Measuring Attribution: a practical framework to select appropriate attribution methods, with cases from ALCP in Georgia, MDF in East Timor, Propcom Mai-Karfi in Nigeria and Samarth-NMDP in Nepal

Synopsis

Measuring impact in private sector development programs is important but also challenging. This guidance paper provides an overview of the most common attribution methods, and offers guidance on how to select the most appropriate attribution method for the diversity of interventions and their context. This paper also documents how four programs have selected and implemented four different attribution methods.

Authors: Hans Posthumus and Phitcha Wanitphon

Date: August 2015

Acknowledgements:

This is one of ten cases that have been developed by Hans Posthumus Consultancy[1]. The preparation of these cases was supported by funds from the Swiss Agency for Development and Cooperation (SDC), provided through the DCED Trust Fund. We would like to thank them for providing the opportunity to work on this case. The cases we describe are drawn from the Alliances Lesser Caucasus Programme (ALCP) in Georgia, Market Development Facility (MDF) in Timor Leste, Propcom Mai-Karfi (PM) in Nigeria and Samarth Nepal Market Development Programme (Samarth-NMDP) in Nepal to which we are indebted. We are grateful, in particular, to Helen Bradbury and Zakairia Tavberidze from ALCP, to Mujaddid Mohsin and Samira Saif from MDF, to Belinda Boateng, Collins Apuoyo, Nurul Azam and Oluwatosin Ariyo from PM, and to Tim Stewart and Sanju Joshi from Samarth-NMDP for their valuable contributions. We would also like to thank Aly Miehlbradt for her valuable input.

This case describes how the programs have addressed a typical challenge in results measurement. The aim of the case is to provide insights that will be useful to other practitioners facing a similar challenge. The authors do not represent the DCED or SDC, nor do the views expressed in the case necessarily reflect the views of the DCED or SDC.

[1] The HPC consortium was led by Hans Posthumus (HPC) and consisted of Aly Miehlbradt (MCL), Ben Fowler (MSA), Mihaela Balan, Nabanita Sen (OU), Phitcha Wanitphon and Wafa Hafiz (H&S)

Table of Contents:

1. The challenge of measuring attribution

Measuring the impact in Private Sector Development (PSD) programs is important but also challenging. The DCED Standard for Measuring Results defines attribution as '*the ascription of a causal link between observed (or expected to be observed) changes and a specific intervention',* and control point 4.1 requires that '*attributable changes in all key indicators in the results chains are estimated using methods that conform to established good practice'*[2].

Many PSD programs have taken an indirect approach to achieving sustainable impact at scale. They often aim to enable their target beneficiaries to increase their income, or they aim to create employment opportunities. Many factors influence these and other higher-level impacts that a program aims to achieve for its target beneficiaries. External factors, for example the weather, the use of other inputs and services, and market dynamics may have an effect on the intermediate changes that exist between the activities of the program and the final impact on the target beneficiaries.

Measuring the changes that are attributable, or not, to the interventions is crucial. We only know that our intervention is successful if we are able to link the changes we measure to our activities. If we don't deal with attribution, we are not able to use the information to inform our management decisions. If we don't deal with attribution, the impact that we are reporting will not be perceived by the donors as credible impact achieved by the program. Programs are not aiming for scientific rigor but for plausible attribution. The measurements must be practical and doable, ensuring it can and will be done. Most importantly, they should be used to inform management and donors to allow them to take decisions.

This guidance paper provides an overview of the most common attribution methods, and offers guidance on how to select the most appropriate attribution method to assess the impact at the target group level for the diversity of interventions and their context[3]. This paper also documents how four programs have selected and implemented four different

[2] The DCED refers to the OECD-DAC definition
[3] The guidance is provided by the authors, and describes how they would select the most appropriate attribution method.

attribution methods. This paper focuses on attribution: it does not document other important elements in measuring such as monitoring intermediate changes in the early stages of the intervention and sampling of baselines and end-lines, for which reference is made to other guidance papers and cases on the DCED website.

2. Attribution methods

The results chain is the first important instrument that helps us to determine whether higher impact level changes are due to our activities. For each intermediate result in the results chain, between the activity level and the impact level, we verify whether and why this change is due to the previous change or not. If it is, causality continues. If it is not, the causality ends and there is no attribution. This verification is, however, not a simple yes or no, but involves a quantitative aspect. How many target beneficiaries are using the service (scale) and how much of the change (depth) is due to the activities?

As visualized below, the final impact (in numbers) is in practice often less as we only continue to the next intermediate change with '*those that benefited from the previous change*'.

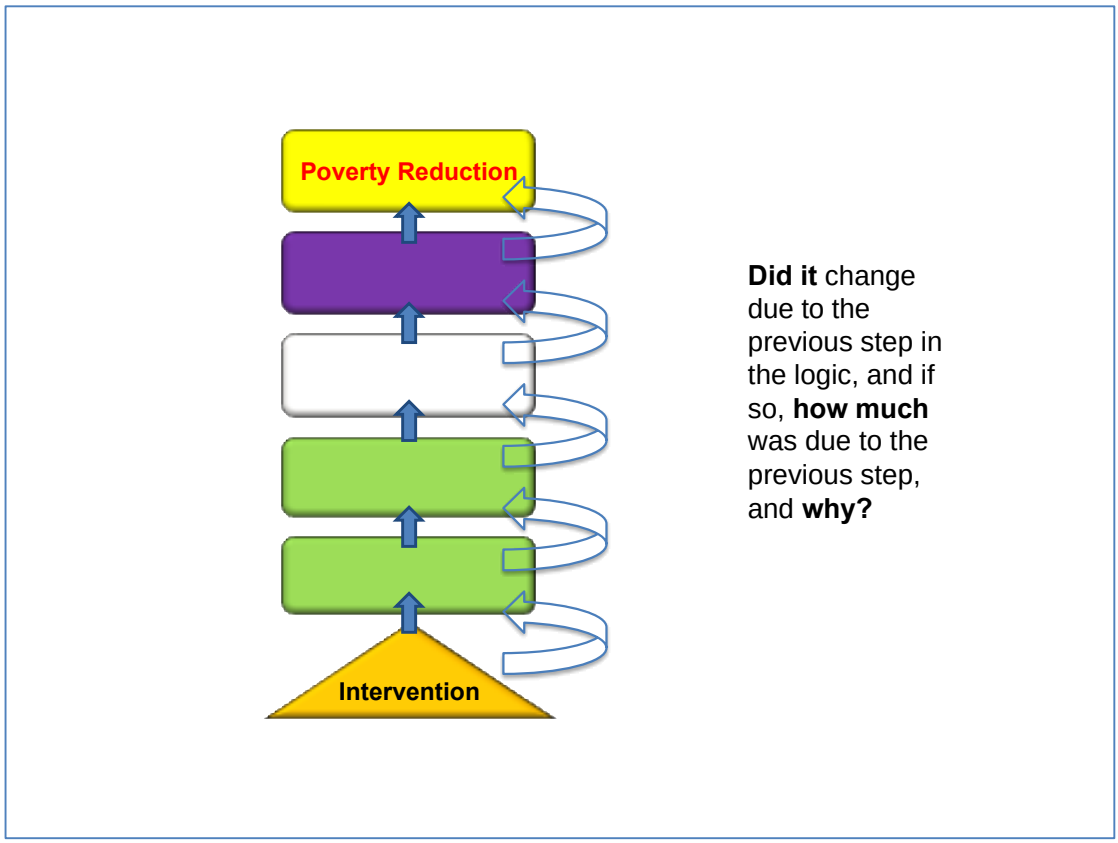

Figure 1 Causality

For those that benefited, we need to assess the depth of the change which is the change due to the intervention. The main challenge is to estimate the counterfactual. The counterfactual, as illustrated in Figure 2 below, is 'what would have happened had we done nothing'. Because we did intervene, we often don't know what the counterfactual would have been. There are various options to determine which attribution method is the most appropriate for the specific intervention, given its context.

Figure 2 The counterfactual

In some cases it is not that difficult to estimate the counterfactual at all, but in many other cases it is a bit challenging to estimate the counterfactual. However, thinking carefully through the intervention logic, taking the context into account, selecting the most appropriate attribution method before constructing baselines, can often lead to doable and informative impact assessments. Below we give an overview of the most common attribution methods and when they can be applied.

2.1 Before and After Comparison (BAC)

This method assumes that there are no other influencing factors; in other words, the counterfactual is simply the base-line situation. The impact is simply the difference between the end-line and the base-line (A-B), as visualized in Figure 3.

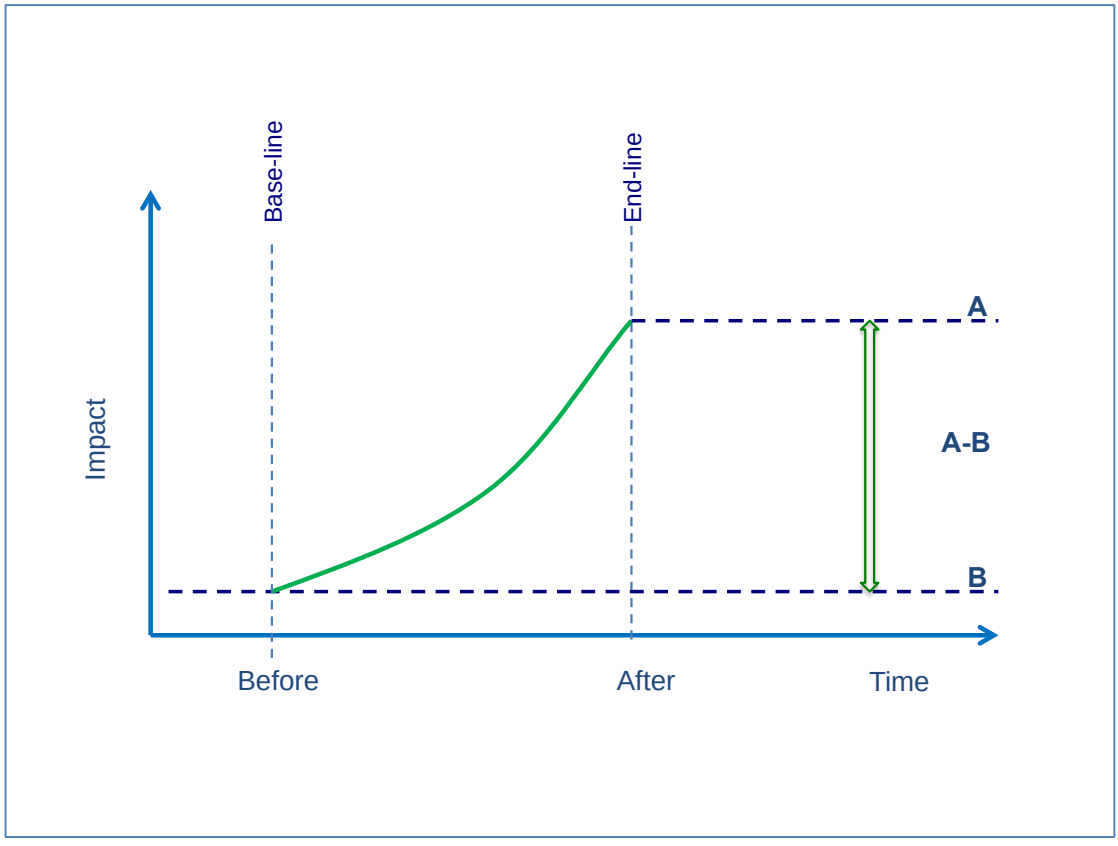

Figure 3 Before and After Comparison (BAC)

This is often found to be the easiest method to apply, yet it is often also the most questioned method. In many programs, interventions lead to changes at the lower (output or market uptake) levels and in most cases there are various intermediate changes between the activity and the impact level that may be influenced by other factors.

In only a few cases there are no other influences, for example when introducing a new crop that is grown on previously fallow lands. Any income derived from this new crop is due to

the intervention, irrespective of the actual yields that are of course influenced by the weather or other factors.

In most cases however, we want to understand how and why this change is happening. A simple BAC is then not enough, but if we include questions to seek the opinion of respondents on the reasons for changes, it is more informative and we call this a Before and After Comparison with Opinion (BACO) as explained below.

2.2 Before and After Comparison with Opinion (BACO)

Similar to the Before and After Comparison (BAC), this Before and After Comparison with Opinion (BACO) method also assumes there are no, or very few influential factors, and that these can be neglected or isolated. The main difference is in the additional 'opinion'. Besides the quantitative information we obtain on the base- and end-line of the beneficiaries, we seek additional qualitative information to confirm (or not) that the changes between the before and after situation are due (or not) to the intervention. If only a portion of the change is due to the intervention, we may use the opinion of the respondents (and key informants) to estimate that portion and to correct the total change.

Often, there are one or two key influencing factors that can be easily isolated, by '*holding them constant*'. For example, the additional income in a manufacturing plant may be due to an intervention to improve productivity in manufacturing furniture, but additional income is also affected by fluctuating market prices (due to other reasons not related to increased productivity). In that case the difference in productivity is assessed, and the additional income is based upon the price and costs at one point in time, rather than at the time of the base- and end-line, when market prices would be different and thus influence the total change.

The Market Development Facility (MDF) in Timor Leste applied this method to measure their impact with the Acelda intervention.

2.3 Quasi Experimental Design (QED)

Quasi Experimental Design (QED) can be applied when the target beneficiaries have potentially benefited, but when a similar group of potential beneficiaries did not have access to the intervention. In that case, the end-line of the non-beneficiaries is equal to the end-line of the beneficiaries if they had not benefited. In other words, the non-beneficiaries represent the counterfactual.

An important precondition to be able to apply the QED, is that both groups, the ones that benefited from the intervention (**the treatment group**) as well as the ones that did not have access to the intervention (**the comparison group**), are exposed to the same influencing external factors. The difference between them represents the change due to the intervention, as shown in Figure 4.

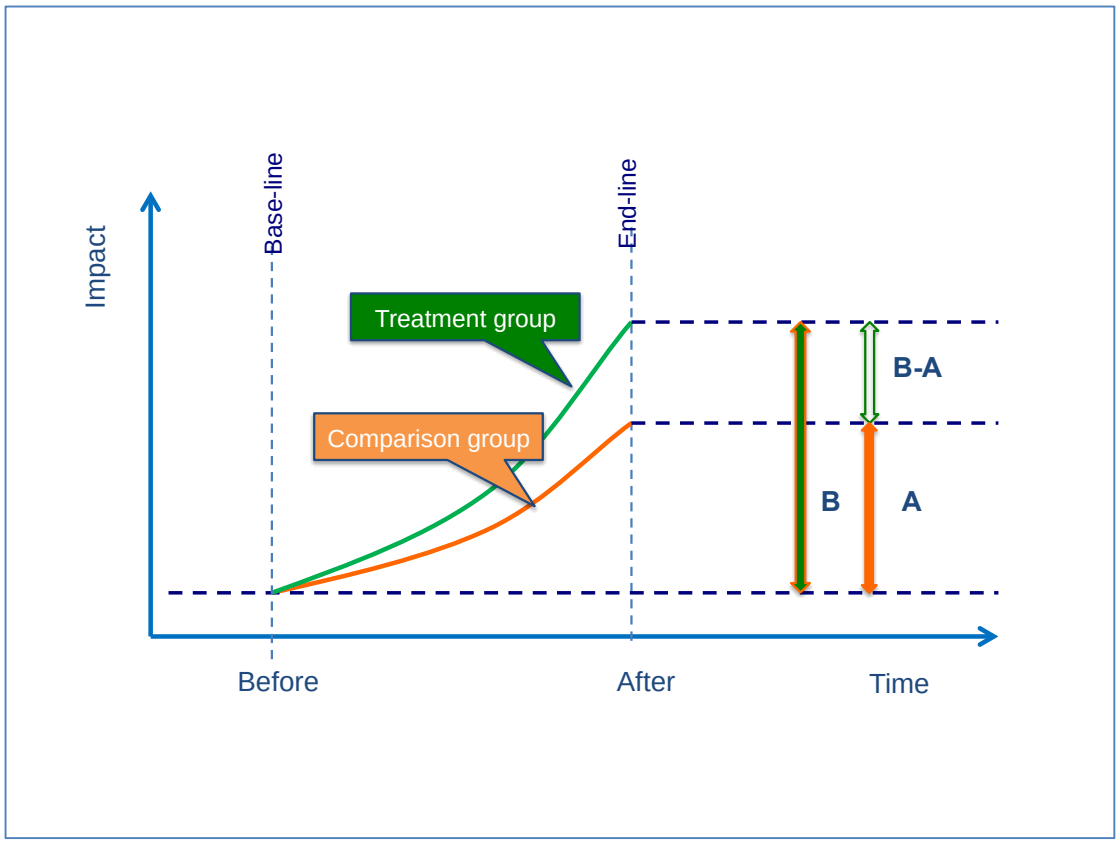

Figure 4 Quasi Experimental Design (QED) – same base-lines

In some instances, the treatment and comparison groups are not identical. As long as they react identically to the influencing factors, the counterfactual can be calculated in the same way as shown in Figure 5.

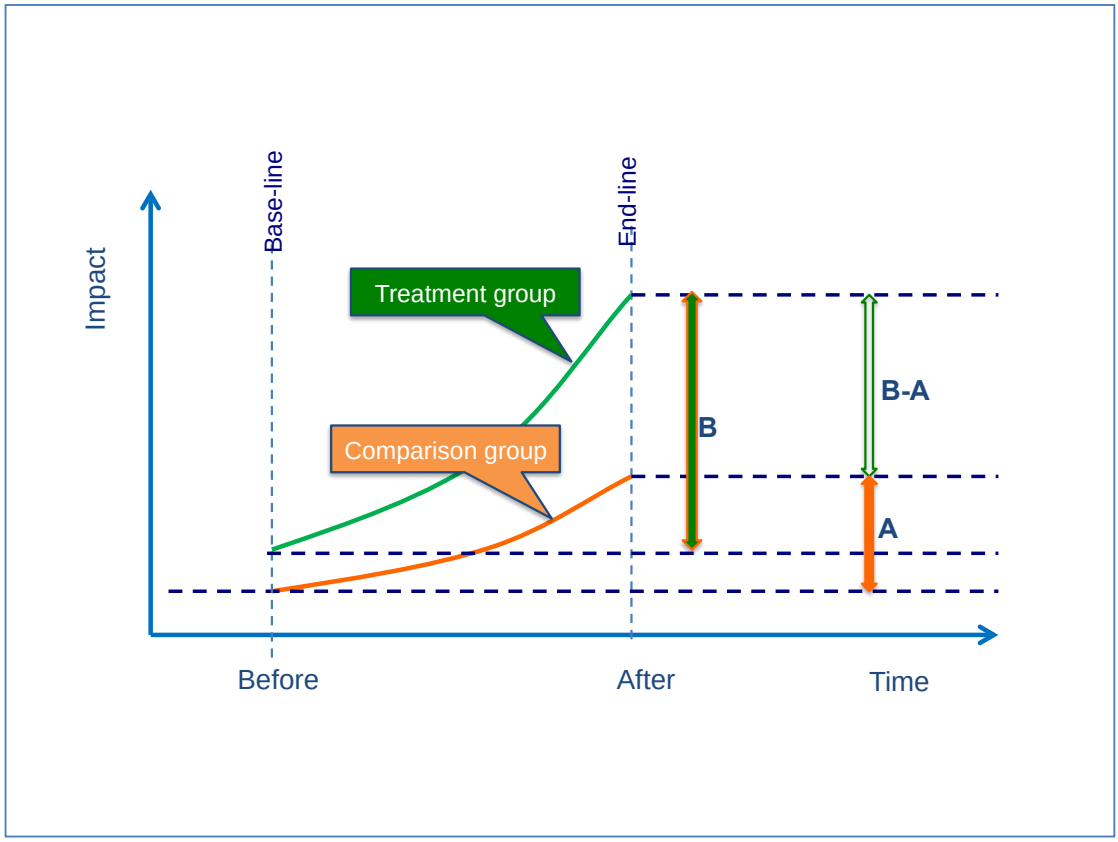

Figure 5 Quasi Experimental Design (QED) – different base-lines

In this case, the Difference-in-Difference (DiD), the difference between the after and before situation of the treatment group (B) minus the difference between the after and before situation of the comparison group (A) represents the impact due to the intervention (B-A). However, the assumption that both groups react identically to the same external factors is an important assumption that needs to be verified.

QED is used when we know in advance who will be exposed to the intervention, and who will not, thus creating our **treatment and comparison groups** respectively. In market development programs, we often don't know in advance who will benefit which makes it difficult to construct the baseline and presents a challenge.

The allocation into either **treatment or comparison group** is not random and therefore implies a risk that both groups are not identical and do not react identically to the external factors. It is important to ascertain how distorted the comparison might be. If the distortion is limited, in other words, the change experienced by both groups due to external factors is much bigger than the difference of change between the treatment and comparison groups, we can accept this small 'error'.

For example, one intervention introduces new seeds but yields vary enormously every season due to the weather. Suppose the treatment group consists of more mature farmers and the comparison group consists of less mature farmers, and the ability of the more mature farmers to apply new seeds leads to slightly higher yields than the less mature farmers applying the same new seeds. We can accept this and describe the method and its limitations in the report. If however the change due to external factors is much less and the

difference in change between the two groups much bigger, we shouldn't use this comparison group.

The Samarth Nepal Market Development Programme (Samarth-NMDP) applied this method to measure the impact of one intervention in the ginger sector.

2.4 Comparison Groups (CG)

The Comparison Group (CG) method is very similar to the QED method, estimating the Difference in Difference (DiD) between those that benefited (**the user-group**) and those that did not benefit (**the non-user group**)[4]. We apply the Comparison Group method when we don't know in advance who will benefit and who will not benefit. The allocation of respondents is thus only done during the end-line situation, when **users** are allocated into the **users-group** and **non-users** are allocated to the **non-users-group**.

The main complicating factor is the potential self-selection bias. Self-selection occurs if some potential beneficiaries decide to apply, while others decide not to apply. The reason for their different reaction is due to a characteristic that distinguishes one from the other. This implies that respondents in both groups do not react in the same way to the intervention.

As with QED, it is important to judge the potential difference in impact between the **user-group and the non-user group**. For managers, it is often crucial to find out if and why these two groups react differently to the same business model. It informs management why the innovative business models work for some beneficiaries and not for others.

Propcom Mai-Karfi (PM) in Nigeria applied this method to measure the impact of one intervention in the tractor market.

2.4 Comparing Trends (CT)

Analyzing trends, and especially changes in trends, can help to assess the impact of an intervention. In some cases, the interventions will affect all respondents, making it impossible to compare the potential beneficiaries with others that did not benefit. If we do have sufficient historical data, for example for a time period spanning several years or even decades, we can compare the trend (inclination) after the intervention (Beta) with the trend before the intervention took place (Alpha) as shown in Figure 6. The difference between them is thus due to the intervention.

[4] We use the terms treatment- and comparison group for QED and user- and non-user groups for CG

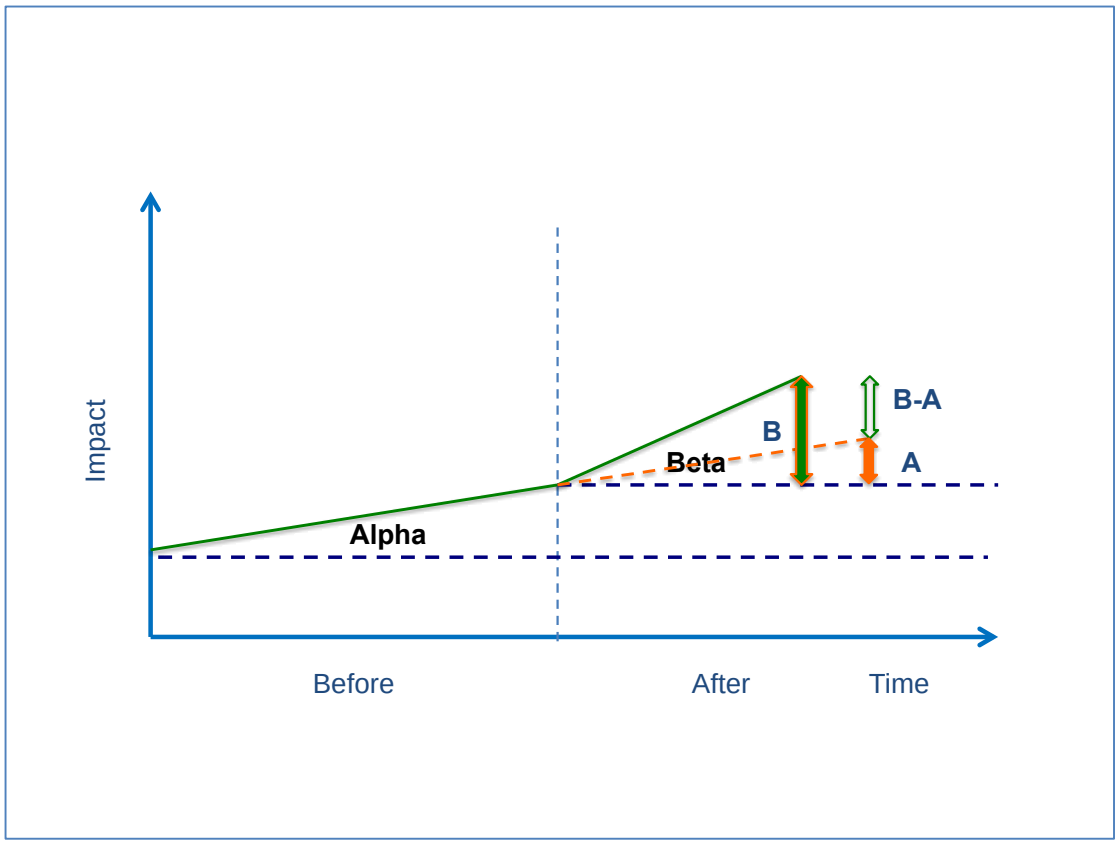

Figure. 6 Comparing trends (CT)

There are a few limitations and implications when using this method. The data we use is then often secondary data, implying the need to verify whether it is reliable, while at the same time it implies that the data are often more aggregated and less specific than we would often need. For example, export volumes of exported fruits are probably easier to obtain than detailed information on the quality and variety of these exported fruits.

Statistical information is also often not very up to date which means that we have to wait and can't assess the impact. Changes in trends can't be assumed to be due to the intervention only. Often, changes in the trends are caused by external factors. Any trend analyses should thus seek the opinion of key informants, beneficiaries, market actors, researchers and others to investigate, and if need be, take into account, the influence of other factors on the end-line situation.

Obviously, we can only use trend analyses if there is a trend meaning a reasonably steady pattern of change over a number of years, not when annual numbers are fluctuating enormously and there is no trend. Trend analyses are thus not much used when assessing impacts. It is rare that an intervention will affect all respondents at the same time and to the same degree and we often lack the relevant historical data and information to differentiate between the change due to the intervention and that due to other factors.

2.5 Regression Analyses (RA)

If one understands the relationship between two variables and can estimate by how much a certain dependent variable would change (for example, the yields of a crop) when the independent variables change (for example, the use of fertilizer and the application of

certain cultivation practices), one can use this relationship to estimate changes of the dependent variable (for example the change in yields) by measuring the independent variable (such as the use of fertilizer and cultivation practices).

Figure 7 shows the relationship between two variables

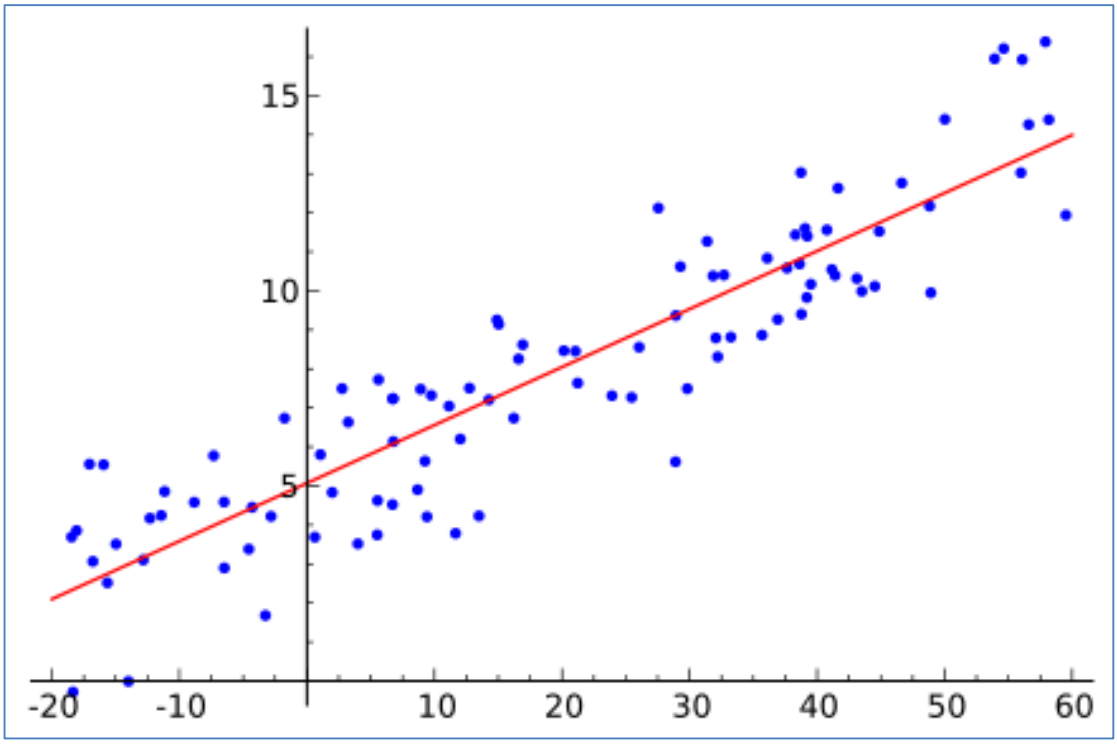

Figure 7 Regression Analyses (RA)

Each blue dot shows where a particular sample is: one independent variable is plotted on the horizontal axis, the dependent variable is plotted on the vertical axis. With sufficient data (blue) points, the general relationship between these dependent and independent variables can be expressed as the red line.

Once it has been demonstrated that a change in one variable causes a change in the other, then future measurement becomes easier. If the programme has demonstrated that a certain increase in fertilizer application leads to a certain increase in yield, in future it could just measure the change in fertilizer application, which is cheaper and easier to assess.

However, in order to understand and obtain sufficient certainty in these relationships, we need to do rather intensive research. This investment pays off if a program plans to be active in one or two sectors (crops) for a longer period.

2.6 Randomized Control Trial (RCT)

A Randomized Control Trial (RCT) randomly allocates (potential) beneficiaries into either **the treatment group** or **the control group**[5]. In other words, the program then decides who will

[5] We use the terms treatment- and control group for RCT, treatment- and comparison group for QED and user- and non-user groups for CG

benefit and who will not benefit (comparable to flipping coins), in order to be able to compare the two groups and assess the impact. There are implications that need to be considered if we wish to use RCTs.

First, denying access is often considered unethical. The target beneficiaries, often the poor, potentially benefit from the access. Providing access to some (who as a result may become less poor) and denying access to others implies that the others remain poor for the sake of measuring.

Secondly, providing access to some and denying access to others is also often impossible in market development programs. The program is not in control, but the market actors decide. The program supports market actors in creating a new business model, offering, for example, a new service or input that triggers target beneficiaries to take it up (or not).

There are not many PSD programs that have applied RCTs, and more information on why RCTs are not so much applied, is discussed in the paper 'Why randomized control trials really are the 'gold standard' for private sector development' produced by the ILO Lab.

3. Selecting the appropriate attribution method

The specific context of an intervention determines what the most appropriate method is. Figure 8 shows a selection aid that can be used to decide which of the commonly used attribution methods is the most appropriate.

This paper does not discuss the management decision that is required at the outset, namely, to decide how important the assessment of this individual intervention is for the program, and thus how rigorous the measurements should be (for learning and reporting). Such decisions may affect the level of rigor, but never the selection of the attribution method.

For example, if external factors do influence the change, it's better to do a smaller Quasi Experimental Design rather than to do a bigger Before and After Comparison. A Quasi Experimental Design comparing 30 farmers that potentially benefited with 30 other farmers that did not benefit, is more informative than a larger Before and After Comparison with 60 farmers that all potentially benefitted. The resources that are needed for both assessments are roughly the same.

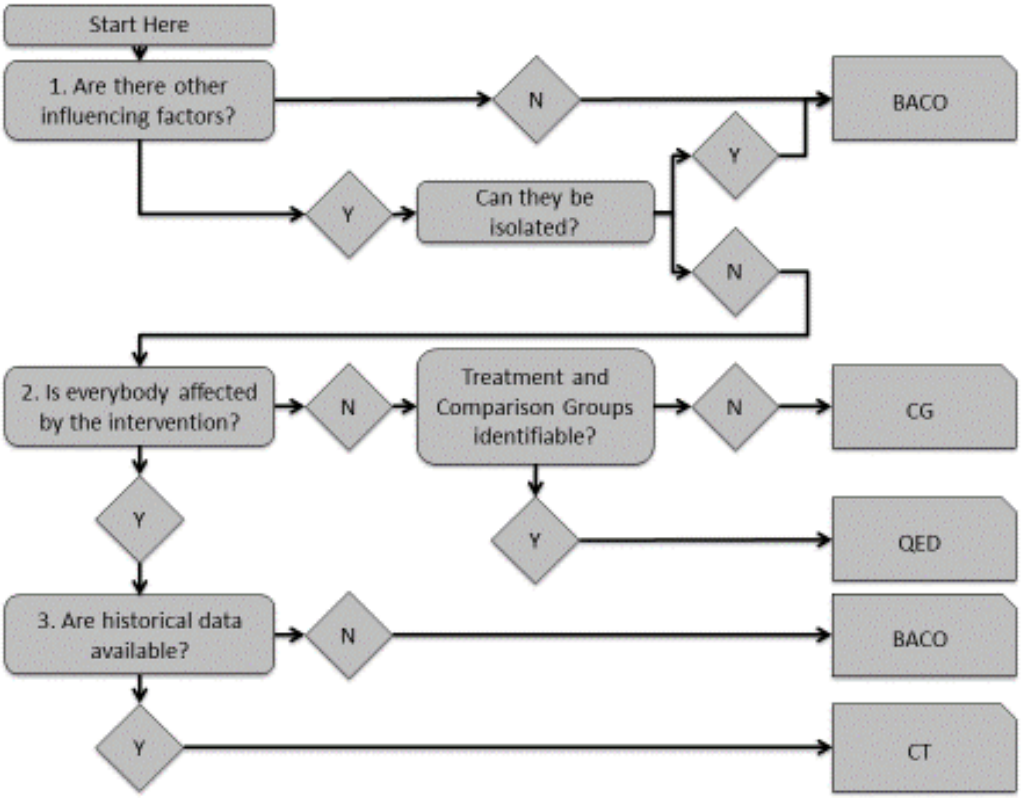

Figure 8 Attribution Selection Aid

1 The first question to answer is: '**Are there other influencing factors** besides the intervention?'.

> *If the answer is no*, there aren't other influencing factors, or they have a very minor influence, the most appropriate attribution method is the Before and After Comparison with Opinion (BACO). Read about the practicalities when applying a BACO in the case study 'The intervention of MDF with Acelda in Timor Leste'.
>
> *If the answer is yes*, there are other influencing factors, then the next question is '**Can these factors be isolated**?' If they can be isolated by putting these factors on hold or freezing them, the most appropriate attribution method is still the Before and After Comparison with Opinion (BACO), holding the external factors constant in the assessment calculations. If they can't be isolated, we have to answer the second main question.

2 The second question to answer is: '**Is everybody affected by the intervention?**'

> *If the answer is no*, not everybody is affected by the intervention, ideally we can search for the most appropriate comparison group to estimate the counterfactual. The next question to be answered is then: ' **Are the treatment and comparison groups identifiable at the start?'** If we are able to construct a treatment group that will have access to and use the new service or input and we are able to construct a comparison group that will not have access at the start of the intervention, then we can opt for the Quasi Experimental Design (QED) method. If this is not feasible, then we can opt to apply the Comparison Group method, allocating respondents in either the user-group or non-user- group after the intervention and at the time of the end-

line survey. In both cases, the impact is estimated using the Difference-in-Difference logic. The pros and cons of each method are described under 2.3 and 2.4 respectively.

More information is available on the practical experiences of and how Propcom Mai-Karfi applied the CG to measure the impact of an intervention in the tractor market.

If the answer is yes, we have to answer the third question.

3 The third question to answer is: '**Are historical data available?**'

If they answer is yes, there are reliable and sufficiently detailed data available, we could Compare Trends (CT) to assess whether there is credible causality between the intervention and the visible change in trends.

If the answer is no, we may have to resort to a Before and After Comparison with Opinion not because it is an appropriate method for this intervention, but because there are no other appropriate alternatives. It does however imply that we have to understand, and communicate, the limitation of the applied attribution method. The quality and credibility of the reported impact will be dependent on the skills and degree of conservatism applied when estimating the influence of other factors and thus the correction in terms of impact.

4. Considerations in field research

Triangulation

For all of the above attribution methods, we should realize that triangulation will reinforce the process of estimating impacts. For each measurement, we should not be depending on one source and one measurement tool. The more we triangulate, the more likely it is that the estimated impact is a reasonable reflection of the reality. Triangulation enables the program to accept lower levels of confidence, thus reducing sample sizes, and in turn reducing measurement costs. More information on sampling is provided in the sample size calculator.

Combining the impact assessment of interventions, or not

The selection process described above for the most appropriate attribution method is applied for each intervention because different interventions require different attribution methods.

Most of the time, we will opt to measure the result of individual interventions. Interventions often start and thus impact the target beneficiaries at different times. The end-line will then have to deal with potential beneficiaries that have benefited from one or more interventions and that have also benefited for different periods of time, complicating the assessment. Many times, we see that some interventions target the same overall target group, but in reality they each impact only sub-sets of the overall target group. We then run the risk of

ending up with too small sample sizes in each sub-group to be able to draw statistically sound conclusions.

The main advantage of assessing changes due to individual interventions is that the information is available and specific per intervention. If we wished to assess the impact of several interventions at once, it becomes difficult to judge which intervention was more effective.

In a few cases, when several interventions took place at the same time and targeted the same beneficiaries, we may not have had the option to isolate the impacts, hence we have to combine the assessments. Typical examples are when the same members of a farmer association gain access to new seeds, make use of new services and apply better cultivation techniques, all due to the interventions with the association to provide these inputs and services.

The Alliances Lesser Caucasus Programme (ALCP) in Georgia was operational in a relatively small geographical area with relatively few target beneficiaries that were affected by many interventions, hence they opted to measure the impact by combining the impact of those interventions.

5. Benefits of measuring attribution

Selecting the appropriate attribution method is crucial. Measuring impacts with methods that have not appropriately assessed the attribution of changes to an intervention leads to management making erroneous decisions because their decisions are based on incorrect information. For example, management might decide to upscale an intervention given the positive changes measured, but not realizing that these are mainly caused by external factors.

Choosing the most appropriate method however may also lead to cost savings, since thinking through the intervention logic and its context, may lead to more efficient (and more informative) methods. Last but not least, only impact that can credibly be attributed to our interventions should be reported to donors and used to evaluate what has been achieved by the program.

6. Practical examples of assessing attribution

- The intervention of MDF with Acelda in Timor Leste, illustrating the use of a before and after with opinion (BACO) method.
- Samarth-NDMP intervention in the ginger sector in Nepal, illustrating the use of a quasi-experimental (QED) method.
- The Alliances Lesser Caucasus Programme (ALCP) in Georgia, illustrating how a single impact assessment could assess attribution for multiple interventions.
- Propcom Mai-Karfi (PM) intervention in the tractor market in Nigeria, illustrating the use of comparison groups (CG)