Applying Experimental and Quasi-Experimental Methods IN COCOA SUSTAINABILITY PROGRAM EVALUATIONS

Meghann Jones
Senior Vice President, Ipsos

Dr Robert Petrin
Director of Data Analytics and Applied Sciences, Ipsos

Manuel Kiewisch
Global Cocoa Life Monitoring & Evaluation Lead, Mondelēz International

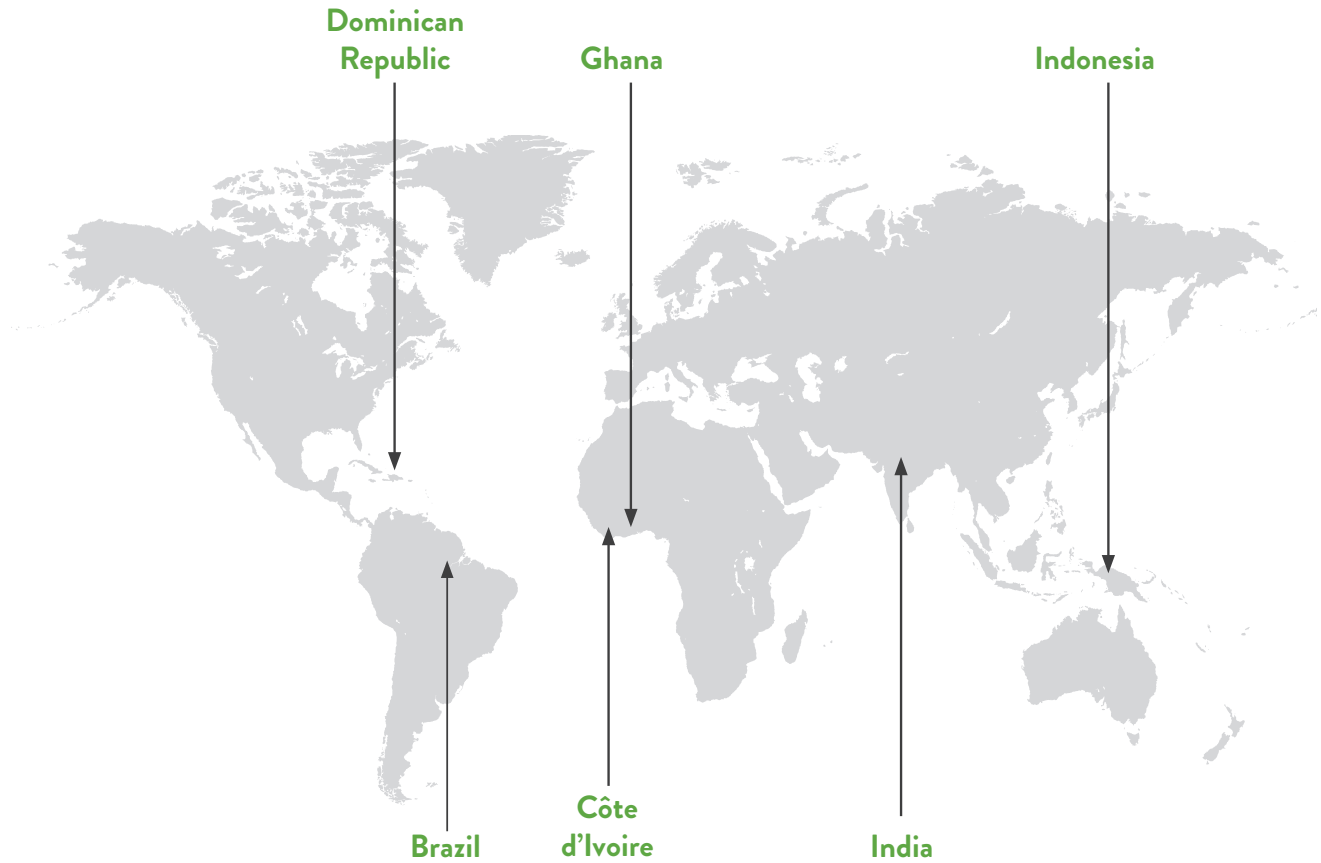
Ferdy Fabian
Measurement and Evaluation Manager, Cocoa Life, Mondelēz International

A joint report by Ipsos and Mondelēz International Cocoa Life

Introduction

Since 2016, Ipsos has worked with Mondelēz International's Cocoa Life program to design and implement measurement and evaluation methodologies that give insight into the program's impact. The Cocoa Life program works with thousands of smallholder farmers, their communities, and their families in Côte d'Ivoire, Ghana, Indonesia, Dominican Republic, India and Brazil, by providing access to training for on-farm and off-farm skills, building platforms within communities to steer their own development, assisting with funding for community projects, and partnering to protect natural resources.

This paper shares learnings from implementing measurement and evaluation studies to understand the impact of Cocoa Life interventions in cocoa origin countries. The information included in this paper is intended to be useful to others working in the field, seeking to develop their own approach to understanding how interventions affect the communities that participate in them.



Understanding Impact

The problem of attribution – whether or not an observed change is the result of an intervention - is an enduring challenge in program evaluation. To understand the impact of a program, you need to know what would have happened in the absence of the program (i.e. the counterfactual), to compare the observed outcomes of the intervention with outcomes in a group that did not receive the intervention.

The gold standard approach for doing this in the hard sciences, such as medical research, is using a randomized control trial (RCT). However, there are many barriers to implementing effective RCTs in the social science context related to implementation, scalability, and ethics.

The theoretical design of an RCT can be different from the implementation on the ground, where challenging conditions in the local context may make it difficult to follow the evaluation protocol. For example, the design of the intervention itself may make it difficult to achieve true randomization between the treatment and control group. Attempting to implement an intervention in one community or set of households, but not in neighboring communities or households, may be difficult or impossible because of logistical or governance problems.

Furthermore, keeping some communities or individuals in a control group may present ethical issues if the implementers of the program do not want to prevent the control group from experiencing the benefits. Not including communities or individuals in a program simply for monitoring and evaluation purposes is considered questionable practice among some experts and practitioners.

One way to overcome some of the challenges of RCTs while maintaining the same benefits of randomization in principle is to use quasi-experimental designs. These designs create or identify, post-hoc, an artificial control group, known in this context as a comparator group. However, in practice, they present their own challenges. For example, given the paucity of rigorous demographic and community profile data available for many of the contexts in which cocoa livelihoods programs operate, it can be challenging to identify viable comparator sites that match the treatment site. This means that comparing the difference between a treatment group and an artificially constructed comparator group could produce significantly different results to a comparison with a treatment group and a truly randomized control group.

In addition, since several treatment interventions are usually offered to farmers at the same time where cocoa sustainability is seen as a critical challenge by the chocolate industry, it is less and less possible to find a group of farmers that will not have received any intervention at all. As it is very difficult to isolate the effects of one intervention over another one running concurrently, it may be difficult to get a true picture of the effects of the program of interest.

Finally, regardless of whether an experimental or quasi-experimental design can be implemented to give a clear picture of the intervention's impact, evaluators should remain mindful that these methods often yield limited insight into how or why an intervention does or does not work as intended. In this light, it's important not to rely on any single method for understanding the effects of an intervention, to ensure that all interventions have a strong theoretical basis and that data is triangulated from multiple sources to understand impact. RCTs can have a place in evaluation design if the limiting factors described below can be controlled and if any level of additional certainty provided by an RCT is absolutely critical for the evaluation.

The case studies that follow show the process Cocoa Life went through to overcome these design challenges in evaluating the impact of its programs in Indonesia and Ghana.

What is a Randomized Control Trial (RCT)?

An RCT is a study in which the subjects are randomly distributed into groups that are either subjected to the intervention or not. Those that are part of the intervention are called the treatment group, while those that do not are considered the control group.

Following an intervention, the outcomes of the two groups are compared and, because randomization should mean the groups are the same in all dimensions other than receiving the treatment, any difference in outcomes can be attributed to the effects of the intervention.

What is Propensity Score Matching (PSM)?

When study participants are not randomized to treatment or control/comparator group (or, when the initial randomization gets disrupted), the PSM method manually balance the treatment and comparator samples on observable characteristics by matching similar individuals. If the matched sample is sufficiently balanced, we can then use the comparator group to determine whether statistically significant differences in outcome variables are truly attributable to the intervention.

The Cocoa Life Experience:
Two Case Studies

Ipsos has worked with Mondelēz International on impact studies for Cocoa Life since 2016. Among several studies that have taken place in four origin countries, two experimental/quasi-experimental studies have been conducted.

The first, a quasi-experimental design, was implemented in Indonesia, where available demographic data was used to identify communities that matched Cocoa Life communities entering the program in 2015 on observable characteristics, meaning they could be used as a valid comparator group. The second was in Ghana, where a RCT was attempted with a group of communities and cocoa farmers who entered the Cocoa Life program in 2016.

Case Study 1: Using a Quasi-Experimental Approach in Indonesia

For the early Cocoa Life program entrants in Indonesia (Cohort 1, entering the program in 2015), a quasi-experimental impact study design was used. This involved conducting the study in comparator communities that were matched with Cocoa Life (treatment) communities based on demographic data.

Following baseline data collection, when we compared the treatment communities with the comparator communities on numerous demographic and key variable measures (like total cocoa yield), we found that matching was not sufficient and that comparisons at the community level would not be possible.

However, because the comparator group was still "clean" – that is, they hadn't received the Cocoa Life intervention – we attempted to fix the experiment by using propensity score matching (PSM). Rather than matching and comparing whole communities, PSM matches individuals on key demographic and variable measures.

Our PSM approach was able to successfully balance the groups on individual attributes for only one of the two regions in the study – Lampung. Once our matching was complete for this region, we then used DID analysis to determine whether the observed outcomes could be attributed to the program.

What is difference in differences (DID or DD) analysis?

DID analysis is a statistical technique that mimics elements of experimental methods when an actual experiment is not possible. As with an RCT, it compares the outcomes of a treatment and control group to give an estimate of the impact of an intervention. In the absence of other factors that may have contributed to change in the outcomes for each group, the difference between the final outcomes minus the difference in original state gives the estimates treatment effect of the intervention.

Key learnings:

1. Locating sufficiently accurate and detailed population data to enable the identification of comparator communities is challenging, and without such data, matching can be imperfect and inefficient.
2. Even when using PSM, having accurate and detailed population data to use as matching criteria upfront can help to ensure efficiency. Without this data, a lot of cases are likely to be discarded during the analysis.
3. For individual-level PSM to work, a comparator sample about 25% larger than the treatment sample is needed to ensure that good matches will be found. This creates a sizable – and expensive - requirement for data collection.
4. With both community matching and PSM, the issue of control still exists in contexts where multiple interventions are happening: multiple interventions create a lot of "noise" that is difficult to untangle from the intervention itself.
5. Using PSM is very resource intensive. In addition to collecting 25% more data than with a community matching approach, to conduct the PSM procedures is technically complicated and laborious.

Using a Quasi-Experimental Approach in Indonesia — View from the Ground

"It was only possible to perfectly retain the treatment vs. comparator group identity of farmers during the baseline study. Due to scale up, at least 5 of the 25 comparator group villages became treatment villages - it was not reasonable to deny them. Also, the calculation that comparator villages were in a similar condition to the treatment villages changed over time, as we found certain districts to have been subjected to a big cocoa government programs which led to distortion. Better knowledge of future plans by other actors, and more detailed demographic information, could improve similar studies in the future. The quasi-experiment was an important learning for Cocoa Life, which needs to understand the best way to measure outcomes and impacts."

Fabian Ferdy, Cocoa Life, Indonesia M&E Manager

Case Study 2: Conducting an RCT in Ghana

In 2016, 8,438 farmers entered the Cocoa Life program in Ghana. For the RCT, 565 were randomly selected into a treatment group, and 483 were randomly selected into a control group.

During the Wave 1 data analysis it was discovered that the necessary rigid adherence to the RCT protocol had failed at the implementation stage. Some communities that were initially selected as comparator communities were included in the Cocoa Life program after the study design had been finalized. Thus, they had received interventions and had to be removed from the control group, otherwise this would lead to errors when we estimated the program impact by comparing differences in outcomes between the treatment and control group. Compounding this problem, within the treatment group of Cocoa Life communities, some individual farmers who had initially signed-up for the program decided to drop out. Thus, they couldn't be included in the treatment group.

In fact, as a result of changing agreements between farmer organizations and farmers, it was impossible to ascertain whether there was any farmer in the final registration lists who had not received the intervention and who could be considered part of the comparator group. Therefore, the RCT design had to be abandoned.

Key learnings:

1. RCTs must only be used if there is complete certainty that the intervention protocol will be followed and can be monitored.
2. Centralized record-keeping that responds in near real-time to changes in program settings is necessary to verify the purity of program implementation and ensure the rigor of the findings.

Conclusions

Appropriate Uses of Experimental and Quasi-Experimental Design

The two Cocoa Life cases confirmed some of the known issues with experimental and quasi-experimental design in this context: achieving a valid control or comparator group is practically challenging and resource- intensive, especially for developmentally-focused programs which depend on agile and localized solutions.

This finding aligns with research from governments, academia, and non-profits that shows that it is difficult to capture causality scientifically, even in the best of circumstances, for such interventions. Social programs are not conducted in laboratory settings and, as such, changes in circumstances that are outside the control of a program can invalidate observed results.

However, if an experimental or quasi-experimental design is being used, we recommend always designing for multiple scenarios to safeguard your approach. For example, even if you are designing for matching at the community level comparison, we recommend also designing for the use of PSM and for non-experimental approaches in case something goes wrong with your plan. While this is likely to drive up the cost of a study significantly, it can be justified for answering specific and critical learning questions, and it ensures that if something doesn't go to plan, you still have useful insights from your study.

Furthermore, make sure to build your impact study on a robust theoretical foundation, regardless of other design parameters. Without concrete theoretical hypotheses about how and why an interventions works, it may not be possible to design an effective evaluation framework and indicator set, or to understand how outcomes are linked to program activities and inputs.

For Cocoa Life, the results from the Indonesia quasi-experimental study have been valuable in providing additional information about program impact. Based on the learnings observed, we will not make quasi-experimental design a standard feature of Cocoa Life's impact evaluation. However, we understand what benefits the technique can deliver under the right circumstances and will use that insight to guide other research initiatives as appropriate.

A joint report by Ipsos, Mondelēz International and Cocoa Life