by collecting information on the credit and source of loan, they will be able to identify and assess the magnitude of this problem.

The second type of example is of cluster-based randomisation, a cluster being a grouping of individuals or households at a level which



A latrine project in Bangladesh. Randomisation has to be adjusted to the level of implementation (e.g. community/village instead of individual level).

Photo: Mirva Tuulia Moilanen/The World Bank

makes sense from a point of view of the intervention and outcome of interest (e.g. village, schools, health centres). Many interventions are implemented for example at the community/village level, with the expected benefits also to be captured at that level, requiring village- instead of individual-based randomisation. The other reason for a clustered approach is the large expected spillovers within tightly knit rural communities, which would entail that other individuals within a community where some members are participating in a programme may also benefit from the intervention (e.g. by watching their neighbours and talking to them). 3ie is funding four ongoing impact evaluations focusing on promoting latrine use among rural households in four different states in India. As all four projects are complex interventions involving social demonstrations, workshops, community events and mixed-media communication, there is a high risk of spillover effects of the interventions among individuals and households in the treatment areas. As a result, all four projects have taken a clustered RCT approach in order to avoid these effects. The Odisha team has randomised at the village level while the Karnataka team is randomising at the Gram Panchayat (the village council, which is the lowest administrative unit in rural areas). The projects have chosen to randomise at different levels given differences in distance.



Where withholding the programme from any group of potential beneficiaries is not desirable, stepped wedge cluster randomised trials can be suitable.

Photo: Jörg Böhling

The previous examples have all been designed to respond to the question ‘does the programme work’ by having control groups that do not receive it. Quite often, however, what you want to test out is modifications to an existing programme to see whether adding a design component leads to improved effectiveness or by comparing different additions to judge which is the more effective innovation. This was the background for a 3ie-funded ongoing multi-arm RCT to test innovative modules of farmer extension services and their effect on agricultural productivity in Cambodia, within the Project for Agricultural Development and Economic Empowerment (PADEE). The authors investigate the impact of two additional features to the traditional extension worker model that provide agricultural advice. First, they assess the impact of incorporating Information and Communication Technologies (ICTs) to overcome extension agents’ low levels of technical education and training. The extension agents are provided with tablets equipped with specialised software with information about soil testing, seed recommendations, fertiliser application, and identification and treatment of crop diseases. Second, the authors test whether performance-based incentives can incentivise extension workers to make use of information available in the software to increase their effectiveness. The authors assess the impact of these features

by randomly assigning a group of 20 villages to each of the following treatment arms: 1) regular extensions services, 2) ICT extension, and 3) ICT plus incentives extension. By measuring the value added of components 2) and 3), they compare the effect of the second treatment to the first and the third treatment to the second and first.

Sometimes, programme implementers are interested in whether the dosage of a treatment makes a difference to the measured outcomes. Stepped wedge cluster randomised trials allow for controlling for variations in timing due to random and sequential crossover of clusters from control to intervention. A 3ie supported impact evaluation in Sudan assesses the impact on incidence and prevalence of moderate acute malnutrition (MAM) in children under five years and pregnant and lactating women of different MAM treatment and prevention interventions. The evaluation design uses variation in the timing of introduction of MAM prevention components (such as food-based prevention, behaviour change communication) and home fortification to localities (clusters) where treatment activities were underway. The impacts are assessed by undertaking a cross sectional comparison across clusters, as well as a comparison over time within the same cluster. This is a good example of a methodology that can be employed for robust causal analysis when baseline data are not available, and where withholding the programme from any group of potential beneficiaries is neither desirable nor feasible.

## ETHICAL CONSIDERATIONS

It is sometimes critically viewed that impact evaluation designs require that only some individuals receive the intervention and this brings up ethical concerns. However, randomisation does not necessarily drive the fact that only some individuals receive an intervention; they are particularly well-suited when for financial or logistical reasons the implementation and roll-out is slow or staggered, or when comparable groups are left out for other reasons. This is the reality of most development interventions. Part of what underlies the ethical concern about impact evaluations is the premise that assignment to a comparison or control group implies ‘not receiving a benefit’. This is not necessarily the case for two reasons. First, the comparison group can be receiving a treatment with which another competing intervention is being compared, as we saw in the multi-arm RCT. Second, it is important to examine the assumption that receiving a

development intervention, or more of one, is always a benefit. The reality is that the effectiveness and impact of a large number of development interventions have yet to be proven. When a genuine state of uncertainty exists about the benefits of an intervention, so that in theory it could be harmful or ineffective, there is an urgent need for it to be critically examined. This state of uncertainty is known as equipoise in the medical literature.

On the other hand, in cases where a programme cannot be implemented across all individuals immediately, randomisation of eligible individuals can in fact be perceived as more ethical and transparent than any other allocation mechanism. While the ethical concerns may sometimes be misplaced or exaggerated for the reasons just described, it is nevertheless critically important to always carefully consider the potential ethical issues that may arise when designing and conducting RCTs.

To summarise, the gaps in knowledge about what works when and where in the rural and agricultural development space (check out the evidence gaps in 3ie’s Evidence Gap Maps) are still immense, and the opportunities to utilise RCT-type impact evaluations to answer effectiveness questions abound.

---

**Marie M. Gaarder** is Director of the Evaluation Office and Global Director for Innovation and Country Engagement at the International Initiative for Impact Evaluation (3ie). She has also held senior positions at the World Bank and Norwegian Agency for Development Cooperation. Marie Gaarder has a PhD in Economics from University College London. Contact: [mgaarder@3ieimpact.org](mailto:mgaarder@3ieimpact.org)

**Sven Harten** heads the Competence Centre for Evaluation Methodology and is Deputy Director of the German Institute for Development Evaluation (DEval). He has more than 15 years of experience as a professional evaluator and most recently worked as Senior Evaluation Specialist at the World Bank (IFC). Sven Harten has a PhD in Political Science from the London School of Economics. Contact: [sven.harten@deval.org](mailto:sven.harten@deval.org)

\* With input from Bidisha Barooah, Neeta Goel, Shaon Lahore, Diana Lopez-Avila, Emmanuel Jimenez, Monica Jain, Tara Kaul, and Francis Rathinam.

For a list of references, links to the studies and further information, see the online version of this article at: [www.rural21.com](http://www.rural21.com)