# QUASI-EXPERIMENTAL APPROACHES

Experimental approaches work by comparing changes in a group that receives a development intervention with a group that does not. The difference is then attributed to the intervention. In a full experimental approach, units are randomly allocated to two groups – one that receives the intervention and one that does not. In a quasi-experimental approach non-random methods of assignment are used instead.

Experimental approaches aim to measure the change(s) resulting from a development intervention. This is done by comparing the situation of a target population that received the development intervention with the situation of a similar group that did not. The difference can then be attributed to the intervention. A target population can be comprised of any unit of analysis, such as people, households, communities or organisations.

Experimental approaches work by comparing change with a counterfactual. A ***counterfactual*** is a theory that states what would have happened if a target population had not received a development intervention. In experimental approaches, the counterfactual is developed by assessing the situation of a control or comparison group. These are groups that are as similar as possible to the population targeted by a development intervention, but which were not themselves influenced by that intervention. The counterfactual therefore provides a theory about what would have happened to the target population had the development intervention not taken place.

There are three major types of experimental approach (ActionAid 2016):

- In a full experimental approach, an intervention is provided randomly to units (e.g. individuals, households or organisations) in a target population, thereby creating an intervention group (which receives the products and/or services being tested) and a control group (which does not). Units within the target population are allocated to the intervention and control groups purely on the basis of chance. This kind of experiment is normally called a randomised control trial (RCT). RCTs are covered in a separate paper in the M&E Universe.
- Quasi-experimental approaches are similar in that they compare units that are part of a development intervention with those that are not. However, in a quasi-experimental approach the target population is not randomly allocated to the intervention and comparison groups. This means there may be systematic differences between the groups. This is why it is known as a "quasi" experiment rather than being a true experiment.
- A natural experiment occurs when two groups have already been developed before a study is planned. For example, if a new government education initiative was introduced in one district but not a neighbouring district – dividing school children into those that are included in the new initiative and those that are not – a comparison of progress between the two districts would be seen as a natural experiment.

**Control and Comparison Groups**

Technically, the term ***control*** group should only be used when applying full experimental approaches such as an RCT, where groups have been allocated randomly.

Any other kind of group used to develop a counterfactual should be referred to as a ***comparison*** group. Quasi-experimental methods always use comparison groups rather than control groups, and this convention is used throughout this paper.

However, within the CSO community the two terms are sometimes used interchangeably, albeit incorrectly.

The purpose of an experimental approach – whether full, quasi or natural – is to test whether an intervention has had a measurable effect or not by comparing the situation of an intervention group with the situation of a control/comparison group. This works best when the intervention and the control/comparison group are identical in every way except for the fact that one received a development intervention and the other didn't. In quasi-experimental approaches the groups are not always identical, and therefore they are sometimes seen as less robust than full experimental approaches such as RCTs. However, quasi-experimental approaches are often considered more practical, and far more CSOs have used quasi-experimental approaches than have ever used RCTs.

## *When to use quasi-experimental approaches*

Any experimental approach is best suited to interventions where there are clear, predicted, measurable outcomes. Examples often include projects or programmes in the health, education and livelihoods sectors, which can lead to measurable changes over quite short periods. By contrast, governance or empowerment work may not be as suitable for experimental approaches as the outcomes of such work are often intangible, complex or contested.

Another factor to consider is the complexity of an intervention. Experimental approaches tend to treat an intervention as a single cause. It is therefore easier to apply an experimental approach to a single project than to a complex programme with many different and over-lapping components. The simpler the intervention, and the more measurable the predicted outcomes, the easier it is to apply an experimental approach.

INTRAC's advice to CSO practitioners is to consider using an experimental or quasi-experimental approach when:

- it is possible to clearly measure the intended results;
- the size of the target population is sufficiently large (most experimental models are based around quantitative inquiry, and require large sample sizes);
- it is possible (and ethical) to form an appropriate control or comparison group;
- the CSO has the necessary research resources and expertise, or can afford to buy it in; and
- the benefits of implementing the experimental or quasi-experimental approach outweigh the costs.

Experimental and quasi-experimental approaches can be costly, and are perhaps most useful when there is a clear rationale for using the results – for example if considering whether to scale up or replicate a pilot study, or if attempting to assess whether a particular approach works or not.

The choice of whether to carry out a full experimental approach, such as an RCT, or quasi-experimental approach depends on several factors. Firstly, an RCT is generally acknowledged to be the most robust experimental approach, but it is not possible to carry out an RCT unless it is planned from the very start of an initiative. (This is because an RCT relies on the random allocation of a target population to intervention or control groups.) By contrast, it is sometimes possible to implement a quasi-experimental method even if it was not planned at the start of a project or programme (White and Sabarwal 2014).

Secondly, there are ethical reasons why NGOs might not want to randomly assign people or organisations to intervention or control groups, as this would mean withholding products and services from people simply for the sake of measurement. Under such circumstances a quasi-experimental design would be more appropriate. For example, a common practical method of resolving the ethical issue of withholding treatment is to form a comparison group out of people who are due to enter a project or programme in the future.

Thirdly, there are often practical, political or logistical challenges, such as the need to phase in the geographical roll-out of a programme, which could also prevent randomisation at the start of the programme (ibid). In any case where it is not possible or desirable to carry out an RCT, a quasi-experimental approach may be a valid alternative.

## *How it works*

There are many different types of quasi-experimental approach, some of which are described in the following section. It is beyond the scope of this paper to give a detailed methodology for each of these. Broadly, however, for CSOs, a quasi-experimental approach normally works in the following way.

- The first step is to thoroughly understand the nature of a project or programme, its target population group, and its objectives. A data collection / analysis methodology (or methodologies) is chosen, and a set of indicators is selected that represent the predicted, desired change that it is hoped will be brought about by the project or programme.
- Next, a sample of units (people, households, organisations, etc.) is selected from within the target population to form an intervention group. A comparison group is also selected. The comparison group should be as identical as possible to the intervention group. Often this means choosing units from similar locations to the target population, or with similar profiles.
- If the quasi-experimental approach is designed to use quantitative methods of analysis then a power calculation needs to be performed. This calculation is used to determine the sample sizes needed to be able to detect the expected differences between the two groups.
- In most circumstances a baseline is carried out on the intervention and comparison groups. The baseline should be conducted using the same methodology and in the same way for both groups. For example, if a survey is used then the survey should ask the same questions to the intervention and comparison groups, and in the same order, to avoid any potential bias in the responses.
- A second round of data collection is then applied after the project or programme has finished (or partway through) using the same tools or methodologies as for the baseline, and applied in the same way. Ideally, the same representatives of the target population and comparison group should be re-contacted, although this is not always practically possible.
- Finally, results are compared across the two different groups, and the difference is attributed to the intervention (the project or programme). Sometimes, results can be disaggregated by different factors such as age, gender or disability.

In some circumstances a retrospective baseline may be developed. This can sometimes be done through using secondary data sources, such as previous surveys that have been carried out in the same area, or government statistics. It can also be done by asking people to recall what the situation was prior to an intervention. If a retrospective baseline is used there is only one round of data collection, rather than two.

It is very important when using any kind of experimental approach to report on all the details of the exercise, including information on group formation, sampling, data collection and analysis. Any assumptions made also need to be clearly recorded. This is because the success of an experimental approach – as with any experiment – depends heavily on the process itself. It is therefore vital that anyone reading a report based on an experimental approach can clearly see how results were generated, and how they led to any conclusions or recommendations.

## Different types of quasi-experimental approaches

In many ways, the key challenge of quasi-experimental approaches is how to avoid (or mitigate) the selection bias. This is the fear that those who are in a target population affected by an intervention are systematically different as a group from those who are not. If so, any difference in results between the two groups may be due to these systematic differences rather than the intervention (White and Sabarwal 2014).

There are many different techniques for creating a valid comparison group. Four are most commonly mentioned in development literature, and these are described below.

- Non-equivalent Groups Design (NEGD)
- Propensity Score Matching (PSM)
- Regression Discontinuity Design (RDD)
- Reflexive Comparisons

***Non-equivalent Groups Design (NEGD).*** NEGD is probably the most frequently used quasi-experimental approach used in the social sciences, and is certainly the most common method used by CSOs. The aim is to identify comparison groups that are as similar as possible to the target population. But the comparison groups normally exist as groups prior to the development intervention. For example, if children in a school or classroom form part of a target population than a comparison group could be developed from a similar school or classroom. If working with communities then a comparison group could be developed from people living in a village that is not targeted by a project or programme but is very similar to one that is (see Trochum 2006).

The main drawback of NEGD is that it is never possible to be sure that the intervention and comparison groups are entirely similar, which is why studies based on NEGD are often less reliable and require more careful interpretation than studies based on RCTs. In other words, prior differences between the groups could affect differences in results measured at a later date. Studies based on NEGD almost always consist of a baseline and follow-up, so differences at baseline can be assessed as well as differences during or after an intervention.

***Propensity Score Matching (PSM).*** PSM is a common method used to select a comparison group after data collection has taken place. It attempts to directly match individual units (individuals, households, organisations etc.) that have received an intervention, with those that have not. Ideally, it would be possible to directly match units according to different characteristics. For example, if a household in an intervention group consisted of a husband and wife, aged between 30-40, with two boys and a girl, and owning two hectares of land and three cows, then it would be ideal to have an exactly similar household in the comparison group.

In practice this is not usually possible to do. Instead, PSM uses a set of statistical analysis techniques to create a comparison group that is as similar as possible to the sample in the target population, across all the different characteristics (see Banerjee and Duflo 2009). This results in the formation of two groups that have similar average characteristics. The assumption is that the groups are therefore close enough that results will not be biased.

***Regression Discontinuity Design (RDD).*** RDD can only be used when the target population is selected based on meeting a certain threshold (for example, if people only qualify for a project if they are living on less than $1 a day, or have a body-mass index (BMI) of less than 16). In this case, those above and below the threshold may be very different. So, for instance, if looking at prevalence of diseases, a set of people with a BMI of less than 16 could not reasonably be compared with a comparison group of people with a much higher BMI.

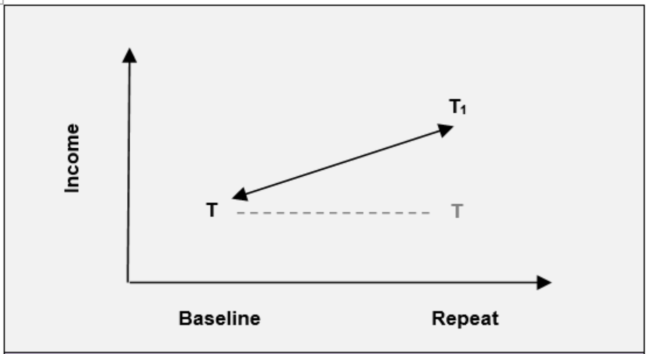
The answer is to compare units that lie just either side of the threshold. For example, if the threshold of inclusion in a project is living on less than $1 a day then people living on $0.98-0.99 a day (who qualify for a project) are probably not much different from people living on $1 or $1.01 a day. Therefore a valid comparison group could be formed of people just above the threshold (White and Sabarwal 2012).

***Reflexive comparisons.*** In a reflexive comparison study, there is no comparison group. A pre- and post- test (baseline and repeat study) is done on a set of units, and the change between the two is attributed to the project intervention. The rationale for calling this a quasi-experimental study is that the units act as their own comparisons. For example, in a project looking to improve farmers' crop yields a sample of the farmers at baseline will not have received any inputs, and can therefore be a comparison group for the same sample of farmers afterwards (see Banerjee and Duflo 2009).

Many CSOs use baselines and follow-up studies to assess change. However, many would be surprised to know that some consider these to be quasi-experimental designs! The main criticism of reflexive comparisons is that they are often unable to distinguish between changes brought about by an intervention and changes due to other effects.

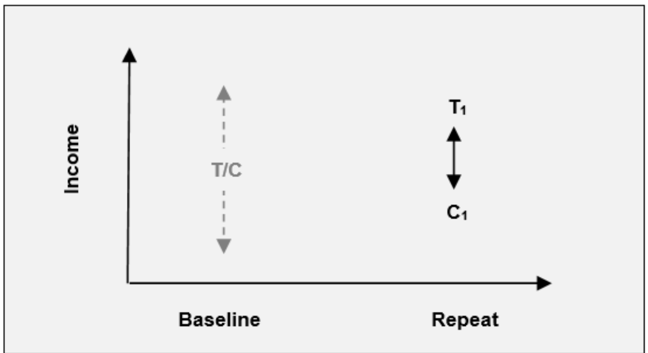
## Analysing the results of a quasi-experimental study

In a simple study using a baseline and follow-up (see reflexive comparison above), the change attributed to an intervention is assumed to be the situation after the development intervention compared to the situation beforehand. The diagram below shows an example taken from a project that is intended to increase the disposable income of households. The result of the intervention is calculated as the situation at the end of the project (the average income of the households) minus the situation at the beginning.



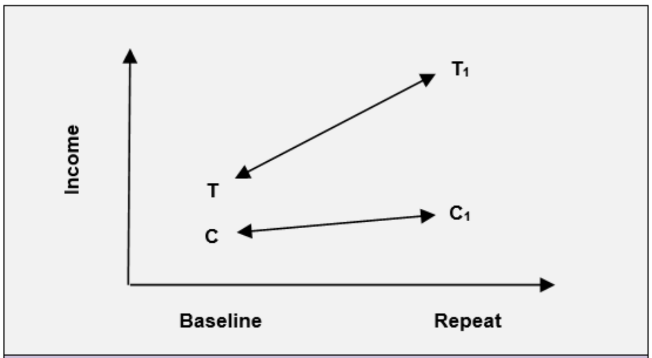

The target population is measured at the baseline (T) and again during the repeat study ($T_1$). It is assumed that any changes to the target population were caused by the intervention. The change attributed to the intervention is therefore the income at $T_1$ minus the income at T.

When using a study with a control or comparison group that is extremely similar to intervention group, as is the case with some RCTs, it can be assumed that the baseline would be the same for both groups. In these cases the result of an intervention is calculated as the situation in the intervention group minus the situation in the comparison group. No baseline is necessary




The change attributable to the intervention is always the situation in the target population ($T_1$) minus the situation in the control/comparison group ($C_1$). This will apply wherever the baseline is, so there is no need to do a baseline.

Where there are doubts about how similar the two groups are – as is always the case in NEGD, and often the case in PSM or RDD – a system known as ***difference-in-differences*** is used. The difference-in-differences approach first compares the change against the baseline for the target population. It then compares the change against the baseline for the comparison group. Finally it estimates the result of the intervention as the change in the situation of the target population minus the change in the situation of the comparison group.




The target population is measured at baseline (T) and again during the repeat study ($T_1$) and the difference calculated. The comparison group is also measured at baseline (C) and again during the repeat ($C_1$) and the difference calculated. The final change is calculated as the difference between $T_1$ and T minus the difference between $C_1$ and C. Hence the name 'difference-in-differences'.

## Strengths and weaknesses

The main strengths of quasi-experimental approaches were described earlier. They can provide evidence of change that is more robust than evidence produced without a control or comparison group. They allow CSOs to develop a counterfactual – an estimate of what the situation would have been without the intervention. They can be planned and applied after an intervention has started – unlike RCTs – and can be used in situations where full experimental designs cannot. They are often easier to set up than RCTs, and may require less expertise and resources (although this is not always the case).

However, they also have several shortcomings.

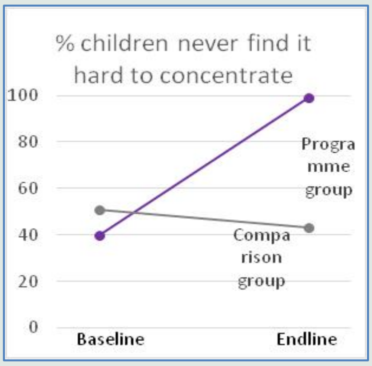
- Firstly, in common with RCTs, quasi-experimental approaches attribute changes directly to interventions without considering how the change was produced. Therefore they are unable to always provide explanations of how change came about (Stern et. al. 2012). This challenge can be partly resolved if alternative, more explanatory methods are used alongside the quasi-experimental approach.
- Quasi-experimental approaches can help answer the question of what changed over a specific time and in a particular environment. But because they do not investigate how or why changed happened, they cannot always be used to make wider generalisations. Again, this challenge can be resolved by using additional methods where appropriate.
- In common with RCTs, quasi-experimental approaches tend to suit interventions where there is a clear, logical link between cause and effect, and where effects are designed to be achieved over short- to medium-term time spans. Yet many CSOs carry out work in highly complex environments, where contribution rather than attribution is considered key, and where links between cause and effect are not always linear.
- Quasi-experimental approaches can be difficult and costly to apply, and are often more complex to analyse and interpret than RCTs. This means specialist

expertise may be needed. Quasi-experimental approaches tend to be based around the collection of data from large numbers of individuals and households, sometimes over long-time periods. Smaller CSOs may need to buy-in specialist knowledge to design and run them, or may not have the resources to implement them at all.

INTRAC believes that there are times when experimental and quasi-experimental approaches are very useful. In these cases, they should be done properly, with appropriate expertise, and with whatever resources are needed. But as with any tool or methodology, the limitations as well as the opportunities should be thoroughly understood beforehand.

**Case study: Malawi school meals programme**

Mary's Meals works in a quarter of all schools in Malawi, providing nutritious lunchtime meals to school children with the aim of increasing school enrolment and attendance, as well as supporting children's learning outcomes and enjoyment of school. An impact study was developed with support from INTRAC. An RCT was not appropriate because randomising pupils receiving meals was neither ethically nor practically feasible. Instead, a Non-Equivalent Groups Design was developed where 10 programme schools were matched with a group of 10 similar schools in a neighbouring district. Mary's Meals had plans to roll out the programme in the comparison schools, although respondents were not made aware of this (to avoid them giving the answers they hoped would result in a school meals programme).




A baseline was conducted via a face-to-face survey with children, teachers and community members to gather data for different indicators at output and outcome levels. This was repeated at two time-periods after the programme started. Sample sizes for children were around 350 in each group at each time period. A difference-in-differences approach was used to analyse the results, and statistical tests used to estimate the effect sizes and the significance of observed changes. Because respondents were clustered in schools, further tests were carried out at school level.

Significant differences and medium to large changes were recorded for key outcomes, such as classroom hunger, the ability to concentrate, whether children were joining in lessons, and whether they wanted to go home early. The exception was children's reported happiness at school, which showed more variation at baseline and much smaller changes (although still statistically significant) because of the programme – indicating that other factors had a greater influence over happiness. The cross-over shown in the chart (see above) means Mary's Meals could be relatively confident that differences between the intervention and comparison groups at the end of the programme were not due to biases in initial selection. The quantitative analysis was supplemented by qualitative methods that helped to support or refute the findings of the study.

**Source: Mary's Meals (2016)**

## Further reading and resources

Another paper in the M&E Universe series deals with Randomised Control Trials (RCTs). There is also a paper on the related topic of sampling.

**Randomised control trials**

**Sampling**

The Research methods Knowledge Base has a website devoted to many different forms of data collection and analysis, and there are many pages dealing with quasi-experimental approaches. The first of these can be found at http://www.socialresearchmethods.net/kb/quasiexp.php. The paper by White and Sabarwal (2014) referenced below is a good guide to different methods, although quite technical for people with a non-statistical background.

## References

- ActionAid (2016). *Evaluation Technical Briefing Note #9; How to choose evaluation methodological approaches (draft)*
- Banerjee, A and Duflo, E (2009) 'The Experimental Approach to Development Economics', *Annual Review of Economics* 1:1.1-1.28.
- Mary's Meals (2016). *Malawi Impact Assessment, Year one report.*
- 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.
- Trochum, W (2006). *Research Methods Knowledge Base. The Non-equivalent Groups Design.* *http://www.socialresearchmethods.net/kb/quasnegd.php*
- White, H and Sabarwal, S (2014). *Quasi-Experimental Design and Methods.* Methodological briefs, impact evaluation No. 8. UNICEF, September 2014.

*Author(s):*
*Dan James, Anne Garbutt and Nigel Simister*

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.

**M&E Training & Consultancy**

**INTRAC's team of M&E specialists offer consultancy and training in all aspects of M&E, from core skills development through to the design of complex M&E systems.**

**Email: info@intrac.org** **Tel: +44 (0)1865 201851**

**M&E Universe**

**For more papers in the M&E Universe series click the home button**