RANDOMISED CONTROL TRIALS

In a Randomised Control Trial (RCT) different units are randomly assigned to separate groups. One group receives a development intervention and the other does not. Changes in the two groups over time are then compared to accurately measure the effect of the intervention. RCTs have been much debated over the past 10 years. Some see them as the 'gold standard' for impact assessment.

The purpose of a Randomised Control trial (RCT) is to test the extent to which planned changes have been achieved within a target population receiving a development intervention. RCTs are a type of experimental approach. Experimental approaches work by comparing changes in a group that receives a development intervention with a similar group that does not. The difference is then attributed to the intervention.

In some types of experimental approach a comparison group is deliberately formed that is identical as possible to the target population receiving an intervention. These are called quasi-experimental approaches, and are covered in a separate M&E Universe paper. But in an RCT different units (such as individuals, households, organisations, schools or districts) are randomly assigned to separate groups. One group receives an intervention and the other does not.

RCTs are deliberately designed to address *selection bias*. Selection bias occurs when a group receiving an intervention is systematically different in some way from a comparison group. In theory, RCTs avoid this bias by randomly appointing units to a group receiving an intervention or to a comparison group. Within RCTs, a comparison group is always known as a ***control group***, whilst the group receiving the intervention is often called the ***treatment group***.

RCTs have a long history, stretching back over a hundred years, and are best known for testing the effectiveness of medicines. In a medical RCT patients with a similar condition are either given a drug or a placebo (something that looks and tastes like a drug but isn't) based on random assignment. Even the patients and researchers do not know which patient is in which group. This is known as *blinding*. Any potential bias is thereby removed, and the effectiveness of the drug can be assessed by comparing the condition of the patients receiving the drug with those receiving the placebo.

Within social development RCTs were mostly used in research studies until quite recently, and were rarely used for monitoring and evaluation (M&E) or impact assessment purposes. They suddenly became popular in the early 21st century, and some see them as the 'gold standard' for impact assessment. As a result many organisations, including CSOs, have been put under pressure to adopt RCTs in different circumstances – whether appropriate or not. For a while there was a heated debate around RCTs, although that seems to have died down a bit.

How it works

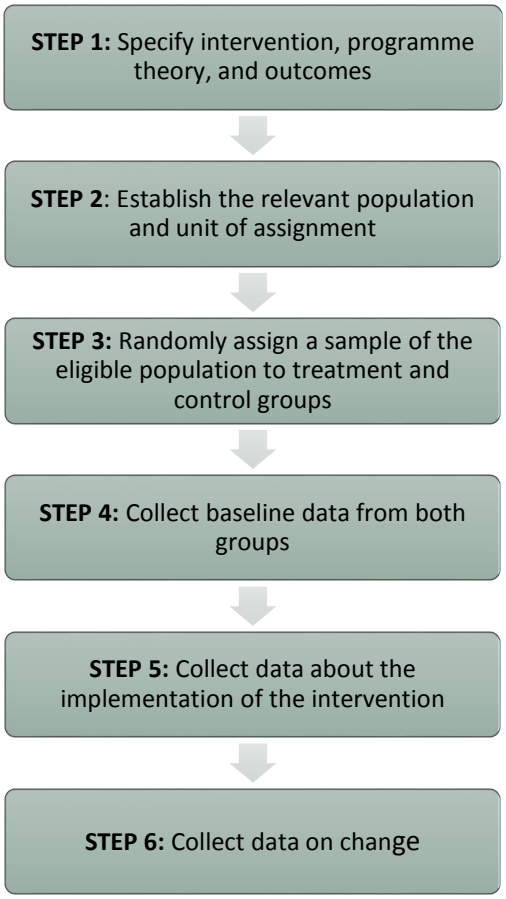
There are different kinds of RCTs, although the steps taken are normally similar (see White et. al. 2014).

- In the most basic form of RCT a treatment group receives a new development intervention, such as financial support or training, and the control group does not.
- In another form of RCT treatment groups may receive different kinds of interventions. For example, one group might receive financial support and another group might receive training. There may still be a control group that does not receive any intervention.
- In a third kind of RCT, different treatment groups may receive different combinations of interventions. For example, one group might receive training, another group might receive financial support, and a third group might receive both training and financial support.

It is important to recognise that people in a control group may still receive existing development interventions. For example, if people are already receiving financial support and an RCT wants to assess the impact of training, the control group could continue to receive financial support.

The basic steps for undertaking an RCT are described below, and in the diagram on the following page. The steps and explanations are mostly based on the guide to RCTs written by White et. al. (2014) for UNICEF. Other guides are available that outline similar steps.

An RCT should always start with a clear articulation of the changes it is hoped will be realised through a development intervention. Usually this involves **developing a theory of change** that articulates what it is hoped will change as a result of the intervention, and how. An RCT works by looking for specific, planned changes, and requires pre-defined outcomes and/or indicators.



STEP TWO

The next step is to **identify the population and the unit of assignment**. Sometimes these are the same. For example, if a project is designed to supply nutritional supplements, children might be randomly assigned to a treatment group that receives the supplements and a control group that doesn't.

However, there are many cases where the unit of treatment and unit of assignment are not the same. For example, it is not normally possible to teach individuals in a school situation without teaching a whole classroom. In this case a cluster RCT design might be more useful. In a cluster RCT design the unit of assignment contains multiple units of treatment. In the example given above, classrooms (the unit of assignment) might be allocated randomly, whilst change would be measured at the level of children (the units of treatment). Equally, it might be more practical to randomly select a number of villages (the unit of assignment) in which to carry out an RCT, instead of randomly selecting individuals (the unit of treatment) across a district.

Some RCTs are even more complex, with randomisation occurring at multiple levels. It is always important, therefore, that an RCT design is clear about the unit(s) where randomisation takes place, and the unit(s) where change is actually measured.

Another issue to consider at this stage is whether change needs to be assessed across an entire population, or whether an RCT also hopes to identify changes amongst different sub-groups, such as boys/girls, people with disabilities or minority groups. This decision is important as it will influence the sample size required for the RCT.

STEP THREE

At this stage the survey sample size needs to be decided. This is quite a complex calculation, known within RCTs as a *power calculation*. In general, the larger the sample size the more certain the findings. However, different factors affect the calculation, and it is normally best left to a trained and qualified statistician.

Once the sample size has been decided, **allocation to treatment and control groups** can be carried out. There are several different ways of randomly assigning a population to different groups.

- One is simple randomisation where different units (individuals, schools, districts, etc.) are chosen at random from a list, using a random number generator.
- Stratified random sampling may also be used (see paper on sampling) to ensure there are the same proportions of different sub-groups (e.g. men and women) in the treatment and comparison groups.
- Another method is called matched pair randomisation, in which different units are matched into pairs with similar characteristics. For example, schools could be listed alongside similar schools in terms of size, location and financial turnover; or individuals could be paired based on gender, age and educational status. One unit from each pair is randomly assigned to the treatment group and the other to the control group.

STEP FOUR

The next step is to **collect baseline data**, usually in the form of a survey. The survey collects basic information (such as household characteristics or socio-economic indicators) from both the treatment and control groups. This helps to ensure that the randomisation process has indeed generated groups that have equivalent characteristics.

In most cases the survey also collects data on the predicted outcomes and/or indicators defined in step 1. For example, an RCT designed to assess the impact of a nutrition programme would measure the nutritional status of children in both treatment and control groups prior to any intervention; whereas an RCT designed to assess an education project might want to assess reading ability. This is in order to make comparisons at a later stage.

STEP FIVE

When conducting an RCT it is considered good practice to **collect data throughout a project or programme**. This is for two main reasons. Firstly, an RCT is designed to measure the effects of a development intervention, but not to explain how those effects came about. Ongoing monitoring of the intervention can help explain how and why changes happened.

Secondly, ongoing monitoring can help identify whether contamination or attrition has occurred, either of which can cause an RCT to produce potentially misleading results. *Contamination* occurs when interventions aimed at the treatment group affect the control group in some way. For example, if an intervention helps raise income within a treatment group, and some of the money is then given to people in the control group, this would be classed as contamination. *Attrition* occurs when units of the population assigned to either treatment or control groups drop out of the RCT after the baseline has been conducted.

The last stage is to conduct a survey on the same treatment and control groups covered by the baseline. This is done to **assess change** within the two groups. The survey should be designed to cover some or all of the same outcomes or indicators that were identified in step 1 and collected via the baseline survey, ideally using the same or similar methodologies. The timing of the repeat survey may vary between different interventions. In some long projects and programmes it may occur partway through; in some it may be at the end; and in some it may be a while after the intervention finishes. The important thing is that sufficient time has elapsed to observe the anticipated changes.

Within RCTs, there are two main methods of calculating the change brought about through a development intervention. The more common is to use the *difference-in-differences* method. In this method the changes against the baseline for the treatment and control groups are calculated separately. Then the impact of the intervention is estimated by calculating the difference between the two.

RCTs without a baseline

Theoretically, it is not always necessary to conduct a baseline when implementing an RCT. If there are enough units then it is possible to assume that the treatment and control groups have similar characteristics. This is because different units are randomly assigned to the two groups, so there is no possibility of selection bias.

If a baseline is not used then the difference-in-differences method cannot be used. Instead, the situation of the treatment group and control group is measured, using the outcomes and/or indicators defined in step 1, and any difference is then attributed to the intervention. This is only valid if the randomisation process generated almost identical groups. The method relies on an assumption that the situations of the treatment and control groups before the intervention were virtually identical.

RCTs are chosen as a research or impact assessment method precisely because they are considered one of the most rigorous methodologies. It is therefore particularly important to record the full methodology and findings at the end of the process. This should cover the theory of change used or developed, the process of implementation, the random allocation method, the sampling strategy, and anything else that would help a decision-maker judge the quality of the study and the importance of the findings.

Strengths and weaknesses

In the right circumstances an RCT delivers an experiment that is free from bias, and capable of clearly measuring the impact of an intervention in a way that most other methodologies cannot. RCTs are regarded as better than quasi-experimental approaches because they eliminate the risk that treatment and control groups are systematically different from each other. At best, RCTs can provide high-quality evidence that can form the basis of decision-making (see case studies below).

Case study: Insecticide-treated Bed Nets

A study by Cohen and Dupas (2008) in Western Kenya used RCTs to assess how prices affected the use of insecticide-treated bed nets (ITNs). 20 similar health centres were randomly assigned to four different treatment groups, offering ITNs at different prices. The control group did not offer ITNs. The RCT showed that when clinics asked for small levels of payment the demand for ITNs decreased, as might be expected. But the data also showed that those who paid more for the nets did not use them any more often that than those who paid less or nothing at all. The findings of the study were used to support the argument for free or subsidised ITN distribution.

Case study: Cash Incentives for Tree Planting

A project in Zambia, supported by a Zambian NGO (Shared Values Africa), aimed to look at the factors that determine how climate-smart agricultural technologies are adopted by small-hold farmers in developing countries. Through conducting RCTs, the project identified the causal relationships between input costs, incentives, and farmer characteristics on the one hand, and outcomes, including programme take-up and tree survival rates on the other. The project found clear evidence that subsidising adoption, and providing cash incentives for sustained effort on land use change, has a positive impact on tree survival rates. Early indications were that the project findings influenced the design of input subsidies in Zambia (CDKN 2013)

However, RCTs are not applicable in many situations, and have some known limitations. Amongst these are the following.

- RCTs are generally considered effective at assessing whether something has worked or not. They are less good at explaining why something has worked. If this is the purpose then they need to be accompanied by alternative methods (Stern et. al. 2012).
- RCTs are conducted in a particular location, context and time. It is not always possible to say with any confidence that the results would be the same elsewhere. This problem can be partly reduced if an RCT is accompanied by a valid theory of change that helps predict and explain why changes have occurred. Alternatively, a number of RCTs can be carried out on the same subject in different times and places.
- RCTs rely on the generation of statistically significant findings. There needs to be a large enough sample size to pick up the effects of an intervention with enough precision (White et. al. 2014). There also needs to be a large enough target population to allow randomisation.

RCTs cannot be done with just a small- or medium-sized number of cases.

- Because they rely on the random assignment of a target population to either treatment or control groups, RCTs must be planned from the start of a project or programme. They cannot be designed or implemented after a development initiative has started.
- RCTs are usually designed to assess how a single cause brings about a specific effect. They are not suited for development interventions operating in more complex settings where change cannot easily be measured, cannot be anticipated at the start of a development intervention, or may be contested.

RCT debates

In theory, RCTs should be treated no differently from any other tool or methodology. They should be used rigorously where appropriate, and not at all otherwise. However, for about a decade RCTs were the centre of intense debates within the social development community. During this time many CSOs believed they were pushed into using RCTs against their better judgement, as a condition of funding from institutional donors. The debates appeared to fall into three main areas.

Firstly, many CSOs over the past decade have been required to carry out RCTs where it was not appropriate to the context, they did not have the skills or resources to carry them out properly, and/or the people demanding RCTs did not properly understand the opportunities and limitations themselves. This is illustrated in a blog by Michael Patton (2014), which he claims represents many other examples, and is backed up by INTRAC's own experiences.

> *"At an African evaluation conference, a program director came up to me in tears. She directed an empowerment program with women in 30 rural villages. The funder, an international agency, had just told her that to have the funding renewed, she would have to stop working in half the villages (selected randomly by the funder) in order to create a control group going forward. The agency was under pressure for not having enough "gold standard evaluations." But, [the director] explained, the villages and the women were networked together and were supporting each other. Even if they didn't get funding, they would continue to support each other. That was the empowerment message. Cutting half of them off made no sense to her."*

Secondly, RCTs have in the past been seen by some as a panacea for all occasions. Yet RCTs are best suited for dealing with quite narrow questions of whether or not a particular intervention with a measurable effect works in a specific location and over a specific timeframe. Some believe that an over-emphasis on RCTs risks ignoring some of the larger issues addressed by CSOs, many of which are concerned with areas such as empowerment, mobilisation, capacity development and policy influencing – areas less susceptible to accurate measurement and therefore less conducive to RCTs.

Thirdly, there are some ethical issues related to RCTs, which are a direct consequence of the randomisation approach. Many tools and methodologies require the formation of control or comparison groups that do not receive the same development interventions as the treatment group. But RCTs are the only methodology in common usage in which stakeholders are randomly assigned to these different groups. Many CSOs consider this to be unethical, and in conflict with their mission to help poor and marginalised people.

In part, this is due to the lack of ethical guidance covering M&E and impact assessment in international development. Barahona (2010) points out that most areas of research involving human subjects have strict codes of conduct and ethical rules, but that these are not always present in social development.

There are two major responses to these arguments. The first is to stress the need for better evidence in social development. White (2011) points out that almost all new interventions are basically experiments, and there should always be a need to test out new interventions. If an intervention is new then it is not known whether it works or not, in which case an RCT does not involve withholding proven benefits from a control group. Ultimately, he argues, it could be considered more unethical to roll out large programmes of work, spending millions of pounds on something that doesn't work. These arguments are more valid for RCTs used to assess new or innovative interventions than those used simply to evaluate existing interventions for accountability purposes.

The second response is more practical. There are many ways in which control groups can be formed that raise less ethical dilemmas (see Ambroz et. al. 2013). The most common of these is called pipeline randomisation. This means the order of entry to a project or programme is randomised, but all stakeholders are included eventually. For example, if a programme is rolled out to 20 communities in each of its first three years, the order can be randomised so that some communities act as the control group in the first year, and then receive the intervention later on.

When should CSOs implement RCTs?

The general advice for CSOs can be summarised as follows (see Stern et. al. 2012, White 2011, White et. al. 2014). An RCT should be considered if:

- there is one primary cause and one primary effect (or outcome) of interest;
- it is possible (and ethical) to develop a control group;
- contamination between the treatment and control group can be controlled;

- there are sufficient numbers to enable statistically significant sample sizes;
- the need is for causal analysis (i.e. what was the effect of an intervention) rather than an explanatory analysis (i.e., how did the change happen);
- there are clear, predicted, measurable changes;
- the RCT can be planned before an intervention begins; and
- the organisation implementing the RCT has the skills, time and resources necessary to undertake the trial properly.

For CSOs, clearly, there is a limited range of interventions in which RCTs will be useful, and most of these probably involve straightforward service delivery. RCTs cannot easily be applied in complex programmes or programmes with multiple levels. RCTs are not useful for projects or programmes where change cannot easily be predicted (e.g. empowerment or mobilisation work) or where it is hard to measure (e.g. capacity development or governance).

INTRAC believes that RCTs are not appropriate when only applied to demonstrate accountability to donors. However, there are times when RCTs can usefully be employed to test the effectiveness of new or innovative projects and programmes, or to compare the impact of different interventions. In these circumstances, the purpose of the RCT should be to inform learning, with clear and visible mechanisms for translating that learning into improved performance. For example, an RCT may be appropriate when making decisions on whether or how to mainstream or scale-up a pilot project.

Where RCTs are carried out they should be properly planned, implemented and resourced, with due concern for ethical standards. In most circumstances RCTs should also be accompanied by alternative, qualitative data collection and analysis methodologies that can help explain any changes identified. Used in the right way, and with the right mix of methodologies, RCTs are a valuable tool that can help deepen understanding of the changes brought about by development interventions.

Further reading and resources

Two other papers in the M&E Universe deal with the related subjects of quasi-experimental approaches and sampling.

The authors of this paper found the most useful guide to RCTs (for beginners) to be the guide written by White et.al. (2014) in the references below. This can be found in many places on the internet.

References

- Ambroz, A; Shotland, M and Siddiqi, H (2013). *Randomized Control Trial (RCT)*. BetterEvaluation. Retrieved September 2013 from http://betterevaluation.org/plan/approach/rct
- Barahona, C (2010). *Randomised Control Trials for the Impact Evaluation of Development Initiatives: A statistician's point of view.* ILAC working paper 13. Institutional Learning and Change (ILAC) Initiative, October, 2010.
- CDKN (2013). *Project Impact Review, AAGL0009k*. Climate and Development Knowledge Network, 2013.
- Cohen, J and Dupas, P (2008). *Free Distribution or Cost-Sharing? Evidence from a randomized malaria prevention experiment.* Brookings Institute Global Economy and Development Working Paper 11. Quoted in Sherbut and Kanji (2012).
- Patton, M (2014). *Week 47: Rumination #3: Fools' gold: the widely touted methodological "gold standard" is neither golden nor a standard.* http://betterevaluation.org/blog/fools_gold_widely_touted_methodological_gold_standard. Retrieved on July 25th, 2016 from the Better Evaluation website.
- Sherbut, G and Kanji, N (2012). *One Size Does Not Fit All: Choosing methods to inform area development*. Draft, 30th April 2012. Article to be submitted to Development in Practice
- Stern, E; Stame, N; Mayne, J; Forss, K; Davies, R and Befani, B (2012). *Broadening the Range of Designs and Methods for Impact Evaluations: Report of a study commissioned by the Department for International Development (DFID), Working paper 38*., April 2012.
- White, H (2011). *An Introduction to the Use of Randomised Control Trials to Evaluate Development Interventions.* Working paper 9. International Initiative for Impact Evaluation, 2011.
- White, H; Sabarwal, S and de Hoop, T (2014). *Randomized Controlled Trials (RCTs): Methodological briefs: impact evaluation no. 7.* United Nations Children's Fund (UNICEF), September 2014.

Author(s):
Nigel Simister and Dan James

INTRAC is a not-for-profit organisation that builds the skills and knowledge of civil society organisations to be more effective in addressing poverty and inequality. Since 1992 INTRAC has provided specialist support in monitoring and evaluation, working with people to develop their own M&E approaches and tools, based on their needs. We encourage appropriate and practical M&E, based on understanding what works in different contexts.