Development Effectiveness

Impact-Evaluation Guidelines
Technical Notes
No. IDB-TN-198
December 2010

Designing Impact Evaluations for Agricultural Projects

Paul Winters
Lina Salazar
Alessandro Maffioli

Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank

Designing Impact Evaluations for Agricultural Projects

Impact-Evaluation Guidelines

Paul Winters
Lina Salazar
Alessandro Maffioli

Inter-American Development Bank
2010

http://www.iadb.org

Paul Winters: winters@american.edu; Lina Salazar: lsalazar@iadb.org; Alessandro Maffioli: alessandrom@iadb.org

Designing Impact Evaluations for Agricultural Projects

Abstract

Paul Winters* Lina Salazar** Alessandro Maffioli ***

With input from Boris Bravo-Ureta, Steve Boucher and Conner Mullally****

The purpose of this guideline is to provide suggestions on designing impact evaluations for agricultural projects, particularly projects that directly target farmers, and seek to improve agricultural production, productivity and profitability. Specific issues in evaluating agricultural projects are addressed, including the need to use production-based indicators and to carefully consider indirect or spillover effects that are common in agricultural projects. The guideline considers the challenges of conducting impact evaluations of agricultural projects as well as the methods for assessing impact. Issues of collecting agricultural data for an impact evaluation and how to put together the overall design strategy in an evaluation plan are also covered. The guideline concludes with three case studies of impact evaluations designed for a technology adoption project in the Dominican Republic, a forestry/technology project in Nicaragua, and a crop insurance project in Peru.

JEL Classification: H43, O12, O13, O22, Q12, Q18

KeyWords: Impact Evaluation, Agriculture, Technology Adoption, Development Effectiveness, Dominican Republic, PATCA, Nicaragua, APAGRO, Crop Insurance, Peru, PROCAMPO, Mexico.

---

* Associate Professor, Department of Economics, American University, Washington, DC; winters@american.edu

** Research Fellow, Strategy Development Division, Inter-American Development Bank, Washington, DC; lsalazar@idb.org

*** Lead Economist, Strategy Development Division, Inter-American Development Bank, Washington, DC; alessandrom@iadb.org

**** Boris Bravo-Ureta, University of Connecticut; Steve Boucher, University of California at Davis; and Conner Mullally, University of California at Davis.

Table of Contents

1. Introduction

While impact evaluations have become widespread in the last decade and the methods of impact evaluation widely known, they are much more common in the social sectors where the indicators of impact tend to be more easily identifiable and there is a longer history of conducting such evaluations. With respect to agricultural projects, a recent study by the IDB (2010) finds that the coverage of agricultural impact evaluations is limited in most areas and even where there is a greater critical mass of evaluations, such as in land titling and technology adoption projects, additional information on these types of interventions would be helpful.

This conclusion is mirrored in a recent review of agricultural impact evaluations by Del Carpio and Maredia (2009), which identifies a total of 85 papers that can be defined as true impact evaluations; that is, which attempt to establish a clearly defined counterfactual. Of those, only eight use experimentally designed approaches while the remainder use quasi-experimental or non-experimental approaches, suggesting that few of the evaluations were designed prior to project implementation. Further, although the papers use evaluation techniques, in many of the cases the papers assess broader changes in rural areas (e.g. expansion of rural banks, contract farming, and new varieties such as Bt cotton) rather than explicit government agricultural policies or projects. Overall, Del Carpio and Maredia (2009) conclude that the evidence on the effectiveness of agricultural projects in developing countries is quite thin.

Although this is the case, the interest in conducting impact evaluations of agricultural projects has increased, with pressure coming from two quarters. First, economists have increasingly emphasized the use of randomized trials to determine the effectiveness of development assistance (Banerjee, 2007; Duflo and Kremer, 2005). Although there is some debate over whether randomized trials are the only valid approach (Ravallion, 2009), economists seem to agree on the value of carefully collecting data to evaluate the impact of development projects and in the importance of using carefully constructed datasets and empirical approaches to identify impact.

Second, within the development community the general increased emphasis on evaluation has begun to move from the social to the productive sectors. The IDB's recent study of the development effectiveness of agricultural projects reflects this new emphasis. The study, in fact, shows that the number of impact evaluations of agricultural projects has increased in recent years and that the IDB has expanded the number of agricultural impact evaluations and included more impact evaluations as part of project design. Further, the meta-analysis conducted by Del Carpio and Maredia (2009) noted above clearly reflects the World Bank Impact Evaluation Group's increased emphasis on evaluations in productive sectors. Finally, the United States' Millennium Challenge Corporation (MCC) has also contracted a number of studies assessing the impact of agricultural projects using randomized trials.[1]

While methods of evaluation are well known, specific issues in evaluating agricultural projects make them different than evaluations done in the social sector. First, agricultural projects are generally designed to improve production or the returns to agriculture and therefore impact evaluations of agricultural projects focus on production-based indicators such as gross margins, crop prices, yields, productivity, agricultural investment, spending on agricultural inputs, technology adoption, changes in land use patterns, crop and varietal diversification and food for home production. Collecting this type of information can be challenging, beginning with the definition of the sample unit: in fact, while production is often linked to multiple plots and crops, the decision-making process takes place at the household level. The challenge is even greater when attempting to evaluate the impact of a project on different types of households, such as smallholders and largeholders, who quite often have very distinct production systems.

Second, impact evaluations often focus on examining a series of indicators in order to obtain a picture of the overall average effects of the intervention as well as the mechanism by which these effects were obtained. In analyzing agricultural production, the relationship between inputs and outputs or profitability is often examined through production or profit functions. Presumably, agricultural projects have an impact not just on production inputs and outputs, but on how they are used and combined. This needs to be considered in an evaluation.

Third, agricultural evaluations are often complicated by indirect or "spillover" effects that are due to the transferring of new technologies and management practices from project participants to non-participants. In fact, agricultural interventions, particularly for technology adoption projects, often explicitly aim at facilitating spillover effects. While these factors often increase the impact of the operations, they complicate the evaluation design by making hard to find an "uncontaminated" counterfactual. Additionally, since much of the influence of the project may be through these spillover effects, a correct assessment of the project requires considering

[1] On the MCC's website, see Agricultural Impact Evaluations at http://www.mcc.gov/pages/sectors/impact/agriculture

how to identify these spillover effects when they exist and to ensure a reasonable, uncontaminated counterfactual (Angelucci and De Giorgi, 2009; Angelucci and Di Maro, 2010).

While the data and evaluation design issues require special consideration, evaluating agricultural projects can be done using standard approaches with some modifications. The purpose of this guideline is to provide suggestions on designing impact evaluations for agricultural projects. In particular, the guideline focuses on agricultural projects that directly target farmers, usually poor farmers, and generally seek to improve agricultural production, productivity and profitability.

For this purpose, the guideline includes a general discussion of the data and methods that can be used to evaluate agricultural projects as well as three case studies of evaluations that have been designed to evaluate agricultural projects in Latin America. As part of the general discussion, Section 2 discusses the logic of agricultural projects and common indicators to assess impact. Section 3 discusses the specific challenges of conducting impact evaluations of agricultural projects. The possible methods for assessing impact are presented in Section 4 while Section 5 addresses issues of data collection. Section 6 provides information on how to put together the overall design strategy in an evaluation plan. Following this, the next three sections (7-9) provide case studies of impact evaluations designed for a technology adoption project in the Dominican Republic, a forestry/technology project in Nicaragua, and a crop insurance project in Peru.

2. Agricultural Projects and their Expected Impact

While governments and donors support a number of projects that affect the agricultural sector, the focus in this guideline is on projects or components of projects that directly target farmers, with a primary objective of improving agricultural production, productivity and profitability in order to increase agricultural income. This may include projects with other objectives, such as improving management of the natural resource base, but a key focus is on improving farmer welfare through agricultural production. Projects that meet this criterion include those that focus on agricultural extension and technology transfer, land titling and access, investment in or management of irrigation systems, farmer credit and insurance access, natural resource management, and market access among others.

To successfully identifying the important set of final and intermediate indicators requires considering the logic of the intervention model; that is, how is the project going to improve the well being of farmers? This should be reflected in lists of indicators used such as those found in a project's Results Matrix often used by development organization such as the Inter-American Development Bank. The indicators should have a vertical logic that show how project investments alter farmer behavior and, in due course, lead to an impact on farmer well being. Ultimately, these sets of indicators create a series of hypothesis regarding how the project will be successful and if an evaluation is carefully designed, these hypotheses can be tested. Box 1 provides an example of the indicators used to assess the logic of a project in Ecuador, the *Plataformas de concertación*, that links poor potato farmers to high-value potato market.

Box 1: Assessing the Logic of the Ecuadorian *Plataformas*

Initiated in the Ecuadorian Sierra in 2003, the *Plataformas* were alliances between small scale farmers and a range of agricultural support service providers (government, non-governmental organizations and universities) that sought to increase profits of potato-producing smallholders by linking them to high-value potato markets such as restaurants, supermarkets and processors. By linking smallholders directly to high-value purchasers, the project circumvented the wholesale market thereby allowing benefits of the new agricultural economy to accrue to producers. This approach seeks to lower transaction costs and to improve overall cost effectiveness of smallholder production through creating a support system to facilitate smallholder entry into these high-value markets.

To enable success, the *Plataformas* provided smallholders with training through Farmer Field Schools (FFS), which focused on helping producers meet the demands of high-value markets and generally supporting potato production. To meet market quality standards, there was a concern that participating farmers would increase the use of pesticides to avoid the risk of their product not being accepted by buyers. To address this concern, the FFS also emphasized Integrated Pest Management practices designed to minimize the use of pesticides while managing pest populations at acceptable levels. Along with training, new potato varieties that met the needs of the high-value market were provided. Finally, a strong emphasis was placed on organizing smallholders to meet the grades and standards of the high-value potato market in a timely manner.

In their impact evaluation, Cavatassi et al (2010) found that the *Plataformas* increased gross margins of beneficiary farmers suggesting that it is possible to link smallholders to high-value potato markets. The evaluation sought, however, not to just determine if smallholders increased gross margins, but the mechanism by which this occurred. One key aspect was that it increased prices received by potato farmers by 30 percent. However, the *Plataformas* was also found to increase yields and the value of potatoes sold suggesting more potatoes were produced and more marketed. The increase in yield was partially due to an increase in the use of inputs suggesting that some of the gain came from improved productivity. The impact of the project was therefore found to be partially due to the fact farmers became linked to high-value and partially as a result of improved production practices.

The basic logic of the aforementioned agricultural projects is to intervene in such a way that ultimately leads to higher and more secure income from agriculture. This can occur through a number of mechanisms including through increased investment, greater input use, adoption of new technology, lower input prices or higher output prices, reduced risk, greater long-term sustainability, etc. For example, a land-titling project improves tenure security and may induce greater investment. It may also lead to more land sales which shifts production to more efficient farmers. An agricultural extension project may lead to technology adoption that alters input use, increases productivity and improves the long-term management of the land. A crop insurance project provides an instrument to deal with risk allowing farmers to invest in more profitable yet potentially riskier activities and to avoid the decumulation of assets when bad outcomes occur. A project that links smallholders to high value markets may increase input costs, but might lead to much higher prices for output and thus greater profits.

All of these projects ultimately lead to higher and more secure farm income in the short or long run, but the mechanisms by which they do so differs. Given this is the case, along with measuring higher impact indicators such as farm income, gross margins or crop profitability, in evaluating farmer-targeted agricultural projects it is important to also measure these intermediate outcomes such as input use, productivity, prices, etc. Examining these outcomes allows for an assessment of the mechanism by which a project was able to achieve success, or why it failed to do so. Further, in many cases, projects seek to have secondary benefits in terms of improved health or environmental outcomes, better resource management and greater knowledge. These secondary effects are often as important for the project as the effects on farmers and should be included as part of the data collection. Table 1 provides a summary of some of the key indicators that might be used for evaluating the impact of agricultural projects.

Table 1. Indicators for Assessing Project Impact on Targeted Farmers

Impact Indicators	**Measure**
Agricultural income	$
Food security	Food security index
Agricultural Profits	$ per hectare
Gross Margins	$ per hectare
Productivity	Total factor productivity
Yields	Output per hectare
Mechanisms of impact	
Price of output	$ per unit
Value of harvest	$
Value or percent of harvest lost	$ or percent
Value or percent of harvest for home consumption	$ or percent
Value or percent of harvest sold (on farm, local market, exports)	$ or percent
Transaction costs	$ or units of time
Input costs	$ per hectare
Costs of key inputs (seed, fertilizer, etc.)	$ per hectare
Family labor used	days per hectare or value of days per hectare
Cost of paid labor ($/ha)	$ per hectare
Farm-owned animal traction	days per hectare or value of days per hectare
Costs of paid animal traction	$ per hectare
Farm-owned machinery (tractor, sprayer, etc.)	days per hectare or value of days per hectare
Costs of rented machinery (tractor, sprayer, etc.)	$ per hectare
Costs of rental land	$ per hectare
Adoption of key technology (seed, practice, etc.)	Yes=1, No=0
Costs of technical assistance	$
Secondary effects (health, environment, knowledge, etc)	
Chemical use (pesticides, fertilizers, etc.)	Yes=1, No=0
Chemical applied (pesticides, fertilizers, etc.)	$ per hectare
Relative toxicity	Environmental impact quotient (EIQ) per hectare
Protection for chemical application (gloves, poncho, etc.)	Yes=1, No=0
Knowledge of toxicity	Yes=1, No=0 or index
Index of crop, variety, etc. diversity	Shannon, Berger-Parker etc. index
Natural resource practices (terrace, biological barrier, etc.)	Yes=1, No=0
Land "appropriately" managed	percent of land "appropriately" cultivated
Land Cultivated (Agricultural frontier)	percentage or hectares
Knowledge of resource management practices	Yes=1, No=0

While the full logic of an agricultural project should be considered, certain indicators can be more readily attributed to a given project and an impact evaluation focuses on these results. Projects may also contribute to achieve some wider scope results, such as a reduction in poverty rates, which may be very difficult to attribute to the project. Additionally, different indicators require being measured and estimated at distinct time intervals. For instance, adoption of new practices is often a short-run measure while a change in productivity is a medium- to long-run measure. In considering indicators, the timing of measurement and the possibility of being able to attribute the effects to the project should be considered.

3. Challenges in Evaluating Agricultural Projects

The standard challenge of any impact evaluation is determining what would have happened in the absence of the project. To truly understand the impact of a project on a given indicator, information would ideally be available on project beneficiaries with the project and those same beneficiaries without the project. The indicator could then be compared between these two states to see if the project had an impact. Of course, beneficiary farmers cannot be simultaneously in the project and out of the project making it necessary to find a substitute group of farmers to act as the counterfactual—that is, what would happen in the absence of the project. To be a legitimate counterfactual, this counterfactual, or control, group would need to be *exactly* like the project beneficiaries, or treatment group, except they would have not received the project. Thus, any differences in the indicator could be attributed to the project.

Creating a counterfactual through identifying a reasonable control group and ensuring that an identified impact can be attributed to a project is always a challenge. The next section discusses evaluation methods that allow for attribution. Before doing this, this section discusses a couple key challenges specifically related to the evaluation of farmer-targeted agricultural projects.

One common issue with evaluating agricultural projects is that they often involve self selection of participants. For example, agricultural extension projects usually interact with self-formed groups of farmers. Land titling projects often offer the option to title land and farmers choose whether to take that option. Projects that offer credit or insurance to farmers may encourage farmers to take-up the offer but only some of them will do so. Self selection implies that only certain types of farmers may choose to participate in a given project. If an evaluation attempts to determine the impact of a project by comparing those that chose to be in the project to those that did not, difference in the indicator of interest may reflect not only the impact of the project, but also any innate differences between participants and non-participants. Suppose the better farmers in a region decide to participate in an agricultural extension project—that is, farmers that are innovative and like to experiment with their production to see what works best. Such farmers are likely to have higher yields even without the project. A comparison of yields between these innovative, treated farmers and non-participant, control farmers is likely to show higher yields for the treated farmers due to the project but also due to the fact the farmers are innovative. The problem is that it is hard to know how much of the yield difference is due to the project and how much to the differences in farmer type. This makes any estimate of project impact biased since the estimate cannot solely be attributed to the project.

Clearly, selection is also an issue if farmers with certain attributes are chosen by the project to participate. If a project focuses on farmers with limited land access, those with larger landholding are unlikely to be a good comparison. However, these attributes tend to be observable since the project must observe them to identify who will participate. With careful evaluation design, particularly if done in combination with project design prior to project implementation, it is possible to create a reasonable counterfactual and avoid biased estimates of impact. Self selection can also be managed, as discussed in the next section, but tends to be more complicated.

Along with selection issues, another critical issue for agricultural projects relate to indirect or spillover effects. Spillover effects refer to impacts of a project on households or individuals who are not directly targeted by the project. Spillover effects include externalities and interaction effects as well as general equilibrium effects.[2] Externalities and interaction effects are common in agriculture since agricultural practices, both production practices as well as mechanisms for interacting with the market, are often transferred from farmer to farmer. In fact, agricultural extension and technology transfer projects are often designed to have spillover effects. For instance, they may seek to train a limited number of farmers in a certain technology to create a critical mass of knowledge that spreads to other non-participant farmers. General equilibrium effects of projects refer to how a project can have broader effects on the local economy. For example, projects can lead to changes in prices for inputs or for a particular output, which would be considered a general equilibrium effect.

Spillover effects create two challenges for conducting impact evaluations. First, if a project leads to spillover effects on the identified control group, this may lead to incorrect estimates of impact. Consider a technology, such as a new variety of seed that increases yields. Suppose a project has been put in place to randomly distribute the seeds to farmers in a predefined region. Those selected to receive free seeds will be the treatment group and those not selected the control. This would seem like a good evaluation design since with a large enough sample the random assignment should ensure control and treatment are similar in everything, but having received the project benefits. The problem comes if those selected to receive the new

[2] See Angelucci and Di Maro (2010) for a detailed discussion of spillover effects.

variety share it with their neighbors who were not selected. This may occur immediately or over time as the value of new variety becomes known. If seed has been shared with the control group, the control group is receiving the benefits of the new variety and any evaluation will underestimate the impact of the project—the evaluation would be comparing a treatment group who received full benefits of the new variety to a control group who at least received some benefits. The control group is “contaminated”.

A second issue is that these spillover effects can be quite large and often exceed the direct effects of the project, especially in agricultural projects where technology is easily transferred. Given this is the case, failing to recognize these effects can dramatically underestimate the overall impact of a project. In some cases, an evaluation that fails to recognize the spillover effects of a project may lead to the project’s discontinuation, while if the spillover effects were properly considered the project would have been viewed as very successful. The challenge is then to design an evaluation that captures both direct and spillover effects and calculates the overall project effect. The next section discusses how to design an impact evaluation to avoid contamination of the control group via spillover effects and to capture these effects in the overall estimate of impact.

4. Evaluation Design and Methods to Employ

A number of approaches can be taken to evaluate agricultural projects. Providing the details of these approaches is beyond the scope of this guideline and the objective of this section is to provide an overview of the options available.[3] Once an option is selected for the evaluation, further details on that approach should be obtained from other sources.

One factor that should be kept in mind in designing an evaluation strategy is the close link between the data collected for an evaluation and the methodology that will be employed to ensure the impact estimate is unbiased. Since data collection for project impact tends to be representative samples of treated and control households or individuals, statistical methods, particularly coming from the econometrics literature, are used to identify impact. The best method for assessing impact for a given project depends on the data available. As a general rule, the better the data, the less sophisticated the econometric procedures that are needed to analyze the data. Because such models necessarily rely on certain assumptions, there has been an increasing emphasis on collecting better data to avoid having to use more complicated econometric procedures. Part of the emphasis in this guideline is to promote the design of impact evaluation *ex ante* along with the project design to ensure that the best data possible has been collected.

Broadly, the approaches to impact evaluation can be divided between experimental and non-experimental approaches. Experiments refer to projects in which the treated and control groups are randomly assigned to ensure they are alike in all ways except in the treated group having received the project. In realty, perfect experiments are difficult to design and even when well done often face issues in implementation. As such, adjustments are often needed in the analysis to ensure that impact can be correctly identified, including using methods that might be considered non-experimental. With this in mind, the distinction is kept here with the primary difference between the two approaches relating to the manner in which the treatment and control are assigned. Experimental approaches are defined as those that attempt to randomly assign treatment and control groups prior to project implementation, even when this attempt may not provide a perfect experiment. Non-experimental approaches are then defined as those that do not

[3] Detailed information on some of the approaches to impact evaluation can be found on the IDB's Development Effectiveness website under Publications: http://www.IDB.org/en/topics/development-effectiveness/development-effectiveness,1222.html

randomly assign treatment and control, but take an approach that allows a control group to be created that represents a reasonable counterfactual to the treatment group.

4.1 Experimental Approaches

Recall that the key issue in impact evaluation is that it is not possible to know what would have happened to project participants in the absence of the project. This would be the ideal counterfactual. The closest to this ideal would be a group that meets all the criteria to participate in the project and has a similar range of characteristics to the participants. The best way to obtain this group is to randomly assign those that meet the eligibility criteria to a treatment group, who will receive the project, and a control group, who will not receive the project.

For example, suppose a government is concerned that liquidity constraints are limiting the ability of poor farmers to purchase inputs and to be sufficiently productive to maintain food security and improve their well being. The government's solution to this problem is to provide unconditional cash transfers under the assumption that this will relax liquidity constraints and allow farmers to purchase inputs. Ideally the hypothesis that providing cash to farmers induces input purchases would be verified in a pilot project and a carefully designed impact evaluation. To identify impact in the pilot region, poor farmers would need to be identified using some sort of criteria such as cultivating no more than two hectares of land and earning more than 50 percent of their income from agricultural activities. Once the eligible farmers are identified, they can be randomly assigned into a treatment group who will receive the cash transfer and a control group who would not. Assuming the project is large enough, on average the treated and control farmers should be alike in all ways except one received the project and the other did not (see Box 2). Both groups met the eligibility criteria and the ones that received the project were selected not based on characteristics of the farm or because they wanted to participate, but randomly thereby removing any systematic differences between the groups. With a large enough sample of treatment and control farmers, data can be collected on each to determine the impact of the project on input use, food security and farmer well being.

Box 2: Why Randomization Produce the Best Estimate of the Project's Impact

Some formalization can help to better understand why randomization is the best option for evaluating the impact of an agricultural project. The impact of a treatment for an individual *i*, noted δ_i, is commonly defined as the difference between the potential outcome in case of treatment and the potential outcome in absence of treatment:

$$\delta_i = Y_{1i} - Y_{0i}$$

In general, an evaluation seeks to estimate the average impact of the project obtained by averaging the impact across all the individuals in the population. This parameter is known as Average Treatment Effect or ATE:

$$ATE = E(\delta) = E(Y_1 - Y_0)$$

where $E(.)$ represents the average (or *expected value*).

Another quantity of interest is the Average Treatment Effect on the Treated, or ATT, which measures the impact of the project on those individuals who participated (where treatment status $D=1$ if participated and zero otherwise):

$$ATT = E(Y_1 - Y_0 \mid D = 1)$$

The problem is that these parameters are not observable, since they depend on counterfactual outcomes. For instance, using the fact that the average of a difference is the difference of the averages, the ATT can be rewritten as:

$$ATT = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 1)$$

The second term, $E(Y_0 \mid D = 1)$, is the average outcome that the treated individuals would have obtained in absence of treatment, which is not observed. However, we do observe the term $E(Y_0 \mid D = 0)$, that is, the value of Y_0 for the untreated individuals. Thus, we can calculate the difference:

$$\Delta = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 0)$$

What is the difference between Δ and the ATT? Adding and subtracting the term $E(Y_0 \mid D = 1)$:

$$\Delta = E(Y_1 \mid D=1) - E(Y_0 \mid D=1) + E(Y_0 \mid D=1) - E(Y_0 \mid D=0)\Delta$$
$$= ATT + E(Y_0 \mid D=1) - E(Y_0 \mid D=0)$$
$$\Delta = ATT + SB$$

The second term, *SB*, is the selection bias: the difference between the counterfactual for treated individuals and the observed outcome for the untreated individuals. If this term is equal to 0, then the ATT can be estimated by the difference between the mean observed outcomes for treated and untreated:

$$ATE = E(Y \mid D=1) - E(Y \mid D=0)$$

However, in many cases the selection bias term is not equal to 0 (as mentioned before, farmers with a higher propensity to innovate may self-select into a technology transfer project and would have higher agricultural outcomes even without the program). In these cases, the difference in means will be a biased estimator of the ATT. The main goal of an evaluation is to ensure that the selection bias is equal to zero in order to correctly estimate the parameter of interest.

In an experimental design, the assignment to treatment is determined by a purely random mechanism. For example, one could assign a number to each eligible individual and select the treated individuals by lottery.

The main advantage of random assignment is that it guarantees that the treatment status (D) is uncorrelated with any other variables, both observable and unobservable, and, as a result, the potential outcomes will be statistically independent of the treatment status. In technical notation:

$$(Y_1, Y_0) \perp D$$

This means that with random assignment, all the characteristics of the individuals are equally distributed between treated and untreated groups (i.e., they are statistically equivalent). On average, the groups will be identical, except for the fact that one of them received the treatment. This implies that:

$$E(Y_0 \mid D=1) = E(Y_0 \mid D=0)$$

which allows one to replace the left-hand side (unobservable) with the right-hand side, which is observable, to estimate the ATT. Thus, experimental design ensures that the selection bias term is zero, and therefore, the impact of the project can be estimated as a simple difference between the average outcomes between groups. The impact may also be estimated, with identical results, by running a linear regression of the outcome on the treatment status variable and a constant:

$$Y = \alpha + \beta D + \in$$

where β captures the impact of the project.

In practice, running this type of experiment in the field is often complicated by the need to simultaneously meet project requirements and address political and ethical concerns. A common problem is with the idea of having a control group that will not receive the project. The view that farmers are being used for experiments can create negative perceptions. However, in most cases randomization can contribute to a positive perception of the project by building a reputation of transparency and fairness. In fact, projects often have limited resources and are unlikely to be able to serve all their potential beneficiaries. Under these circumstances, the fairest way to allocate resources is through a lottery where every farmer has an equal chance of obtaining the project benefits. This would clearly appear to the farmers a much fairer allocation process compared to alternatives such as social connections or geographic proximity.

Experiments may not be politically possible or it may be that everyone in a given region will eventually receive the benefits. In these cases, another alternative is to use a "randomized phase in". Projects are usually completed over a number of years and only some farmers benefit each year—the project is phased in over a few years with each phase helping a set of farmers. Rather than choosing the order of farmers based on geography, politics or other considerations, the order in which they are treated may be randomly chosen. The first group that receives the benefits is the treatment group and those who will get project benefits in the future will be the control group—at least temporarily. This allows an experiment to be constructed based on how the project is rolled out. An analysis of project impact can be done prior to the control group receiving the project. As noted below, for this approach to work it has to ensure that there is no strategic behavior on the part of those who are receiving treatment at different points in time.

Another common complication is the difficulty in randomly assign individual farms to treatment and control groups. There may be political difficulties in providing support to one farmer and not their neighbor. There may be concerns over spillover effects and contamination (discussed below). Randomization may then be done at a higher geographic unit such as the community or municipality. While this is a common practice, it also is not without pitfalls some of which are noted below.

With a well done experimental design, the statistical analysis of the data is relatively straightforward. For any indicator, a simple test of difference in means can be done. If the difference is found to be statistically significant, this suggests the project had an impact. Using a regression approach (which is equivalent to a simple t-test), the following regression can be estimated:

$$Y_1 = \beta_0 + \beta_1 * P_1 + \varepsilon_1 \quad (1)$$

where, Y_i is the impact indicator of interest for farmer $i=1...N$

P_i is equal to one if the farmer is in the treatment group and zero if the farmer is in the control group

ε_1 is the error term.

With this estimation, the coefficient β_1 provides what is referred to as the average treatment effect. This is the average difference between treatment and control farms for the indicator of interest and is thus the impact of the project. Equation (1) can be estimated for each impact indicator of interest.[4]

[4] The appropriate regression model depends on the type of indicator variable. For example, if it is continuous then an ordinary least squared model is appropriate and if it is a discrete variable, such as a zero-one variable, a probit or logit model would be used.

4.1.1 Challenges in Using Experimental Approaches

The challenge in using experiments to assess impact stems from the fact that perfect experiments are difficult to design and even when they are possible, it is critical to verify that they worked reasonable well. The reasons why perfect experiments are difficult to design include the following:

1. *Small Sample Size*

   A key to a good experiment is to have a sufficiently large sample to analyze a project impact. In Section 5, there is some discussion of how to calculate the necessary sample size. But for a number of reasons, particularly budget limitations, the sample may not be as large as desired. Further, if randomization is done at a higher level than the farm, such as at the community level, there may be a limited number of communities in the project and, in effect, small sample issues. There may be community factors that make the randomization procedure such that the control is not a perfect counterfactual for the treatment. Having a small sample does not bias estimates of impact, but makes them less precise making it more likely that impact estimates will not be significant.

2. *Randomization not Implemented Properly*

   Although designed well, it may be that the experiment was not properly implemented. For example, the implementing agency may have decided to allow some of the control group to receive some or all of the project benefits. Some groups may have been excluded or included for political reasons. If there is a deviation from the protocol established to create an experiment, it is critical to determine what effect this might have on the creation of a counterfactual.

3. *Failure to Follow Treatment Protocol*

   Even though a project is offering some benefits to farmers for a number of reasons they may not choose to take up the offer. For example, if a voucher is offered to pay for extension services, treated farmers may choose not to obtain those services. On the other hand, farmers may only partially take-up or comply with the requirements of the project. They may only choose part of the technologies offered or attend only some of the activities. The concern is that if only those that fully take-up the project are analyzed then issues of selection will arise—is there something fundamentally different about those who take up the project than those that do? Analyzing only those who took up the project leads to potential problems in impact estimates. Similarly, assuming everyone took up the project, if failing to comply is widespread, can also lead to a mismeasurement of impact.

4. *Attrition and Measurement Error*

   Even if there is widespread compliance with the project, there may be issues in the data collection. Members of the treatment or control group may refuse to answer the survey questionnaires or may answer some questions in an incorrect manner. For example, the control group may refuse to answer since they are not receiving project benefits.[5] The treatment group may answer questions strategically to ensure continuation in the project. This can lead to data problems if this is widespread and systematic—meaning that the reasons for refusing to answer or for answering in a questionable manner are related to whether the farmer is in the treatment or control group.[6]

5. *Experimental Effects*

   Related to the previous point, if farmers know they are part of an experiment, whether in the treatment or control, they may act differently or respond to questions differently. In trials to tests new medicines, this is avoided by using a placebo so that study participants do not know if they are receiving the true medicine or not. In project evaluation, a placebo is not possible making experimental effects a potential problem.

6. *Spillover Effects and Contamination*

   Finally, if a project leads to spillover effects on the control group, even if the control group is randomly assigned, the control group is contaminated and does not represent a reasonable counterfactual.[7] The control group can also be contaminated in other ways. For example, if a randomized phase in is being used, this often means that the control group knows that it will be treated in the future. Knowing this is the case, the control farmers may alter behavior in anticipation of project benefits. For example, if the control group

[5] To avoid attrition from the control group and 'strategic' answers from the treated group it might be better to avoid stating that the data collection is for the purpose of an impact evaluation of a specific project. For communication purposes, stating that the data are collected for the purpose of a research project should be enough.

[6] See Stecklov and Weinberg (2010) for a general discussion of measurement error. http://www.iadb.org/document.cfm?id=35158648

[7] See Angelucci and Di Maro (2010) for a detailed discussion of spillover effects. http://www.iadb.org/document.cfm?id=35173297

knows it will be receiving support to adopt a new technology in the future, the control farmers may delay adoption. This means the control group farmers have altered what it is doing as a result of the project creating problems of it being a good counterfactual.

Dealing with these potential issues requires carefully designing the impact evaluation strategy to minimize the chance or the impact of these problems. It also highlights the great importance of having baseline data from prior to the implementation of the project. Impact evaluation is all about how credible the estimates are. Not having a baseline to show the validity of the randomization process could undermine the overall credibility of the study.

Among other benefits, baseline data allows an assessment of whether an experiment created a reasonable counterfactual. As noted, the expectation is that with random assignment the control group should be on average similar to the treatment group in all characteristics except the treatment group received the project. This can be verified using baseline data. Comparisons of key variables using statistical tests of significant differences are one simple way to check this. If the control and treatment group are alike, baseline characteristics should be similar—so the hypotheses that they are equal should not be rejected. Of course, even with random assignment there is a probability that some variables will differ, but across a range of variables these differences should be minimal. If there are a substantial number of differences, it raises questions about how good a counterfactual the control group might be. Box 3 provides an example of baseline tests of difference.

Box 3: Checking if Experiments Worked

Table 2 provides an example of baseline tests of difference from an experimental evaluation conducted in Kenya by Chakravarty (2009). The evaluation examines the impact of a fertilizer provision project targeting farming households in which one or more members was receiving treatment for HIV/AIDS. The study randomly assigned half of the 540 patients to receive free fertilizer for the 2007 planting season. The summary statistics from the baseline survey shown in Table 2-Panel A compare farm households that received the fertilizer (treatment group) to those that did not (control group). The results show no statistically significant differences between treatment and control in maize production practices prior to the project. This suggests that the experiment did a good job in creating a reasonable counterfactual. Note that this is in spite of the fact 15 households refused to respond (number of observations is 525). Panel B shows information on those households that produced maize in 2006 prior to the baseline. This limits the data set to 477 observations. Even with this data set, tests of difference show no statistically significant difference in the baseline between control and treatment maize producers. Again, the experiment appears to have created a reasonable counterfactual.

Table 2. Baseline Summary Statistics from Kenya Survey

Panel A. All study enrollees (2007), 525 observations

Variable Description	Control	Treatment	Difference	p-value
Produced maize 2006?	0.9	0.92	0.02	0.433
Number crops produced 2006	1.22	1.22	0.00	0.988
Maize Acreage 2006	1.26	1.30	0.04	0.710
Maize Production 2006 (90 kg bags)	15.07	15.53	0.45	0.784
Per acre maize yield 2006 (90 kg bags)	10.50	10.69	0.19	0.734

Panel B. Respondents who produced maize in 2006. 477 observations

Variable Description	Control	Treatment	Difference	p-value
Used planting fertilizer 2006	0.88	0.90	0.02	0.476
Per acre planting fertilizer 2006 kg	58.02	54.98	-3.05	0.826
Used top dressing fertilizer 2006	0.10	0.07	-0.03	0.506
Per acre top dressing fertilizer 2006 kg	3.85	5.12	1.27	0.441
Per acre total fertilizer 2006 kg	61.88	60.10	-1.78	0.899
Seed expenditures 2006 Ksh	1657.60	1658.81	1.21	0.993
Qty maize sold 2006 (90 kg bags)	7.76	7.37	-0.39	0.764
Income from maize sales 2006 KSh	7684.78	7289.72	-395.07	0.785

Source: Chakravarty (2009), Table 1

4.1.2 Adjusting to Problems

If an assessment of the experiment shows problems with the created counterfactual, it may be necessary to employ a non-experimental technique. These approaches are described in the next section. One common approach to addressing problems with a counterfactual is to use a difference-in-difference or double difference (DD) approach. As shown below, this approach helps control for (time-invariant) difference between the treatment and control group at baseline. If the assessment of the experiment shows potential problems, the DD approach can help resolve these problems and allows for a reasonable assessment of impact. If there are no baseline differences, using a DD approach gives the same answer as the standard approach shown in equation (1). Given this is the case, it is common for evaluators to use a DD approach even if the counterfactual seems reasonable since it can help identify impact, but rarely causes problems.

Another common problem with experiments comes from the failure of beneficiaries to comply with the treatment protocol. In such cases, the evaluator must decide whether the problem is substantial enough to merit using a non-experimental approach. If the problem is minor, the intent-to-treat might be used for the evaluation—that is, the analysis is done as if all the beneficiaries did comply. The benefit of doing this is it maintains the experiment and self-selection issues do not enter. The downside is that if there is widespread non-compliance, or partial compliance, the benefits of full receipt of the project are likely to be underestimated. If non-compliance is widespread, an alternative approach is to use an instrumental variable approach where the instrument is the intent-to-treat. This is discussed more fully below.

Box 4 summarizes some of the key messages about using experiments for agricultural projects.

Box 4. Key Messages about Experiments and Agricultural Projects

1. Experiments represent the best mechanism to obtain a reasonable counterfactual and are possible when agricultural projects target farmers using clear eligibility criteria.
2. From both a political and practical standpoint, there are a number of challenges to implementing experiments although careful planning can help overcome these challenges.
3. Baseline data is invaluable for assessing the success of an experiment and for adjusting approaches if the data is imperfect. Baseline data should always be collected.
4. Even if randomization does not create the perfect counterfactual, it is usually possible to use a non-experimental method to evaluate the project impact.

4.2 Non-experimental Approaches

If an experiment is not possible, a number of non-experimental methods can be used to assess impact of agricultural projects, including using a double difference approach. The underlying logic of these approaches is the same as in an experiment. To assess impact, it is necessary to identify a counterfactual and then to take measures to ensure the estimate of impact is free from bias. The appropriate method to be used depends on the available data and the details of the project. If secondary data is to be used for an evaluation, the best non-experimental approach depends on the type of data available. If data is to be collected and the evaluation is being designed as part of the project, the best method depends on project details, including who will be eligible for the project and how the project will be rolled out. The best approach requires considering the available options and selecting the best possible option or set of options.

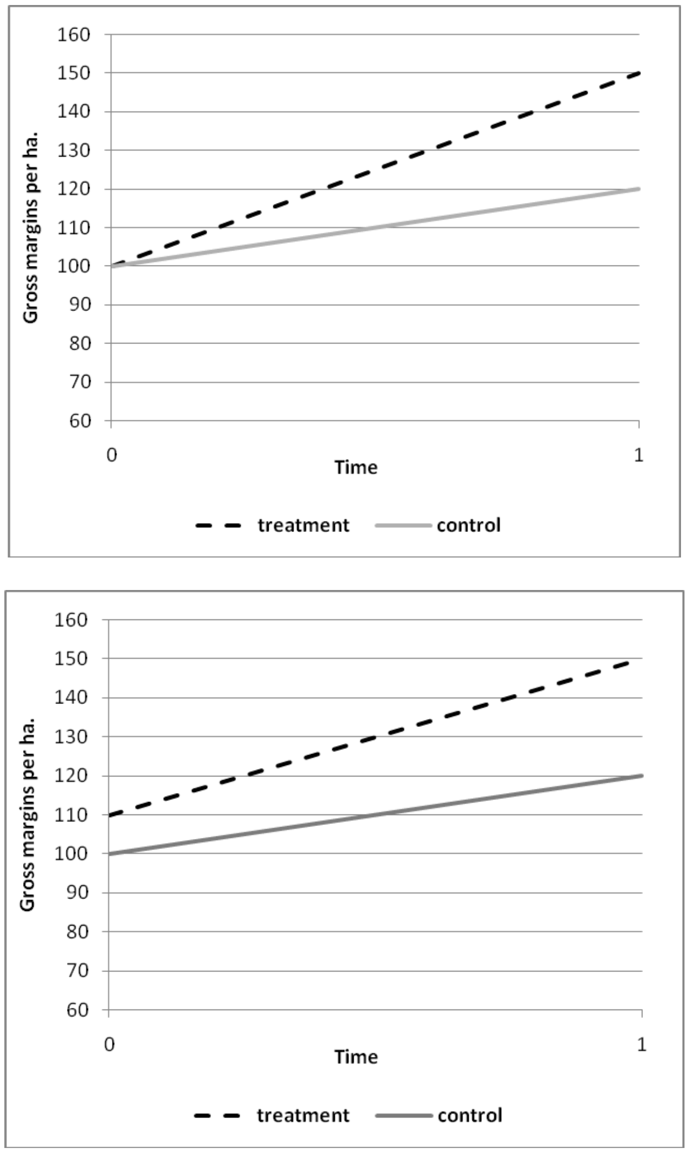
Going through each of the available non-experimental approaches in detail is beyond the scope of this guideline. Instead, each of the options is discussed and the underlying principles behind each approach are noted.

4.2.1 Difference-in-difference

The DD approach is one of the most popular non-experimental techniques in impact evaluation since it allows controlling for some types of selection in a straightforward and intuitive way, as long as baseline data is available. In a DD model, the relevant comparison is changes in the indicator over time. Thus, the comparison in a DD model is between the trends in the control group from before and after the project versus the trends in the treatment group. The double difference then refers to the difference over time (the first difference) and difference between the control and treatment (the second difference). If the trends are significantly greater for the treatment group (in a statistical sense), this suggests that the project had an impact. Thus, the DD estimator combines cross-sectional and over-time variation to correct for differences between groups when treated and controls start from different levels.

Figure 1 illustrates how a DD works. Panel 1 shows the case of an experiment that created the ideal counterfactual, where at baseline (time zero—just prior to treatment) the indicator of interest, agricultural gross margins per hectare, is the same for control and treatment. According to the figure, gross margins in both groups were US$100 per hectare. For the control group, the figure shows gross margins increased by US$20 over the period of the project. This increase reflects a general increase in the eligible population in the indicator. For the treatment group, gross margins increase by US$50, which is the general time trend ($20) plus the project impact. The impact is the difference between these trends ($150-$100)-($120-$100)=$30. Note that since the counterfactual was well constructed, the estimate of impact could have simply been estimated as described Section 4.1 comparing the control and treatment after the intervention using the post-treatment data. This brings about the equivalent results of US$30 as found using the DD.

Figure 1: Double difference



Panel 2 illustrates the case where there are some initial differences between the control and treatment groups at the baseline. In particular, gross margins are US$10 ($110-$100) larger for the treatment group just prior to treatment, suggesting that for some reason prior to the project implementation the treatment group gets a greater return from a hectare of land than the control group. The estimate of impact should reflect the fact that treated farmers started out slightly better off than the control group. By using the DD, a comparison of trends in the treatment versus control shows an estimated of impact of ($150-$110)-($120-$100)=$20. Note that in this case, the estimate of impact using a comparison of treatment and control after the intervention would have produced a biased estimate of impact. For this particular example, the estimate would have been US$30, which would be upwardly biased estimate of the true impact US$20.

The DD approach can also be estimated using a regression approach provided there is baseline and post-treatment data for treatment and control groups. If there is, the following regression can be estimated:

$$Y_i = \beta_0 + \beta_1 * P_1 + \beta_2 * T + \beta_3 * P_1 * T + \varepsilon_i \quad (2)$$

where Y_i is the impact indicator of interest for farmer $i=1...N$

P_i is equal to one if the farmer is in the treatment group and zero if the farmer is in the control group

T is equal to zero if at baseline and one after treatment;

ε_i is the error term.

With this estimation, the coefficient β_1 controls for initial differences between control and treatment, β_2 controls for general trends over time, and β_3 provides the estimate of impact or the average treatment effect. The main advantage of this second specification is that it can be easily extended to the case with several groups and several time periods by adding group and time-period dummy variables.

The fundamental assumption of the DD estimator is that the control-group trend is identical to the trend that the treated group would have had in the absence of treatment. While this assumption is not testable, its validity should always be carefully discussed to ensure that the DD properly estimates the impact of the program. If data for several pre-treatment years is available, one straightforward way to assess the validity of this assumption is to analyze whether pre-treatment trends were equal between groups. While this does not formally prove the identification assumption (which, as mentioned, is not testable), the equality of pre-treatment trends suggests that the treated and control groups are, indeed, comparable and thus reinforces the credibility of the estimations.

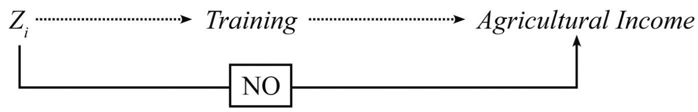
Under this identification assumption, the great benefit of using a DD is that it controls for unobservable differences in baseline characteristics of treatment and control households, thus minimizing potential biases in impact estimates. DD estimates only address time-invariant differences in control and treatment groups. This means that if there are changes that occur over time that affect one group and not the other this cannot be controlled for using this approach. The evaluator must be confident that such changes did not occur to be sure impact estimates are reasonable.

Along with using a DD approach, in estimating equation (2) additional conditioning variables are often included in the regression. These variables are baseline characteristics that are linked to the outcome of interest. If an experiment creates a reasonable counterfactual, the inclusion of these variables should not influence the estimate of impact although it can improve the precision of estimates. If the counterfactual is imperfect, the addition of conditioning can help control for baseline differences in observable characteristics like the DD. As with the DD, the inclusion of conditioning variables can help adjust for problems in assessing impact, but does not bias estimates so it is common for such variables to be included.

4.2.2 Instrumental Variables and Encouragement Design

The instrumental variable approach is a technique that is often implemented when a project includes some degree of self-selection and there is a concern that unobservable differences between those that received the project and those who do not might lead to biased estimates of impact. This is a rather frequent problem in agricultural projects where farmers self-select themselves into the project. For instance, let's suppose that a project aims to increase agricultural income for farmers who self-select themselves into a training course for seed and fungicide management. In this case, it is expected that some intrinsic characteristics that determined farmers' participation in the training might also affect their income, such as their leadership capabilities, their entrepreneurship abilities, their motivations, their drive to succeed, etc. Any comparison between participants and non-participants will include the impact of the project as well as the intrinsic characteristics of participating farmers. More formally, in such cases it is not correct to use a regression to estimate the causal effect of project participation (P_i) on the variable of interest (Y_i) such as described in equation (1). This is because the unobserved characteristics, captured by the error term (ε_i), would be correlated with both income (Y_i) and training (P_i).

To deal with this problem, an instrument (Z_i) or set of instruments is required that can serve to predict project participation (relevant), but is not correlated with the outcome variable (exogenous). In this example, a variable that predicts training participation, but does not affect agricultural income would be the instrument needed.



If an instrument is found, a two stage least squares (2SLS) procedure can be implemented in order to estimate project impact. In the first stage, the effect of Z_i on treatment as follows:

$$P_i = \delta_0 + \delta_1 Z_i + \mu_i \quad (3)$$

where, μ_i is the error term.

In the second stage, the estimated values of treatment ($\hat{P}_i$) from the first stage are used to estimate impact as follows[8]:

$$Y_i = \alpha_0 + \alpha_1 \hat{P}_i + \varepsilon_i \quad (4)$$

Although using instrumental variables solves various problems such as unobserved heterogeneity, finding an instrument after the project has been implemented is not always an easy task. One approach to ensure an instrument is available is to implement the project with a *random encouragement design.* This type of project design implies that some farmers are randomly encouraged to participate through different mechanisms while others are not. For instance, an agricultural insurance program can randomly provide coupons to some farmers to reduce the price of the service and therefore, encourage farmers to purchase the insurance. Also, a project that provides a voucher to finance part of the cost of an agricultural technology can randomly provide coupons that increase the amount of the cost financed by the project if farmers decide to participate in the project (ex. All the vouchers finance 50% of the technology but with the coupon, farmers obtain 5% more financing). In these scenarios, it is reasonable to believe that farmers who received the coupons, or who were *encouraged*, are more likely to participate in the project than those farmers who did not. Given that incentives are randomly distributed among farmers, there is no reason to believe that the encouragement mechanism is correlated to the

[8] Doing this procedure manually will provide incorrect standard errors since they need to be adjusted due to the fact participation is predicted rather than actual participation. It is recommended to use IVREG in STATA to estimate the 2SLS procedure as it calculates the correct standard errors.

outcome variable making it a reasonable (exogenous) instrument. Thus, the incentives can be used as an instrument (Z_i), in the first stage.

As noted, an instrumental variable approach can be used to correct for the bias generated by non-compliance in an experimental design. Non-compliance implies that some farmers who were randomly selected to participate were not treated. For instance, some projects frequently use vouchers that finance certain percentage of the costs of a new technology to promote adoption among farmers. However, even if this voucher is given randomly, farmers are not always able to obtain the benefits as they cannot always afford to finance rest of the costs themselves. It is expected that some farmers are not able to be treated even though they were randomly selected as part of the treatment group. If this is the case, the randomization itself (in these cases referred to as "intention to treat") can serve as an instrument to predict project impact. In other words, the instrument Z_i would be equal to one if the farmer was randomly selected to participate (obtained the voucher) and zero otherwise.

If not embedded in the design of the project, it is quite difficult to identify an instrument that can guarantee both relevance and exogeneity *ex-post*. For example, it is not easy to think about a variable that might affect training participation but does not affect income. A potential candidate could be distance from the farm to the place where the training takes place since it is almost certain that farmers who are located closer to the training facility would be more likely to attend. However, distance from the farm to the training facility may or may not be correlated with agricultural income. In fact, if the training takes place in the center of the community, where all the economic activity takes place and most of the input providers and markets are located, it might be difficult to argue instrument exogeneity. For this reason, identifying an instrument before the project is implemented is critical.

Although the IV approach is a rigorous methodology and tackles biases generated by both time-invariant and time-variant unobservable characteristics of farmers, it also has some limitations. In particular, an IV approach can only estimate a Local Average Treatment Effects (LATE), which means that it's results are relevant only for those whose behavior is affected by the instrument (Angrist, 2001). In the particular case of the experiment with non-compliers, this implies that the results are valid only for the farmers who eventually participate in the program. Therefore, if the original experiment was designed so that the treated and non-treated groups were representative of a certain population, the findings after the correction with IV may not have external validity because the compliers may not be representative of the population of reference anymore.

4.2.3 Propensity-score Matching

The basic idea of propensity-score matching (PSM) is that in absence of an experimental design, assignment to treatment is frequently non-random, and thus, farmers receiving treatment and those excluded from treatment may differ not only in their treatment status but also in other characteristics that affect both participation and the outcome of interest. Among farmers not receiving the project, the PSM approach seeks to find non-treated farmers that are similar to treated farmers, but did not receive project benefits. The approach does this by matching treated farmers to non-treated farmers using propensity scores (described below). In some ways, this technique can be viewed as replicating the project selection process as long as the selection is based on observable factors. By using propensity scores to identify a control group that is like the treatment group, PSM creates what some refer to as a "quasi-experiment" since the control group is *statistically equivalent* to the treatment group.

Specifically, given that the outcome variable cannot be observed for the treated group with and without the project, it becomes necessary to identify a control group in order to infer what would have happened to the beneficiaries without the project. In other words, a group of farmers, units of production or agricultural households are needed who are statistically similar in all observed characteristics to those who received the project. The propensity score allows us to solve this problem by estimating the conditional probability of receiving the project ($P_i=1$) given a vector of observed characteristics (X) (*dimensionality problem*). Hence, all these similarities can be "aggregated" into only one number or score:

$$\Pr(P_i) = \Pr(P_i = 1 \mid X)$$

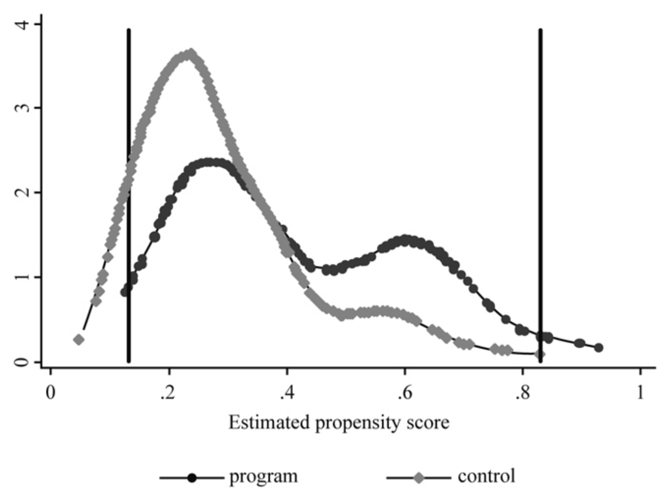
The propensity score or conditional probability of participation is calculated by using a probit or a logit model in which the dependent variable is a dummy variable equal to one if the farmer participated in the project and zero otherwise. The vector of covariates or independent variables should be composed of those characteristics that determined project placement in order to replicate the selection process. To determine these characteristics requires clearly identifying the institutional arrangements that defined selection into the project (Caliendo and Kopening, 2008). Also, it is important to clarify that variables included as independent variables should not be affected by participation in the project—that is, they should be exogenous to project participation. Also, it is important to include variables that allow for some variability and therefore, do not perfectly predict participation or non-participation (Heckman, Ichimura and Todd, 1998).

Two conditions should be fulfilled in order to apply PSM. The first one is referred to as the *Conditional Independence Assumption* (CIA). This assumption implies that after controlling for some set of covariates X the potential outcomes for treated and control (Y_1, Y_0) are independent of treatment status which can be written mathematically as follows:

$$(Y_1, Y_0) \perp P \mid X$$

The implication of this assumption is that participation in the project does not depend on the outcomes. The second assumption refers to the *Overlap Condition*. Once the propensity score is estimated, the next step is to check for an area of common support. Common support refers to the overlap area between the propensity scores of treatment and control groups. In other words, to assure comparability, it is necessary to identify those non-participants with similar propensity scores to the participants (see Figure 2). In Figure 2, the common support is the area located between the vertical lines.

Figure 2: Common Support and Propensity Score Matching



Source: Cavatassi et al (2010)

Assuming that a proper control group is identified so that these two conditions hold and the vector of covariates *X* is observable to the researcher, the propensity score can be used to match statistically similar individuals in order to estimate project impact as follows:

Impact= E [Y_i (t=1, P=1) – Y_j (t=1, P=0)]

where Y_i is the outcome for the treated;

Y_j is the outcome for the non-treated;

$t=1$ represents the period post-treatment;

$P=1$ represents project participation; and

$P=0$ represents non-participation.

The matching algorithm, which refers to the number of non-treated units in the area of common support who will be used for comparison to the treated units, does not necessarily have to be only the one with the closest propensity score (nearest neighbor). In fact, there are many matching algorithms that can be used as robustness checks such as the five nearest neighbors or the kernel density approach.[9] In addition to the closest neighbor, these approaches use information on other neighbors, but often weight these less than the closest score.

Overall, the main advantage of PSM is that it facilitates the identification of a counterfactual when the selection bias to be addressed is clearly due to observable characteristics of the farmers (for instance when it is related to clear administrative rules and there is no self-selection). On the other hand, given that it corrects only for observable characteristics, this approach is a highly data intense and, therefore, its application to agricultural projects could be challenging given the cost of data collection in rural areas. Certainly, PSM alone is not appropriate when unobservable farmers' characteristics might affect both the outcome variables and the program placement. These could be the case when beneficiary farmers self-select themselves in the program because of motivational, attitudinal or skill related characteristics and the outcome variable is also related to their motivation, attitude and skills, such in the case of productivity measures. In these cases, it is always recommendable to at least combine PSM with a DD approach, in order to at least remove the bias due to time-invariant unobservable characteristics (such for instance, motivation). Lopez and Maffioli (2008) and Cerdan-Infantes,

[9] For more information on how to use propensity score matching for impact evaluations see Heinrich, Maffioli and Vazquez (2010) at: http://www.iadb.org/document.cfm?id=35320229

Maffioli and Ubfal (2008) have used PSM and DD to identify the impact of agricultural projects in different settings.

4.2.4 Regression Discontinuity

Regression discontinuity is a non-experimental approach that can be implemented to evaluate projects in which the targeting depends on a specific variable which defines a clear threshold for participation. For example, a project may target only small farmers with less than 5 hectares of cultivated land, farmers below the poverty line or farmers whose agricultural earnings are below the regional average. Based on this procedure, any farmer above this threshold is ineligible and every farmer below the threshold is treated. Of course, the difference between farmers on either side of the threshold may be quite small or may be well understood. This type of targeting suggests a clear threshold that defines project treatment and allows for applying regression discontinuity design. The main idea behind regression discontinuity is that at the margin of the threshold, the assignment to treatment and control is close to be random. In other words, farmers who are in the neighborhood of the cutoff (below or above) are very similar and, therefore, represent a good counterfactual for the treatment group.

There are two types of regression discontinuity: sharp and fuzzy. The sharp discontinuity refers to the type of targeting in which the threshold clearly determines participants and non-participants. For instance, all the farmers who own less than 5 hectares of land receive a voucher to buy fertilizer or all the farmers who earn less than US$2.00 per day receive a subsidy to buy seeds. Assume that x represents the variable that determines treatment (ex. number of hectares of land, income, etc.), with a sharp discontinuity design, there is no value of x at which both treated and control can be found (Imbens and Lemieux, 2008). In other words, treatment is a function of x as follows:

$$P_i \begin{cases} 1 if \ x \leq 5 hectares \\ 0 if \ x > 5 hectares \end{cases}$$

where, $P_i = 1$ if the farmer is treated

$P_i = 0$ if the farmer is not treated

This type of targeting design allows us to use regression discontinuity to estimate the project impact as follows (Angrist and Pischke, 2009):

$$Y_i = \alpha + \beta x_i + \rho P_i + \eta_i \quad (5)$$

Where, Y_i is the outcome variable (ex. Productivity)

x_i is the continuous variable that defines treatment

P_i is a dummy variable equal to 1 if the farmer was treated

η_i is the error term.

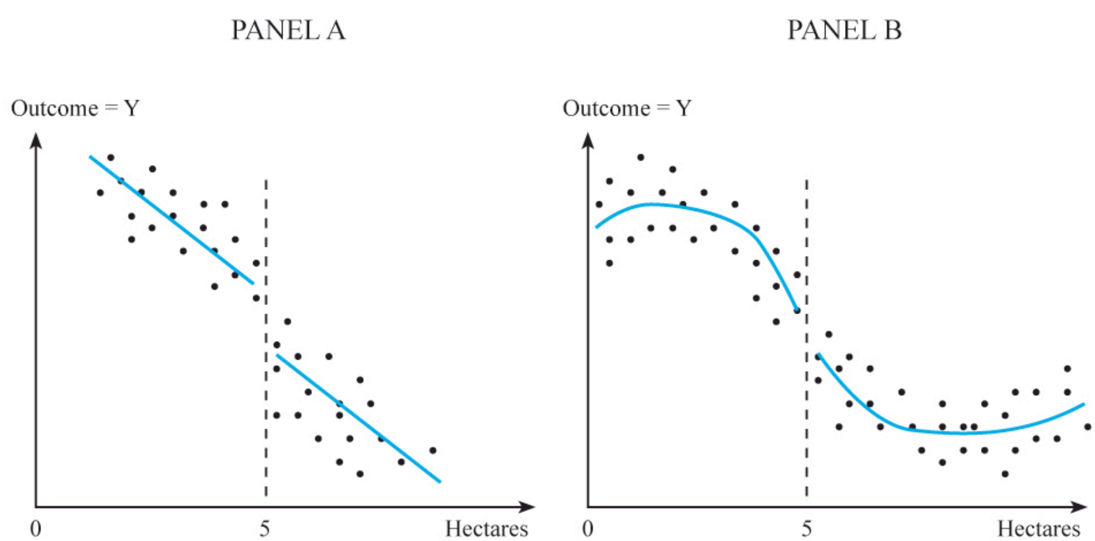
In equation (5) ρ is the coefficient of interest and represents the causal effect of project participation on Y.

In the case in which the relationship between x and Y is not linear, equation (5) can be estimated as follows:

$$Y_i = f(x_i) + \rho P_i + \eta_i \quad (6)$$

Graphically, a plot of the outcome variable Y and the variable x, should show a jump (discontinuity) at the value of x that determines project treatment. Following the example above, at 5 hectares of land (threshold) a discontinuity or a jump on agricultural productivity should be found. This discontinuity is presented in Figure 3, panel A for a linear case and Figure 3, panel B for a non-linear case.

Figure 3: Regression Discontinuity "jumps"



Some of the robustness checks to assure that estimates represent the true impact of the project include estimating the regression for a smaller sample closer to the threshold or "Discontinuity Samples" (Angrist and Lavy, 1999) as well as to estimate the regression for a pre-treatment variable in which case the ρ coefficient should not be significant and the diagram should not show any discontinuity at the threshold (Angrist and Pischke, 2009).

On the other hand, fuzzy discontinuity differs to sharp discontinuity in that the variable x does not perfectly determines treatment and control but influences the probability of treatment. Therefore, the variable x can be used as an instrumental variable to predict treatment and the model can be estimated using two-stage least squares as explained in section 4.2.1.

Regression discontinuity design has not been widely implemented to evaluate agricultural projects despite the fact that many projects define a clear threshold to determine participation such as by land size or income. Other agricultural projects define their interventions geographically which could also facilitate the use of this methodology. For instance, Del Carpio Loayza and Datar (2010) apply regression discontinuity to identify the impact of an irrigation project in Peru by using a geographic cut-off. This methodology is rather powerful as it can be comparable to an experiment in the neighborhood close to the threshold. On the other hand, the main problem of using this approach is that it requires a good number of observations next to the discontinuity in order to draw meaningful conclusions. Besides, as in the case of IV, this methodology can only estimate a local treatment effect, which means that the results are valid only for those participants who are close to the threshold but it might be difficult to extrapolate those findings to other units located far away from it.

4.3 Incorporating Spillover Effects

As noted previously, there are number of reasons to expect that agricultural projects will have spillover effects at least within local communities and possibly beyond that. This not only has the potential to cause problems for identifying the direct project impacts, since it can lead to the potential contamination of a control group, but can lead to underestimates of project impact if not incorporated in the evaluation. If spillover effects are expected, the evaluation design should make efforts to incorporate them into the impact evaluation.

The first step in incorporating spillover effects into an evaluation is to consider the theoretical reasons why spillover effects are expected. As mentioned, the two primary reasons for spillover effects are the existence of externalities and of general equilibrium effects. It is critical to understand why spillover effects occur since it identifies the group of non-participants that may be indirectly affected by the project. These indirect beneficiaries cannot be included as part of the control group since they are "contaminated" by the project. They will, however, become the focus of the analysis of spillover effects. For example, in a technology transfer project, the primary group that is likely to indirectly benefit from a project is non-participants in communities where the project is transferring new technologies. Although not participating in the project, they may learn about the technology or even receive the technology from participants and decide to adopt the technology themselves. If such effects are expected to be substantial, they should be measured.

As with the direct beneficiaries of a project, to capture project impacts on indirect beneficiaries it is necessary to have a counterfactual—that is, a group of farmers with similar characteristics to the indirect beneficiaries but who did not receive indirect benefits. The challenges in indentifying this group are similar to identifying a counterfactual for direct beneficiaries. Following Angelucci and Di Maro (2010), two strategies are considered here: (i) double randomization, (ii) using eligibility criteria. These are considered separately although it is possible to use a combination of these strategies to identify indirect effects. Details of these strategies can be found in their guideline on spillover effects.

Since the objective is to measure both direct and indirect effects of the project, four groups of farmers are required: (i) direct beneficiaries, (ii) a control group for direct beneficiaries (iii) indirect beneficiaries, and (iv) a control group for indirect beneficiaries. In an approach that uses double randomization, eligible farmers are randomly assigned to one of those four groups. The direct beneficiaries are those who will receive the project. The indirect beneficiaries are farmers who live in close proximity to the beneficiaries and are likely to benefit from spillover effects—e.g. farmers in the same community. The remaining farmers should not be directly or indirectly influenced by the project and are thus the control group (groups ii and iv). Since all farmers meet the eligibility criteria, the random assignment should ensure that the groups of farmers are alike in all ways except that the direct beneficiaries receive the project and the indirect beneficiaries receive indirect benefits through the beneficiaries. Comparing the direct beneficiaries to the control group and the indirect beneficiaries to the control group provides an estimate of direct project impacts (average treatment effect) and indirect project impacts (indirect treatment effects).

In some cases, indirect beneficiaries are not part of the project because they are not eligible. Even though they are not eligible, they may still be influenced by the program due to externalities or general equilibrium effects. For example, suppose a project targeting poor farmers decides to only provide support to farmers with less than one hectare of land. Farmers with more than one hectare of land living in the community with these poor farmers may indirectly benefit from the project because they may learn about a new beneficial technology (externality) or because the project may influence the price of land, labor, inputs or outputs (general equilibrium effect).

To capture the direct and indirect benefits of the project the four groups noted above are still required. In this case, groups (iii) and (iv) do not meet the eligibility criteria for the project and could theoretically only indirectly benefit. Suppose that most communities have farmers with less and more hectare of land and that a project is only going to target those with less than a hectare. Given these circumstances, capturing direct and indirect effects is possible if communities are randomly assigned to treatment and control. Baseline and post-treatment information should be in treatment and control communities, including eligible households and ineligible households. Assuming there are enough communities that there are no issues with sample size, eligible farmers in treatment communities should be similar to eligible farmers in control communities except for having received the project. Similarly, ineligible farmers in treatment communities should be similar to ineligible farmers in control communities except for having received the indirect benefits of the project. Comparing treatment and control eligible farmers will capture direct effects of the project and comparing treatment and control ineligible farmers will capture indirect effects of the project.

As with any evaluation, capturing the impact of a project is easier when experimentally designed. But it is possible to use non-experimental methods as well to capture spillover effects. As with an experimental approach, the key is to identify a control group that can represent a reasonable counterfactual to the indirect beneficiaries of the project. Similar to identifying direct effects, the best approach to use depends on the characteristics of the project and the available data.

4.4 Considering the Methodological Options

This section has highlighted the methodological options available for evaluating impact. As should be clear from the discussion, the methodological approach depends both on the characteristics of the project as well as the data that is available or can be collected. When possible, a common practice among evaluators is to employ multiple methods or different specifications to identify the impact of a project. The project characteristics and the data available of course limit options, but when possible this is considered a good practice since it helps to ensure that the estimated impact is correctly identified. The expectation is that any valid evaluation approach should bring about similar results and reflect the true impact of the project. By using multiple methods or difference specifications, the robustness of the estimate of impact can be checked. In designing an evaluation strategy as part of the project design, it is often advantageous to consider not just what individual approach might be best, but what alternatives might be needed or desirable to verify the accuracy of impact estimates.

In this guideline, before-after comparisons are not considered proper estimators of a project's impacts. Before-and-after, or reflexive, approaches to evaluation compare indicators of impact for beneficiary farmers from before the project is implemented to after the project is in place. Although these types of comparisons are often used in ex-post evaluations, they remain problematic. It is worth noting the issues in using this approach.

Recall that the evaluation problem is one of requiring a counterfactual that represents beneficiary farmers in the absence of the project. In the case of a reflexive comparison, the counterfactual is the beneficiary farmers themselves, but before the project. This may seem like a reasonable approach since clearly beneficiaries at one point in time are similar to themselves at a later point in time. The problem, however, comes precisely because of this time dimension. Comparing beneficiaries before and after a project assumes that no other changes have occurred over time other that the initiation of the project and, therefore, that all changes in impact indicators can be attributed to the project.

This is a strong assumption particularly in agricultural projects. Weather patterns differ every year meaning an indicator such as yields will undoubtedly change over time. Additionally, input and output prices also tend to change every year, which influence production choices and therefore the return to agriculture. A reflexive comparison is going to capture not only the impact of the project, but all of these other changes. Any ex post evaluation using this approach is likely to be biased and it is impossible to separate general time trends from actual project impact. Using such an approach for evaluating the impact of agricultural projects is not advisable.

5. Collecting Agricultural Data

Collecting farm-level agricultural data is complicated by the fact that agriculture is complex, often involving multiple products (crops and livestock), numerous plots, distinct seasons, and a range of inputs. The fact that agriculture is a self-employment activity also means that it is difficult to ascertain income from the activity without carefully determining revenues and costs. The farm household often consumes much of the product of production making the valuation of the output challenging. Further, farmers are rarely involved solely in agriculture and understanding the impact of a project on farmers frequently requires looking at an agricultural households overall livelihood strategy to see if labor or other resources have shifted as a results of the project. The logistics of collecting data from farmers can also be complicated by the fact they tend to be widely dispersed. In this section of the guideline, some suggestions are provided for collecting agricultural data for impact evaluations. Of course in considering the data to collect, it is necessary to keep in mind the indicators previously identified to assess impact and the approach that will be used for identifying impact. The data collected will be used to create variables that are either used as impact indicators or as part of the analysis.

Following the discussion in the previous section, the expectation is that any reasonable impact evaluation will be designed to have at least two rounds of data collection: a baseline and post-intervention data collection. More rounds of data collection are possible and can be quite useful, especially if trying to distinguish short- and long-term project impacts. One of the general rules of collecting data in multiple rounds is to try and maintain the same format for questions and the same type of questions. Changes in the questionnaire can result in differences over time being due to changes in the way questions were posed rather than in changes in the underlying variable of interest. It is critical to start with a well designed questionnaire and to avoid changing that questionnaire over time. A common starting point for evaluating projects is to use a standard questionnaire that has already been administered in the country or in a similar country and has already been field tested.

The data collection for assessing impact of farmer-targeted projects focuses primarily on obtaining information from detailed questionnaires of farmers, including the treated and control group farmers. Other information can also be collected from community-level surveys or market-level surveys as appropriate. Here, the focus is on the data collection via questionnaires administered to farm or agricultural households, although other complementary surveys should be considered. The discussion is divided into two parts. The first focuses on the procedures for data collection while the second discusses the details of the farmer questionnaire.

5.1 Designing Data Collection Procedures

5.1.1 Timing and Periodicity

Two of the main aspects to be considered when designing the data collection procedure are timing and periodicity. The timing refers to the period (month, year) in which the data will be collected. There are three major aspects to be considered when deciding the best timing to collect agricultural data and all of these vary depending on the type of project, the indicators to be measured and the project targeting. First, an agricultural survey should request information regarding the previous agricultural year which usually does not coincide with a calendar year. Then, the survey should be planned accordingly to be collected once the agricultural year has finished. This includes gathering information about the crops, inputs, parcels and costs among other things for all the seasons in the last agricultural year. Administering the questionnaire at the end of the season reduces the recall period and the measurement error as it enhances accuracy on the estimates of inputs used, production sold, prices, and so on.

Second, the baseline should be administered in order to cover information regarding the agricultural year that was not yet affected by the intervention. Even if the intervention is already in place in a given year, the previous agricultural year might not have been affected by the intervention raising the possibility of collecting the baseline after the project has initiated. This is only acceptable if it is clear that the initiation of the project will not have influenced previous production or the manner in which questions are answered.

Third, the surveys should be administered after the harvest for the main agricultural season takes place. Usually, most of the agricultural projects are focused on increasing productivity and agricultural income, therefore, it might be problematic to collect data when the main season crops has not yet been harvested as farmers are forced to predict expected production that might generate inaccurate estimates of agricultural productivity and increase measurement error. On the other hand, if the main focus of a project is to improve access to markets and therefore, increase commercialized production; the data collection must be set up after the majority of the agricultural sales have taken place. In this case, collecting data right after the end of the harvest might not be advisable because the commercialization stage might have not yet started. Then, farmers might not be able to provide accurate values about the amount of production sold, the commercialization costs or the market prices. The follow-up surveys should always be collected in the same period when the baseline was collected.

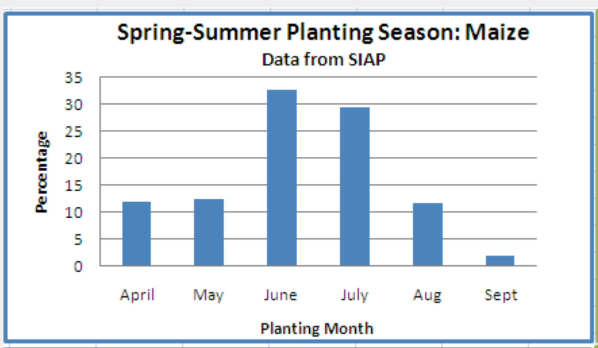
Of course, making multiple visits to the farm over the agricultural calendar to complete the questionnaire is possible. Other approaches such as farmers keeping logs of activities and inputs used may also be an alternative. These approaches allow information to be collected at key points in time (end of the first season, at planting at harvest, etc.) and should improve the accuracy of information collected. If such approached are used, the total information required and the timing of collecting individual pieces of information should be determined ahead of time. Further, it is important to maintain consistency checks to be sure questions refer to the same plot and crop. A benefit beyond less recall error and thus better data is that each visit should be shorter taking less time of the farmer or the farmer's family. However, this approach tends to be much more costly and there is a risk that the farmer refuses to continue answering questions at some stage. Single visits covering the agricultural calendar year tend to be most common given the cost advantages.

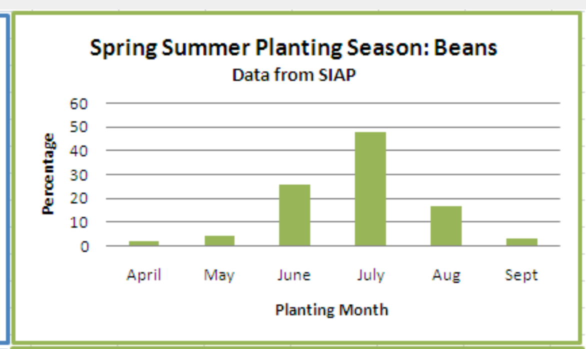
Since the agricultural year does not coincide with the calendar year and harvesting times tend to vary by region and crop, to decide the best timing to collect agricultural data, it is recommended to consult a crop calendar where information regarding planting and harvesting for different crops and regions can be obtained. Box 5 provides an example of how the information from a crop calendar is useful when deciding the timing for data collection.

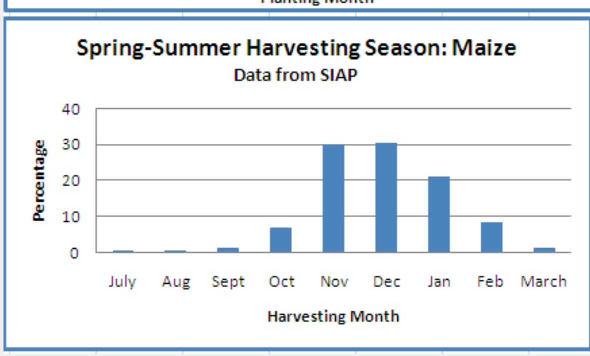
Box 5. Crop Calendar and Data Collection

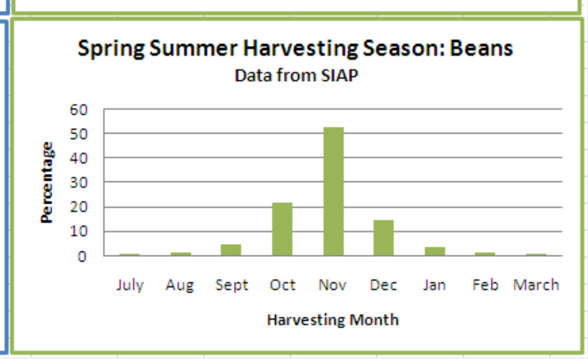
PROCAMPO, the largest agricultural program in Mexico, provides cash transfers to farmers who cultivated one of nine grains or oilseeds prior to the initiation of the North American Free Trade agreement in 1994. The new phase of PROCAMPO involves a comprehensive impact evaluation to be implemented from 2010-2013. The data collection procedure involves collecting information from a nationally representative sample of nearly 10,000 agricultural producers. The timing for administering the survey was considered carefully given the diverse range of products cultivated by the program beneficiaries. In fact, a Mexican crop calendar obtained from the SIAP (Agrifood and Fishery Information Service) served as the main tool to decide when the field work should take place.

The first stage in considering the crop calendar and when to administer the survey was to identify the main crops produced by the PROCAMPO beneficiaries. The most important crops were found to be maize and beans. Therefore, the data collection strategy is largely based on the agricultural cycle associated with these two crops. The second step was to identify the main agricultural season. In Mexico, there are two agricultural seasons for annual crops: Spring-Summer and Fall-Winter seasons. For beans and maize, most of the planting for the Summer-Spring season takes place from April to September and the harvesting from October to February. On the other hand, for the Fall-Winter season, most of the planting takes place from October to February and the harvesting from March to July (SIAP, 2010[*]). The figures below show the distribution for planting and harvesting in the main agricultural season for maize and beans by month. This information suggested that the best timing for collecting agricultural data in Mexico is in February and March right after the completion of the harvest for the main crops.









[*] SIAP: http://www.siap.gob.mx/index.php?option=com_content&view=article&id=10&Itemid=15

The next aspect to consider when collecting agricultural data is periodicity. This refers to the time between the baseline administration and the follow-up surveys. The main factor that influences the periodicity of data collection is the estimated time that it is expected to take for the project to have an impact. This is particularly important when there is a limited budget that includes resources for a baseline and one follow-up survey only. If this is the case, it is crucial to time the follow-up survey after the project impact is expected to occur. Otherwise, the evaluation might not be able to detect any impact and disregard the importance of the project when the actual problem was the timing of the follow-up survey. This requires a broad knowledge of the project and its effects as well as previous empirical evidence.

For instance, it is expected that cash transfers, subsidies or direct support payment projects might have immediate to short-term effects, therefore, administering a follow-up survey two years after a projects is implemented might be adequate. On the other hand, technology transfer and land titling projects are expected to have medium to long-term effects. For the case of technology transfers, farmers need to learn how to incorporate the technology and how to use it appropriately; this might take time to produce the expected results on productivity. In the case of land titling projects, the impact on agricultural productivity is likely to occur through greater access to credit and therefore, investment. Hence, to identify any impacts in productivity, the follow-up survey must be timed accordingly.

5.1.2 Pilot Testing and Survey Preparation

Besides administering a baseline and a follow-up survey, the data collection strategy must also include planning for piloting the survey. The main purpose of pilot testing the survey is to check the validity of the questionnaire by finding questions or words that might be misinterpreted, misread or misunderstood as well as to check the functionality of the questionnaire in the field. Questionnaires often have procedures for quality assurance such as checklists that verify all questions are asked and that questions are consistent across section. For example, if a survey has pages of questions on crop inputs for all plots, it is important to verify all sections ask about all plots. Along with checking questions, data quality assurance protocols can also be assessed.

The number of pilot trials depends greatly on the amount of resources available to finance the data collection. However, it is recommended to have at least two pilot trials before the baseline is implemented and before every follow-up survey. The size of the pilot trials might vary from project to project and depends on the sample size of the survey which, of course, depends on the survey stratification. In general, in order to assess the validity of the questionnaire in different settings, it is important to include a heterogeneous group of producers to make sure that the questionnaire is tested for all types of farmers (subsistence farmers, cash-crop farmers, landowners) in different regions and for different crops. For a localized survey in which there is limited variation in the types of households as few as 25-30 tests of the survey could be done. For a national survey with significant stratification, it is recommended to have 100-150 surveys per pilot test being sure that households in each of the strata are included to a sufficient degree.

The pilot test is also useful to begin considering the logistics for the survey administration. The pilot tests resemble the data collection process in the field so can be used to consider the best manner to organize both data collection and data entry. As a general rule, the sooner the data entry is after data collection the better since any mistakes can be dealt with in the field. A common model for administering a survey team is with a group of approximately five with one person as a leader, one that does the data entry on a laptop in the field and the remaining team full time enumerators. Any mistakes in data collection can be fixed immediately by sending enumerators back out. This procedure can be checked during pilot testing. Every group of interviewers should be familiar with the areas in which their questionnaires will be administered and the routes that will be followed during the data collection process. Also, every group member should have a calendar in which the data collection process is described as a whole as well as for the specific group to which they belong.

One of the most important stages prior to conduct any type of field work is the training of interviewers and supervisors. This can be done as part of the piloting of the survey or separately. The main purpose of the training is to familiarize the interviewers with the survey instrument. To accomplish this, it is important to include in-the-classroom and in-the-field activities to practice with other interviewers as well as with agricultural producers. Also, it should include GPS training, cartography training, PDA training, etc. when any of these instruments is used in the survey implementation. Generally, it is advisable for the interviewers to have some type of knowledge in agriculture as it could be useful to detect any inconsistent information provided by the producer. The length of the training could vary depending on the length and the difficulty of the questionnaire. However, to collect agricultural household data is complex and requires in depth training. Additionally, the training should not only cover activities related to questionnaire administration, but also with respect to the management of the logistics in the field.

Finally, every data collection procedure should have a timeline of activities where all the responsible actors and dates when the activities will be implemented are described. This timeline should be widely known by all the actors involved in the process. Box 6 provides an example of an impact evaluation timeline.

Box 6. Impact Evaluation Timeline: An Example

The timeline in the table below describes the main activities, the responsible actors and the dates in which the activities will be conducted for the PROCAMPO impact evaluation. Overall, three main actors are in charge of the impact evaluation activities: the Inter-American Development Bank (through the Strategy Development Division and the Rural Development, Environment and Natural Disasters Division), the consulting firm in charge of field work activities and INEGI (National Statistics and Geography Institute of Mexico) that will create the sample based on the 2007 Agricultural Census.

ACTIVITIES	RESPONSIBLE	DATES
Baseline Administration		
Questionnaire Design	IDB	Nov 2009-April 2010
Pilot Trials (1st round)	Consultant Firm	Sept-Oct 2010
Pilot Trials (2nd round)	Consultant Firm	Oct-Nov 2010
Sampling Design and Methodology	INEGI	June-Aug 2010
Training for interviewers and supervisors.	Consultant Firm	Nov-Jan 2011
Preparation of materials for survey administration.	Consultant Firm	Nov-Jan 2011
Baseline Administration	Consultant Firm	Feb-March 2011
Data Entry	Consultant Firm	April-May 2011
Baseline Data	Consultant Firm	May 2011
Baseline Analysis	IDB	June-Sept 2011
Follow-Up Survey		
Questionnaire Design	IDB	Nov 2011-April 2012
Pilot Trials (1st round)	Consultant Firm	June 2012-July 2012
Pilot Trials (2nd round)	Consultant Firm	Aug-Sept 2012
Training for interviewers and supervisors.	Consultant Firm	Oct 2012-Jan 2013
Preparation of materials for survey administration.	Consultant Firm	Oct 2012-Jan 2013
Follow-up Survey Administration	Consultant Firm	Feb-March 2013
Data Entry	Consultant Firm	April-May 2013
Follow-up Data	Consultant Firm	May 2013
Impact Evaluation Analysis	IDB	June-Sept 2013

5.2 Questionnaire Design

The survey instrument for every impact evaluation should be carefully designed to gather all the information needed to test the hypotheses underscoring the project logic as well as to construct other control variables for impact estimation. A questionnaire to collect agricultural data should usually be based on the agricultural household model. This model suggests that agricultural production units behave as both a firm and a household and, in the presence of market failure, production decisions cannot be separated from household consumption decisions.[10] Hence, collecting household as well as farm information should be a priority in every questionnaire that aims to analyze agricultural household behavior. In general, the modules to be included should obtain a full picture of the household livelihood strategies, including crop and livestock activities but also non-agricultural activities, migration and transfer receipt among others.

The first step in designing a questionnaire is to consider all the variables needed for the analysis, including, as noted, all the indicators required for assessing impact as well as control variables. Many of these variables may not be measured directly so a series of questions may need to be asked to obtain a single key variable. For example, questions on crop profits or gross margins require having information on revenues and costs of crop production. Questions that directly ask this information are unlikely to be reliable so it requires asking a series of questions related to revenue and costs that help to obtain the information to create the variable required. The second step is to consider all the information that will be needed to create the final set of variables. Ultimately, the usefulness of each question in the questionnaire should be justified and their uses for the analysis clear.

With respect to the agricultural data, the agricultural module should generally be the most extensive. Unless focusing on a specific crop, this module should gather information at the parcel and the crop level for every agricultural season separately. If focusing on a specific crop, it may be possible to reduce other agricultural details, but a degree of information is still required. Again, this must be driven by the variables required for the analysis. Agricultural data often varies by type of crop, therefore, requesting information for annual and permanent crops separately is recommended in order to obtain more precise estimates for input use and production. The agricultural module should request information regarding input use, input amount, input prices, output, output utilization, commercialization and home-consumption,

[10] See Singh, Squire and Strauss (1986) for details of agricultural household models.

shocks and agricultural assets. This module should be complemented with a section where the general characteristics of the parcels are described and the terrain location is sketched by the farmer with respect to the household dwelling. Frequently, it might be useful to measure some percentage of the sampled parcels with a GPS to calculate precise estimates for land size.

Overall, it is important to consider that administering an agricultural household questionnaire is challenging because it is extensive and the level at which the information is obtained varies from module to module. For instance, dwelling characteristics are gathered at the household level, household demographic characteristics are gathered for every household member and agricultural production information is gathered for every parcel and every season separately.

Given that administering an agricultural household questionnaire is already a complex task, it should be carefully designed to be functional in the field. For example, binding the questionnaires and limiting the number of open-ended questions would dramatically increase the interviewers' productivity which ranges from 2 to 5 questionnaires per day. This is important to consider when estimating the budget for data collection. In addition, administering other questionnaires to community leaders or input providers might be useful to complement the information provided by the producers' questionnaire including access to services, infrastructure, social services, access to markets, prices, etc.

Table 3 provides a general outline for an agricultural household questionnaire. The length of the different modules tends to vary depending on the focus of the project. For instance, a project that aims at linking farmers to markets should have an extensive commercialization module while a project that aims to reduce barriers to credit access should have an extensive credit and savings module.

Table 3. Questionnaire Modules: General Description

MODULES	DESCRIPTION
Module 1: Household Demographic Characteristics Sec. 1.1: Socio-demographic information Sec. 1.2: Schooling and Education	• Sec. 1.1 should include questions regarding the sex, age, nationality, education, etc., of the household members. • Sec. 1.2 is only for the household members who are 7 years or older. This Sec. request information regarding literacy and years of education for each of the household members.
Module 2: Parcels Sec. 2.1: List of Parcels Sec. 2.2: Owned Parcels: General Information Sec. 2.3: Rented Parcels: General Information Sec. 2.4: Parcels Borrowed to and from Others: General Information. Sec. 2.5: Parcels Bought and Purchased: General Information Sec. 2.6: Investments	• Sec. 2.1 request basic information regarding each of the parcels sown during each of the agricultural seasons (name, land size, crop planted, type of ownership). Also, it usually provides a space in which the farmer can identify the location of each parcel with respect to the household dwelling. It is important to include information about parcels rented, owned, borrowed from and to others, as well as those planted with annual and permanent crops. This Sec. facilitates the identification of each parcel by the farmer. • Secs.2.2 to 2.5 request information regarding general characteristics of owned, rented to others, rented from others, sold and purchased parcels respectively. • Sec. 2.6 requests information regarding the investments carried out in the different parcels during a certain period of time (ex. 5 years).
Module 3: Agricultural Module ❖ **Main Season Information** Sec. 3.1: Crops: General Information Sec. 3.2a: Seeds : Annual Crops in Main Season. Sec. 3.2b: Fungicides: Annual Crops in Main Season. Sec. 3.2c: Other Inputs: Annual Crops in Main Season Sec. 3.2d: Labor: Annual Crops in Main Season Sec. 3.2e: Production: Annual Crops in Main Season. Sec. 3.2f: Commercialization: Annual Crops in Main Season ❖ **Secondary Season Information** Sec. 3.3a: Seeds : Annual Crops in Secondary Season Sec. 3.3b: Fungicides: Annual Crops in Main Season Sec. 3.3c: Other Inputs: Annual Crops in Main Season Sec. 3.3d: Labor: Annual Crops in Main Season Sec. 3.3e: Production: Annual Crops in Main Season Sec. 3.3f: Commercialization: Annual Crops in Main Season ❖ **Permanent Crops Information** Sec. 3.4a: Production and Commercialization of Permanent Crops Sec. 3.4b: Inputs and Labor Used for Permanent	• Sec. 3.1 provides basic information regarding the layout of the different crops and general information such as type of crop, seasons, sown area, Access to irrigation, etc. • Sec. 3.2a -3.2c request agricultural information regarding the permanent crops planted during the main season. Specifically, information about amount of land planted, amount of inputs used (seeds, insecticides, fungicides, herbicides, insecticides and machinery), cost of inputs and access to inputs is inquired. • Sec. 3.2d request information about the amount of labor used for agricultural production. The questionnaire should include paid and family labor for each type of agricultural chore (ex. Sowing, input application, harvesting, etc). • Sec. 3.2e and 3.2f requests information about the amount of output obtained, commercialized, used for seeds, stored and used for home-consumption. • Secs.3.3a -3.3f are the equivalent to Secs.3.2a-3.2f for annual crops planted during the secondary season. • Secs.3.4a and 3.4b request information regarding the production, input use, labor and commercialization of permanent crops. • Sec. 3.5 requests information regarding the producer's agricultural history in a given period of time (ex. 10 years). This Sec. includes questions regarding the crops planted by the farmer in a given period of time. • Sec. 3.6 requests information regarding shocks faced by the household that might have affected agricultural production in the last agricultural year. • Sec. 3.7 requests information regarding access and

Crops. ❖ **Other Agricultural Information** Sec. 3.5 : Agricultural Production: History Sec. 3.6: Agricultural Shocks Sec. 3.7: Access to Agricultural Assets	ownership of the main agricultural assets such as tractor, harvester, etc.
Module 4: Livestock Sec. 4.1: Livestock: General Accounting Sec. 4.2: Livestock Products	• Sec. 4.1 requests information regarding the purchases and sales of livestock as well as expenditures related to livestock maintenance. • Sec. 4.2 requests information regarding the purchase, sale, costs, home-consumption, etc. associated with livestock products such as milk, eggs, etc.
Module 5: Economic Activities Sec. 5.1: Housework Activities Sec. 5.2: Jobs and Occupations Sec. 5.3: Microenterprise Sec. 5.4: Self-Employment	• Sec. 5.1-5.4 request information about economic activities outside the farm for all the household members including occupation, income, time allocation, etc.
Module 6: Migration Sec. 6.1: Domestic Migration Sec. 6.2: International Migration	• Secs.6.1 and 6.2 request information about domestic and international migration. Questions about destinies, remittances, costs, etc.,
Module 7: Credit, Savings and Loans Sec. 7.1. Savings Sec. 7.2. Loans offered to others Sec. 7.3: Loans requested	• Sec. s 7.1-7.3 request information regarding the types of savings, the loans offered to others, the loans requested by the household members, the interest rate paid, etc.
Module 8: Consumption Sec. 8.1: Vegetable, Fruits and Cereals Consumption Sec. 8.2: Processed Food and Meat Consumption Sec. 8.3: Food Security Sec. 8.4: Other non-Food Expenditures Sec. 8.5: Household Assets	• Sec. s 8.1-8.6 request information regarding food consumption (fruits, vegetables, processed food and meat), non-food consumption (clothes, cleaning supplies, alcohol, cigarettes, meals outside the home, etc), food security and household assets.
Module 9: Other Income Sec. 9.1: Government and NGO Support	• This module request information regarding access to basic government programs by the household members, pensions, NGO support, unemployment insurance, etc.
Module 10: Agricultural projects Sec. 10.1: Agricultural Project Information	• It is usually recommended to include a module on the agricultural intervention. This Sec. should ask questions regarding the participation in the project such as the amount of financing obtained by the project, the number of years in the project, the type of media through which the farmer obtained information about the project, the transaction costs, the transportation costs, etc.
Module 11: Social Capital Sec. 11.1: Access to Social Capital	• This module requests information about the farmers' access to social capital. Questions regarding the participation in cultural and religious activities as well as agricultural organizations and community groups by the household members should be asked.
Module 12: Dwelling Conditions Sec. 12.1: Dwelling Conditions	• This Sec. request information regarding the type of dwelling, distance to important places such as school or hospital, rent, access to public services, etc.

5.3 Basics of Sample Design

The sample design is one of the most important activities in every impact evaluation. Usually, it is desirable for the sample to be designed by a specialized agency such as the National Statistical Office which tends to have access to census data that could be used as a sample frame. Besides, these governmental agencies have access to first-hand and second-hand information that might be useful for sample calculations.

The sampling strategy should be carefully planned to guarantee enough statistical power to allow identification of project impact as well as to analyze the heterogeneity of impact among producers. Assessing the heterogeneity of impact is common in agricultural projects. Different types of projects can have diverse effects on different types of producers and it might be in the best interest of the project team to identify and measure these heterogeneous outcomes separately. To accomplish this, stratifying the sample to assure representativeness within different groups in the population (strata) is advisable. For instance, it might be in the best interest to obtain estimates of the impact by region, land size, type of land tenure or access to irrigation. To be confident that inference for every of these subgroups is statistically valid, stratification is recommended.

To assure enough statistical power to identify project impact, statisticians tend to rely on the power formula. The power formula is composed by four pieces of information: (i) expected effect on outcome variable; (ii) outcome variable's standard deviation; (iii) confidence level (usually 95%), and (iv) statistical power. The power formula for an impact evaluation is as follows (World Bank, 2007):

$$N = \frac{4\sigma^2\left(z_\alpha + z_\beta\right)^2}{D^2} \qquad (7)$$

where D = is the impact on the outcome variable measured as the difference in means,

σ = the standard deviation,

z_α = the critical value of a confidence interval (two tail test=1.96),

z_β = the critical value of the statistical power (two tail test=1.28).

To reduce sample dispersion and therefore, administrative costs; it is a common practice to obtain the sample in two stages. In the first stage the principal sampling units (PSU) are randomly selected and in the second stage, producers within the already selected PSU are randomly chosen. These PSUs are geographical areas previously defined. However, using this two stage process involves correction for intra-cluster correlation which increases sample size (World Bank, 2007). Intra-cluster correlation correction is required because the agricultural production units within the same cluster (PSU) tend to be similar. The formula to correct for intra-cluster correlation is:

$$N_{corrected} = N\left[1+\rho\left(m-1\right)\right] \quad (8)$$

where ρ = the intra-cluster correlation

m = the number of agricultural production units to be interviewed in each cluster

Finally, given that most of the data collection procedures for impact evaluations involve collecting a baseline and a follow-up survey for the same producers, a non-response/attrition adjustment might be needed. The attrition rate refers to the number of agricultural production units that cannot be interviewed in the follow-up survey because these are not found by the interviewers. The non-response and attrition correction varies among countries and regions. For instance, a greater attrition correction might be adequate in a country with higher levels of migration.

6. Writing an Evaluation Plan

To finalize decisions related to data collection (Section 5) and methodology (Section 4) for the impact evaluation of the selected project, an evaluation plan should be written. The plan serves the purpose of planning out in detail the activities required to do the evaluation as well as providing guidance for future project leaders, government officials or evaluators who take over the work to know what was intended. Levy and Cossens (2011) have put together a guideline for writing an evaluation plan and the details of this document should be considered in writing the plan. As noted in the guidelines, the three key technical ingredients of the plan are the choice of indicators, selection of impact evaluation method, and preparation of a sampling and data collection plan. These will form the key components of the evaluation plan.

Along with conceptual discussion of these issues, the plan must also include the practicalities of the evaluation. The three key aspects to be dealt at the design stage to ensure a successful implementation of an impact evaluation are constructing the work plan and identifying the evaluation team, budgeting for an evaluation, and getting buy-in from counterparts. The evaluation plan should be clear on each of these components.

The evaluation plan provides a detailed guide for how the evaluation of the agricultural project will be completed. It should be sufficiently clear so that if project team leaders change or the evaluation team changes, those reading the plan can understand the intended approach to implementing and doing the impact evaluation. The next three sections of this guideline provide examples of evaluation plans from three different cases.[11]

[11] For more examples on impact evaluation plans please refer to the IDB´s Development Effectiveness website: http://www.IDB.org/en/topics/development-effectiveness/development-effectiveness,1222.html

7. Case Study: Dominican Republic Tchnology Adoption Program PATCA II[12]

This section describes the most important aspects of the methodology and the execution associated with the impact evaluation of the Agricultural Innovation Support Program (PATCA II) being implemented in the Dominican Republic. This case study is particularly relevant as it confirms the important link that must exist between the impact evaluation design and the program design. Specifically, this case study presents the first technology transfer program that, to our knowledge, will implement a two-stage randomization in order to evaluate the program direct and indirect impacts.

7.1 The PATCA II Intervention

Initially, PATCA was implemented in 2002 with the principal objective to reduce poverty and enhance the efficiency and competitiveness of the Dominican agriculture. Although PATCA can be categorized mainly as a technology transfer program, this initial phase included three components: (i) technology adoption support; (ii) food safety and (iii) technical assistance for institutional reform. The total cost of the program was US$61 million dollars and US$55 million were financed by IDB.

The first component, which provided financial support to farmers in order to promote technology adoption, funded five different types of technologies: (i) land leveling; (ii) modernization of water and irrigation techniques; (iii) zero tillage or no-till farming; (iv) pasture conservation, and (v) introduction of new tree species. These technologies sought to assist rice, vegetable, fruit and tuber producers as well as livestock breeders. This component provided a total of 15,882 subsidies, 5,234 (33%) for modernization of irrigation techniques, 2,649 (17%) for land-leveling, 3,567 (22%) for adoption of new tree species, 703 (4%) for zero or non-farming, and 3,669 (23%) for pasture conservation and rehabilitation (PATCA Executing Unit in Gonzalez, et. al, 2009).

[12] This section was written by Lina Salazar and Alessandro Maffioli of the Strategy Development Division of the Inter-American Development Bank.

The process by which the executing unit provided the financial support for technology adoption consisted of five stages:

1. The producer decided the type of technology to be adopted.
2. The field agent visited the producer to check the eligibility criteria and verify that the land and geographic conditions were suitable to implement the technology.
3. The application for technological support was approved or denied.
4. If the application was approved, the producer selected a provider from a list previously agreed with the executing unit. A certain percentage of the cost of the technology was financed by the government (60-85% depending on the type of technology and its environmental effects). The rest of the cost was financed by the producer in cash or kind.
5. The field agent verified that the technology was appropriately installed and utilized.

To assess the impact of PATCA, the Office of Evaluation and Oversight (OVE) at the IDB implemented an impact evaluation for the first phase of the program (Gonzalez, et. al, 2009). This analysis used PSM methods to identify the program effect on agricultural productivity. The results suggest a positive impact on rice productivity and breeding. However, the analysis does not find any significant impact on other crops productivity or milk producers.

Although this study represents an important effort to identify the impact of this program, there are four fundamental characteristics of the evaluation that might have influenced the identification of the program impact and therefore, need to be considered. First, the analysis conducted by OVE has a small data set with few observations, which might make difficult the identification of the impact due to sampling power limitations. This also limits the possibility of performing any type of heterogeneity of impact, which is critical to consider in this type of programs. Second, the program was not randomly implemented and the authors do not have a panel data or any type of instrument that allows them to control for unobserved heterogeneity that might influence both program participation and outcome indicators. Third, the data set used does not provide enough information on farm, producer or household characteristics. Fourth, the program implementation does not allow identifying any type of spillover effects.

Given the high demand for PATCA I by agricultural producers, the government decided to implement the second phase of the program (PATCA II). This time, however, the program will exclusively focus on the technology transfer component. Specifically, PATCA II aims at providing financial and technical support for the adoption of technologies associated with irrigation, land leveling, pasture conservation and greenhouse installation. Through the adoption of these technologies, the program intends to increase agricultural productivity and income.

7.2 The Logic of the PATCA II Intervention

This section provides an overview of the logic of the intervention. To accomplish this, in this section the existing literature on government financed technology transfer programs is analyzed. Specifically, a review of the theoretical and empirical support used to justify the implementation of the program will be presented with a focus on the problems to be addressed by the implementation of PATCA II. Second, the logic for implementing PATCA II is described by presenting the main hypotheses associated with the evaluation of program impact, the indicators of impact and the information to be collected in order to test the hypotheses and measure the outcome indicators.

7.2.1 Market Failure and Technology Diffusion Curve

PATCA II can be categorized as a technology transfer program or more broadly as an extension services program. The basic logic for implementing government financed technology transfer programs has been justified, from an economic perspective, by numerous empirical and theoretical studies. These justifications are mainly related to the presence of barriers that impede an optimal process of technology adoption which in turn influences economic growth and development. Jack (2009) categorizes these into two groups: (i) barriers that impede the adoption of technologies that are favorable from a social perspective because these do not provide enough economic benefits for private agents to implement them, and (ii) barriers that limit the adoption of technologies even if these are profitable from a private perspective.

PATCA II seeks to address the second type of barriers. There are numerous obstacles that limit the adoption of technologies that are economically favorable to agricultural producers: (i) credit constraints; (ii) lack of access to information or information asymmetries; (iii) risk aversion; (iv) lack of technology providers of dysfunctional supply; (v) human capital restrictions; (vi) deficient infrastructure; (vii) shortage of complementary inputs, etc. (Feder, Just and Zilberman, 1985). The first three types of barriers are the focus of the discussion as these are the ones that PATCA II aims to tackle.

The first barrier that PATCA II intends to address is related to credit market failures. Due to credit market imperfections, farmers are often liquidity constraints limiting their ability to undertake investments or adopt technologies that increase agricultural productivity and therefore, income. This type of market imperfections are frequent in rural areas in developing countries and various studies provide empirical evidence on the existence of these constraints such as Morduch (1994) for the case of India, Jalan and Ravallion (1999) for China, Carter (1989) for Nicaragua and Barham, and Boucher and Carter (1996) for Guatemala. Although it is traditionally assumed that credit constraints mainly affects technology adoption that requires a major initial investment (indivisible technologies), studies have shown that a negative relationship exists between credit constraints and the adoption of divisible technologies (Feder, Just and Zilberman, 1985). For instance, Simtowe and Zeller (2006) find that having access to credit increases the adoption of new maize varieties by credit constrained farmers in Malawi. Moser and Barrett (2003) show evidence that the presence of liquidity constraints reduces the probability that rice farmers' adopt new technologies in Madagascar. Similarly, Uaiene (2008) and Dercon and Christiaensen (2008) also confirm that credit constraints influence the adoption of divisible technologies.

The second barrier PATCA II helps overcome is related to risk aversion and uncertainty. Overall, risk aversion and uncertainty delay adoption because producers prefer a degree of certainty about the economic benefits associated with the technology before incurring costs. Hence, agricultural producers are more likely to postpone technology investment until they confirm the benefits through other farmers' experience. Various analyses confirm the negative relationship that exists between risk aversion and uncertainty and, technology adoption (Feder, 1980). For instance, Abadi-Ghadim, Pannell and Burton (2005) show that risk aversion reduces technology adoption by chickpea producers in Australia. Besley and Case (1994) develop a model to show that, under the presence of uncertainty, producers are less likely to adopt new high-yielding cotton varieties. The authors find that producers prefer to wait until other producers adopt the varieties in order to learn from them and confirm their expectations about the potential benefits.

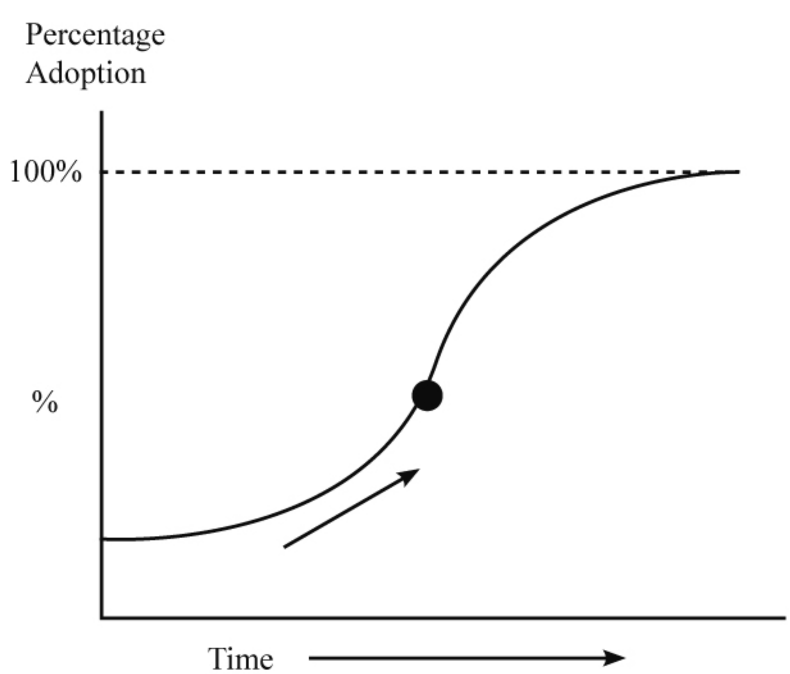
The last barrier that PATCA II aims to address is the lack of information or information asymmetries. This barrier limits the adoption of profitable technologies not only because farmers lack the information regarding the appropriate implementation and utilization of the technology in the field, but also because many farmers do not have information about potential providers, costs, etc. For instance, in Nepal, Joshi and Pandey (2005) show that farmers' valid or false perceptions about other rice varieties influence adoption decisions. The authors conclude that it is important to implement various methods to diffuse information regarding the benefits of the technologies and therefore, shape farmers' perceptions according to reality. The importance of subjective perceptions on technology adoption is also confirmed by Adesina and Zinnah (1993) for the case of rice farmers in Sierra Leone and by Adesina and Baidu-Forson (1995) for the case of Burkina Faso and Guinea. In the same line, Conley and Udry (2004) provide empirical evidence to show the importance of social learning on technology adoption for the case of pineapple producers in Ghana. The authors find that producers change input use patterns once they obtain information about the neighbor's production. Foster and Rosenzweig (1995) also present evidence to confirm that lack of knowledge and imperfect information are major barriers for the adoption of high yielding varieties in India. Specifically, producers with more experienced neighbors are more likely to adopt these technologies in a larger extension of land probably because they obtain information from them regarding technology adoption benefits.

To overcome these three barriers, PATCA II will develop three different strategies. The first strategy is to finance 50% of the cost of the technology through the public sector. This seeks to reduce liquidity constraints. The second strategy is to create a link between supply and demand by offering information about the different providers, costs, locations, etc. This strategy aims to limit information asymmetries. Lastly, PATCA II will provide technical support through field agents who will promote the program in the field, provide technical guidance and supervise the adequate implementation and utilization of the technologies. The last strategy helps to address risk aversion and uncertainty.

Along with these anticipated direct effects of PATCA II, the expectation is that PATCA II will have indirect or spillover effects on non-participants and, in fact such effects are a desirable impact of the program. As noted in section 4.4, direct effects correspond to the effect of the program on the treated producers while the indirect effects correspond to the effect of the program on the non-participating producers and take place when these producers are indirectly affected by the program participants or the program itself. There are various indirect effects that can take place with a technology transfer program including positive externalities and general equilibrium effects (Angelucci and Di Maro, 2010; Angelucci and De Giorgi, 2009; Avitabile and Di Maro, 2007). For the case of PATCA II, it is likely that non-participant producers could learn the benefits of the promoted technologies from treated farmers through direct observation, social interaction or other mechanisms. PATCA II could produce general equilibrium effects due to increases in income that might affect the aggregate demand and therefore, influence local economy outcomes. Finally, PATCA II could generate positive externalities associated with the different technologies that are being promoted.

Overall, the objective of PATCA II is to accelerate the process of technological diffusion and adoption in the rural areas in the Dominican Republic. The process of technological diffusion has been usually associated with an S-shaped curve which presents the relationship between time and adoption rate (See Figure 4). This curve implies that technology is adopted slowly until it reaches a “critical mass” at which the adoption process accelerates. PATCA II intends to accelerate the adoption process in order to reach this “critical mass” at which technology adoption expands on its own. This shortens the time period between initial adoption by innovators and reaching the adoption ceiling. Thus, the social benefits of adoption are felt earlier and accrue to both direct beneficiaries and potentially to non-participants in beneficiary communities.

Figure 4: Adoption Curve



7.2.2 Hypotheses to Test and Indicators

The first stage in any impact evaluation is to identify the hypotheses to be tested, which must be closely linked to the logic of the intervention and the problems that want to be addressed by the program implementation. Since PATCA II seeks to increase income and productivity through technology adoption, the final outcome indicators measure increments in income and productivity. The intermediate outcomes represent the channels through which the final outcomes will be achieved. In other words, how will the adoption of a given technology increase income and productivity? Although the answer can be directly linked to the adoption itself, the technology adopted might also create different paths through which income and productivity is increased. For instance, different types of technologies can increase input and machinery use that will be translated into higher productivity and therefore, higher income. Note because of the potential indirect effects of the program these hypotheses are tested for both direct and indirect program beneficiaries.

Table 4 shows the main hypotheses to be tested in the PATCA II impact evaluation as well as the indicators to be measured in the analysis. These indicators are divided into two groups: final outcomes and intermediate outcomes. The information needed to construct the indicators is also described. As seen, the main variables used to assess impact are from a data collection process designed for this purpose and described below.

Table 4: Hypotheses and Outcome Indicators

Hypotheses	Final Outcome Indicators	Measurement Frequency and Verification Method	Information
PATCA increases, on average, the beneficiaries' agricultural net income with respect to the non-beneficiaries. • Net Agricultural Income ($)	Difference in the percentage change of the average net agricultural income between beneficiaries and non-beneficiaries. $\left(\frac{\overline{Y}^{B}_{2014}-\overline{Y}^{B}_{2010}}{\overline{Y}^{B}_{2010}}\right)-\left(\frac{\overline{Y}^{NB}_{2014}-\overline{Y}^{NB}_{2010}}{\overline{Y}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through the income and agricultural modules included in the survey instrument.	▪ Total inputs used for production ▪ Price of inputs ▪ Price of production sold. ▪ Amount of production sold. ▪ Amount of production used for home consumption ▪ Amount of production used for seeds. ▪ Total labor used for production. ▪ Wages ▪ Price paid for renting/ buying agricultural machinery
PATCA increases the average total factor productivity (reproductive efficiency) of the beneficiaries with respect to the non-beneficiaries. • Total Factor Productivity • Reproductive Efficiency	Difference in the percentage change of the average total factor productivity (reproductive efficiency) between beneficiaries and non-beneficiaries. $\left(\frac{\overline{TFP}^{B}_{2014}-\overline{TFP}^{B}_{2010}}{\overline{TFP}^{B}_{2010}}\right)-\left(\frac{\overline{TFP}^{NB}_{2014}-\overline{TFP}^{NB}_{2010}}{\overline{TFP}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through a detailed agricultural module to be included in the survey instrument.	▪ Amount of inputs used. ▪ Amount of labor used. ▪ Amount of production obtained. ▪ Machinery used and rented. ▪ Other variables needed to calculate a production function including household, farm and producer characteristics.
PATCA increases the average value of production per hectare of the beneficiaries with respect to the non-beneficiaries.	Difference in the percentage change of the average value of agricultural production per hectare (milk production and/or livestock production) between beneficiaries and non-beneficiaries. $\left(\frac{\overline{V}^{B}_{2014}-\overline{V}^{B}_{2010}}{\overline{V}^{B}_{2010}}\right)-\left(\frac{\overline{V}^{NB}_{2014}-\overline{V}^{NB}_{2010}}{\overline{V}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through a detailed agricultural module to be included in the survey instrument.	▪ Amount of production obtained ▪ Land size ▪ Price of production sold.
PATCA increases the average yields per hectare of the	Difference in the percentage change of the average yields between beneficiaries and non-beneficiaries.	Ad-hoc survey to agricultural producers and livestock breeders. Baseline	▪ Amount of production obtained

PATCA increases the average yields per hectare of the beneficiaries with respect to the non-beneficiaries.	Difference in the percentage change of the average yields between beneficiaries and non-beneficiaries. $\left(\frac{\overline{R}^{B}_{2014} - \overline{R}^{B}_{2010}}{\overline{R}^{B}_{2010}}\right) - \left(\frac{\overline{R}^{NB}_{2014} - \overline{R}^{NB}_{2010}}{\overline{R}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through a detailed agricultural module to be included in the survey instrument.	▪ Amount of production obtained ▪ Land size
Hypotheses	**Intermediate Outcome Indicators**	**Measurement Frequency and Verification Method**	**Information**
PATCA increases the average value of inputs used for production per hectare of the beneficiaries with respect to the non-beneficiaries.	Difference in the percentage change of the average value of inputs per hectare between beneficiaries and non-beneficiaries. $\left(\frac{\overline{I}^{B}_{2014} - \overline{I}^{B}_{2010}}{\overline{I}^{B}_{2010}}\right) - \left(\frac{\overline{I}^{NB}_{2014} - \overline{I}^{NB}_{2010}}{\overline{I}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through a detailed agricultural module to be included in the survey instrument.	▪ Amount of inputs used. ▪ Price of inputs used. ▪ Amount of labor used. ▪ Wages ▪ Land size
PATCA increases the average percentage of commercialized production of the beneficiaries with respect of the non-beneficiaries	Difference in the percentage of commercialized production between beneficiaries and non-beneficiaries. $\left(\frac{\overline{U}^{B}_{2014} - \overline{U}^{B}_{2010}}{\overline{Total}^{B}_{2010}}\right) - \left(\frac{\overline{U}^{NB}_{2014} - \overline{U}^{NB}_{2010}}{\overline{Total}^{NB}_{2010}}\right)$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through detailed agricultural and commercialization modules to be included in the survey instrument.	▪ Amount of production obtained. ▪ Amount of production sold in the market ▪ Amount of production used for home consumption.
PATCA increases technological adoption by the beneficiaries and the non-beneficiaries in selected areas. ▪ Technology User (Y/N)	Difference in the percentage of technological adoption between beneficiaries and non-beneficiaries in non-selected areas. (Direct Effects). $\left(\frac{\overline{U}^{B}_{2014} - \overline{U}^{B}_{2010}}{Total^{B}}\right)^{ZB} - \left(\frac{\overline{U}^{NB}_{2014} - \overline{U}^{NB}_{2010}}{Total^{NB}}\right)^{ZNB}$ *Where U means users; ZB selected zone and ZNB non-selected zone. Difference in the percentage of technological adoption between non-beneficiaries (eligibles and non-eligibles) in selected areas and non-beneficiaries (eligibles and non-eligibles) in non-selected areas (Indirect effects or spillovers) $\left(\overline{U}^{NB}_{2014} - \overline{U}^{NB}_{2010}\right)^{ZB} \quad \left(\overline{U}^{NB}_{2014} - \overline{U}^{NB}_{2010}\right)^{ZNB}$	Ad-hoc survey to agricultural producers and livestock breeders. Baseline in 2011 and follow up survey in 2014. The information to measure this indicator will be collected through detailed agricultural and access to technology modules to be included in the survey instrument.	▪ Access to technology (S/N) ▪ Access to providers (S/N)

7.3 Evaluation Methodology

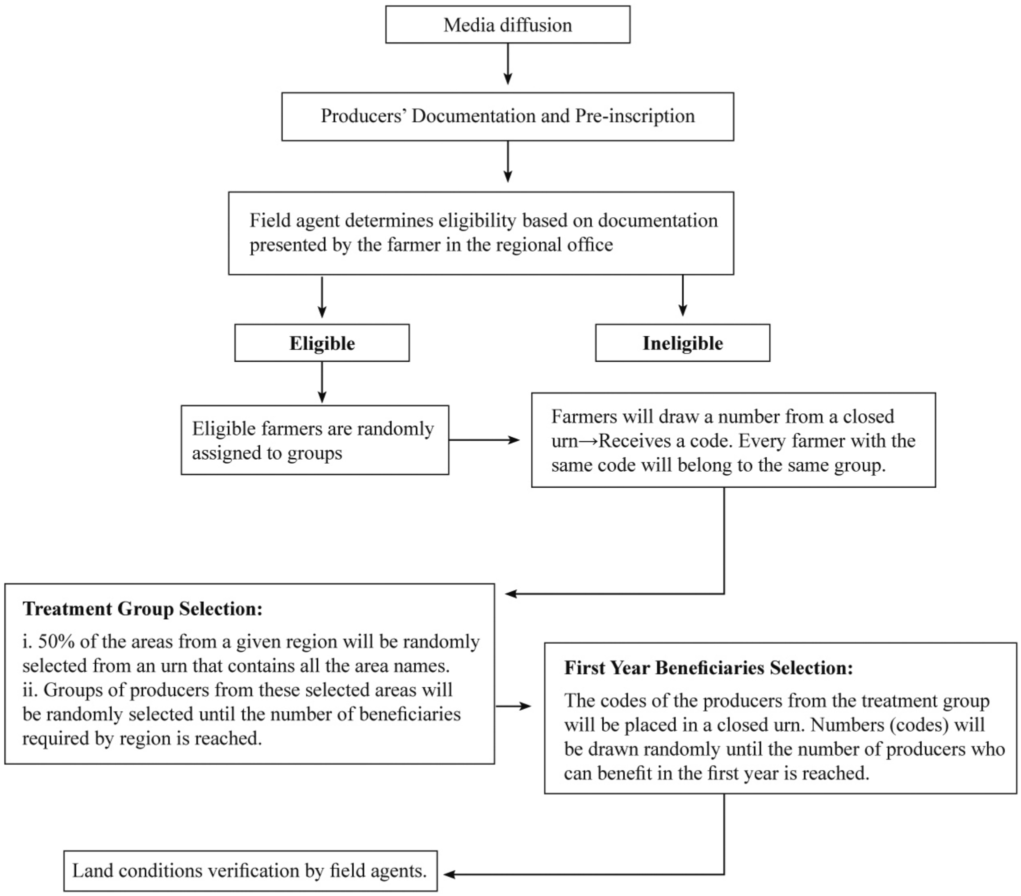
The evaluation methodology in any impact evaluation must be closely linked to the program design. In this particular case, the second phase of PATCA will use a randomization process to assign producers to treatment group. Therefore, an experimental methodology can be used to measure program impact. Randomization is possible to implement in this particular case because program demand is higher than the resources available for funding the technologies. Hence, using an experimental design provides a fair approach to assign financial resources to producers. Below, the process through which producers will be selected as beneficiaries is described followed by the econometric analysis and the sampling strategy.

7.3.1 Selection of Beneficiaries (Annual Randomization in Two Stages)

The process through which producers will be assigned to treatment and control groups was designed to identify the program direct and indirect effects. The identification and measurement of indirect effects represents one of the major challenges associated with this type of programs as it requires careful project design and implementation.

In order to identify the direct effect of the intervention it is necessary to create a control group of producers who are not exposed to the program directly or indirectly (Angelucci and Di Maro, 2010). Technically, this control group is uncontaminated by the program. Given that the program influence tends to be greater for those producers who are geographically closer to the treated farmers, it becomes a priority to randomize at the level of geographical areas. These geographical areas must represent the area of influence or the local economy where the indirect effects are more likely to occur (Angelucci and Di Maro, 2010). This type of randomization allows measurement and identification of the direct effect that can be confidently attributed to the intervention. The failure to design an evaluation that creates an uncontaminated control group might result in biased impact estimates.

Figure 5: Treatment Selection Process: Two-stage Randomization



To create an uncontaminated control group, the randomization of the beneficiaries for PATCA II will be implemented in two stages. In the first stage, the areas that will benefit with the program will be selected. Once the areas are selected, the second stage consists of selecting the producers that will be included in the program (treatment group).The treatment group will be chosen among all the eligible producers who are located in the randomly selected areas and fulfill the following requirements (Figure 5 presents the selection process of the treatment group):

a) Having an identification document
b) Performing an agricultural activity
c) Being a small-medium producer
d) Having legal tenure of the cultivated land

This selection strategy will allow the categorization of farmers into five different groups: (i) eligible beneficiaries in selected areas, (ii) eligible non-beneficiaries in selected areas, (iii) ineligible non-beneficiaries in selected areas, (iv) eligible non-beneficiaries in non-selected areas, and (v) ineligible non-beneficiaries in non-selected areas. Note that this strategy to identifying indirect effects includes the use of double randomization and eligibility criteria since the eligible farmers in the selected areas are randomly assigned to participate in the program (see section 4.4). This means that indirect effects on both those who meet program criteria and those that do not can be assessed. Table 6 clarifies the five groups and the observed effect for each group. Following this, the PATCA II direct effect is equal to (A-C) which represents the direct impact on eligible producers. The indirect effect on eligible producers is equal to (B-C) and the indirect effect on ineligible producers is (D-E). The total effect of PATCA II is the total sum of all the direct and indirect effects: (A-C)+(D-E)+(B-C).

Table 6: Budget for PATCA II Impact Evaluation

Year	Activity	IDB	TC	Loan	Local*	Total
2010	Questionnaire design.	BID (SPD/SDV)	0	0	0	0
	Senior Economist IDB	12,000	0	0	0	12,000
	Junior Economist IDB	12,000	0	0	0	12,000
	TOTAL 2010	**24,000**	**0**	**0**	**0**	**24000**
2011	Sample Design	0	0	10,000	0	10000
	Pilot Trials	0	27,500	10,000	2,500	40000
	Baseline for Producers (4000*US$70)	0	230,000	45,000	5,000	280000
	Baseline for Providers	0	20,000	20,000	5,000	45000
	Field Coordinator	0	50,000	0	0	50000
	Senior Economist IDB (24 days of work and 3 trips)	22,000	0	0	0	22000
	Junior Economist (includes 48 days of work and 3 trips)	27,000	0	0	0	27000
	TOTAL 2011	**49,000**	**327,500**	**85,000**	**12,500**	**474,000**
2012	Baseline Analysis	0	30,000	10,000	2,000	42,000
	Senior Economist IDB (24 days of work and 3 trips)	19,000	0	0	0	19,000
	Junior Economist IDB (includes 48 days of work and 3 trips)	24,000	0	0	0	24,000
	TOTAL 2012	**43,000**	**30,000**	**10,000**	**2,000**	**85,000**
2013	Pilot Trials	0	27,500	10,000	2,500	40000
	Field Coordinator	0	25,000	0	0	25000
	Senior Economist IDB (24 days of work and 3 trips)	22,000	0	0	0	22000
	Junior Economist IDB (includes 48 days of work and 3 trips)	27,000	0	0	0	27000
	TOTAL 2013	**49,000**	**52,500**	**10,000**	**2,500**	**114,000**
2014	Baseline for Producers (4000*US$70)	0	230,000	45,000	5,000	280000
	Baseline for Providers	0	20,000	20,000	5,000	45000
	Field Coordinator	0	25,000	0	0	25000
	Senior Economist IDB (24 days of work and 3 trips)	19,000	0	0	0	19000
	Junior Economist IDB (includes 48 days of work and 3 trips)	24,000	0	0	0	24000
	Other Expenditures	0	30,000	20,000	0	50000
	Data Analysis	0	30,000	10,000	2,000	42000
	TOTAL 2014	**43,000**	**335,000**	**95,000**	**12,000**	**485,000**
	TOTAL COST 2010-2014	**208,000**	**745,000**	**200,000**	**29,000**	**1,182,000**

Due to the small operational capacity of the regional offices which impedes the disbursement of resources to all the producers in the treatment group during the first year of execution, the program will be executed in four years. Producers from the treatment group will receive the benefits in different years during the implementation of the program. To address this issue, the program will have annual randomizations from the treatment group to select producers who will benefit with the program in a given year. Hence, all the farmers will know in advance who will benefit with PATCA during the four years (treatment group), but only the farmers who are selected in the first year will know when they will obtain the benefits. The reason behind selecting all the producers who will belong to the treatment group in the first year prior to the implementation is to remove incentives to delay adoption due to the expectation of being selected in the future. Failure to accomplish this might create negative externalities on adoption decisions by the control group due to a lack of ex ante information. However, the government was concerned about letting future beneficiaries know that they will receive the benefits only until four years after the implementation then, only those farmers who benefit in the first year will know in advance and annual randomizations will determine when other producers will obtain the funding.

Another aspect that needs to be considered when implementing a randomized selection of beneficiaries is the social validity of the process. Usually, governments are concerned about producers' perceptions regarding the transparency of the selection process. Hence, in order to validate the selection process, the program will be broadly advertised at the national and regional levels using mass media as well as other mechanisms to inform the farmers about the second phase of PATCA. Additionally, farmers and community leaders will be invited to participate in the randomized process of selection in order to guarantee the transparency of the process.

7.3.2 Econometric Analysis

As with every impact evaluation the most important challenge is to identify an appropriate counterfactual that allows us to infer what would have been the outcome for the treatment group with and without the program. However, given that PATCA II will implement a randomized selection of beneficiaries, the identification of the program impact is considerably simpler than without a randomized selection.

As mentioned in Section 4, a simple experimental design implies that producers will be assigned to control and treatment group randomly prior to program implementation. Once producers are assigned to each group, all the producers in the treatment group will receive the program and producers in the control group will not receive the program. An experiment implies that, without the program, the outcome variable should have been the same, on average, for control and treatment group. The program impact could be estimated (at least initially) following equation (1) by finding the first difference:

$$Y_i = \beta_0 + \beta_1 * PATCA_i + \varepsilon_i \quad (9)$$

This coefficient (β_1) would represent the true causal impact of the program. However, to implement this type of experimental methodology would require that all producers in the treatment group receive the program and all the producers in the control group do not receive the program. It is desirable to verify that ex ante characteristics between control and treatment are the same by administering a baseline survey or using second hand information such as a census.

Even with the randomization to be confident of the results, other actions will be taken for a number of reasons. First, as noted in section 4 since baseline data will be collected running the model as a double difference to check whether the trends in the treatment and control groups differ can help to address any remaining differences in the treatment and control that were not adequately dealt with in the randomization. The results should remain the same but this provides a robustness check.

Second, it is very likely that the selection process will not result in a simple experiment as there might be some producers in the treatment group who will not receive the program. In fact, once the farmer is selected into the treatment group the field agents need to corroborate that all requirements are fulfilled and that land characteristics are appropriate in order to implement the technology. Besides, to benefit from PATCA, the producer needs to finance 50% of the technology and there is no guarantee that all the selected farmers will be able to pay even if an ex-ante economic analysis is performed for each producer. This issue is widely known in the literature as non-compliance.

The instrumental variables technique noted in Section 4 allows us to deal with the issue of non-compliance. In this particular case, the most appropriate instrument is the randomization itself as it is correlated with program participation but not correlated with the outcome. In other words, the instrument will be a dummy variable (Z_i) equal to 1 if the farmer is selected into treatment group and 0 if the farmer is not selected. Then, program impact is estimated using two stage least squares. In the first stage, the participation in PATCA is estimated as follows:

$$PATCA_i = \delta_0 + \delta_1 Z_i + \mu_i \quad (10)$$

In the second stage, the predicted probabilities are used to estimate the impact of PATCA on different outcomes as follows:

$$Y_i = \alpha_0 + \alpha_1 PATCA_i + \varepsilon_i \quad (11)$$

This type of estimation will allow us to identify the program's impact while dealing with unobserved heterogeneity.

Note that the final empirical models will be run to assess the three impacts noted above, including the direct impact, the indirect impact on non-participant eligible farmers and the indirect impact on non-participant ineligible farmers.

7.3.3 Sampling Strategy

In terms of the sampling strategy for the case of PATCA II, the main issue is to identify direct and indirect effects therefore, as with the econometric methodology; the sampling strategy is closely related to program design. As mentioned, the selection strategy will classify the farmers into five excludable groups. Hence, the sampling strategy needs to guarantee enough number of producers for each group to be able to identify program direct and indirect effects as well as heterogeneous impacts among different groups of farmers. In other words, the sample needs to be stratified. To accomplish this, the sample should be representative for each group: (i) beneficiaries; (ii) non-beneficiaries eligible in selected areas; (iii) non-beneficiaries ineligible in selected areas; (iv) non-beneficiaries eligible in non-selected areas and (v) non-beneficiaries ineligible in non-selected areas.

To reduce sample dispersion, the sample will be drawn in two stages. In the first stage, the primary sampling units (PSU) will be randomly selected. In this case, the primary sampling unit is the geographic area at which the spillover is expected to take place. In the second stage, the producers will be randomly selected from the sample of already selected areas. The two stage sampling is standard practice because it facilitates data collection and reduces administrative costs. However, as noted in section 5.3 it requires a sample adjustment caused by the intra-correlation within clusters (World Bank, 2007).

Generally, it is recommended to use the most recent Agricultural Census as the basis to develop a sample frame. However, the Dominican Republic does not have this information available and therefore; it is necessary to develop in the field a listing of all producers located in the selected areas. This list of producers will be used as a sample frame.

As mentioned in Section 5.3, four pieces of information are required to calculate a preliminary sample size: (i) expected effect on outcome variable, (ii) outcome variable standard deviation, (iii) confidence level and (iv) statistical power. To calculate the first two pieces of information it is required to identify the main impact indicators. For PATCA II, the main impact indicator is agricultural productivity and it is the outcome variable that will be used to calculate the sample size. The third and fourth pieces of information refer to the sample capability to reduce miscalculation of program impact. The critical values associated with a normal distributed variable are 1.28 and 1.65 respectively. These four pieces allow us to calculate a preliminary sample size using the following equation (7) in Section 5.3.

Given the lack of information available regarding the outcome variable for the universe of agricultural producers in the Dominican Republic, second-hand information is used to calculate the values of the expected effect on outcome variable and the outcome variable standard deviation Specifically, values are used that are obtained from a study conducted by González et al. (2009) who calculate the value of production per hectare of cultivated land to evaluate the impact of PATCA I. Different calculations using the authors' values suggest a standard deviation of US$4,011. If an average impact of 10% in the value of production is assumed, the difference between the average value of production between both groups would be equal to US$1400. Replacing these values in equation (7) the following sample size is obtained:

$$N = \frac{4(4,011)^2(1.96 + 1.28)^2}{1400^2} \approx 345$$

Assuming an intra-cluster correlation equal to 0.05 and about 16 producers per PSU, following equation (8) the corrected sample is:

$$N_{corregido} = N[1 + \rho(m - 1)] = 345[1 + 0.05(16 - 1)] \approx 600$$

The preliminary results suggest a representative sample for each of the five groups equal to 600 and a total sample of about 3,000 producers. Additionally, it is important to consider the possibility of future attrition and non-response. Hence, assuming a 20% rate of non-response, the total sample would be somewhere between 3,500 to 4,000 producers that should be equally divided into five groups. This would imply about 800 producers per group.

7.4 Data Collection

7.4.1 Timing and Periodicity

For the PATCA II, because the main objective of the survey is to collect detailed agricultural data, the best timing to collect the survey is right after the main crop harvest takes place, which is March. This reduces measurement error because the farmers' recall period is shorter and producers are more aware of the amount of inputs used and the total output obtained. Besides, this avoids the possibility of producers estimating future yields and sales which could be rather inaccurate.

With respect to the periodicity, the data collection strategy for the PATCA II impact evaluation requires two rounds of survey data. This will allow us to have a panel data to estimate double differences in order to control for time invariant unobservable characteristics, time tendencies and initial differences. Also, having baseline data will allow us to check the validity of the randomization strategy. The first round of data will then be collected in 2011 in the month of March. The second round will be administered during the same month in 2014. The information collected will correspond to the agricultural years of 2010 and 2013 respectively.

7.4.2 Questionnaire

The questionnaire to collect first-hand data to evaluate a program should be carefully designed to gather all the information needed to test the aforementioned hypotheses, construct the impact indicators and any other variable needed as control. The survey instrument to be administered for the impact evaluation of PATCA II is composed by more than 300 questions, divided in eleven modules built upon the *LSMS-Integrated Surveys on Agriculture* from the World Bank (World Bank, 2010). Specifically, the modules collect information regarding: household demographic characteristics, parcel characteristics, detailed agricultural information on inputs and production, livestock production, commercialization, economic activities, migration, credit and savings, other income, social capital and dwelling conditions. The questions included are carefully thought to produce the information required to assess the hypotheses presented in section 7.1 (Table 4) as well as to create other control variables needed for the econometric analysis. Also, questionnaires administer to community leaders and technology providers will complement information regarding prices, access to inputs, access to infrastructure, etc.

7.5 Impact Evaluation Budget, Products and Activities

7.5.1 Budget

Table 6 presents the preliminary budget that has been estimated for the implementation of the PATCA II impact evaluation. The numbers obtained to construct the budget were obtained from information provided by firms that collected previous agricultural surveys in the Dominican Republic. Overall, the total cost of the impact evaluation is estimated in US$1,182,000. The impact evaluation cost will be financed with four different sources: local government, IDB loan, IDB resources and a technical cooperation from the IDB. Most of the resources will be allocated to finance the data collection. Specifically, US$575,000 will be allocated to administer the two surveys to agricultural producers (US$280,000 per round).

7.5.2 Products

In addition to budget specification, it is useful to identify the main products that will be obtained with the impact evaluation. Usually, products generated for this purpose can also be used to conduct other types of studies or other impact evaluations. Therefore, it could be useful for government officials and international organizations to communicate the different products that will be obtained in order to identify future economies of scale associated with the availability of information. The products of the PATCA II impact evaluation can be categorized into intermediate and final. The products include a preliminary analysis using the baseline data and the final impact evaluation analysis using both rounds of survey data. The documents will be finalized by August 2012 and December 2014 respectively. The intermediate products include the questionnaires to producers, providers and community leaders as well as the sampling methodology report, pilot trials report, the field work report and the data sets.

7.5.3 Timeline and Activities

One of the most important aspects regarding an impact evaluation is to create a timeline for the different activities. Table 7 presents the PATCA II impact evaluation timeline. The first column describes the activities followed by the main responsible actors and the dates when the activity will take place. In general, the main responsible actors are the SPD/SDV and INE/RND divisions from IDB, the executing unit (PATCA), the field supervisor and the firm responsible for collecting the data. The IDB will be in charge of the scientific design along with the questionnaires design and the scientific supervision. Also, it will provide support on the data analysis and data collection. The executing unit will be mainly responsible of the monitoring aspect and will provide assistance in the data collection and field supervision. Finally, the field supervisor will manage the field work that includes the pilot trials, the training activities and the survey administration along with the consultant firm.

Table 7. Timeline for PATCA II Impact Evaluation Activities

ACTIVITIES	RESPONSIBLE	DATES
Baseline Administration		
Questionnaire Design	BID (SPD; RND)	Sept-Oct 2010
Pilot Trials (1st round)	Consultant firm to perform field work and field supervisor.	Jan 2011
Pilot Trials (2nd round)	Consultant firm to perform field work and field supervisor.	Feb 2011
Sampling Design and Methodology	Consultant/BID (SPD; RND)	Feb 2011
Training for interviewers and supervisors. Preparation of materials for survey administration.	Consultant firm to perform field work and field supervisor.	Jan-Feb 2011
Baseline Administration	Consultant firm to perform field work and field supervisor.	March 2011
Data Entry	Consultant firm to perform field work and field supervisor.	April-May 2011
Baseline Data	Consultant firm to perform field work and field supervisor.	June 2011
Baseline Analysis	Consultant/BID (SPD; RND)	July-Dec 2011
Follow-Up Survey		
Questionnaire Design	BID (SPD; RND)	Sept-Oct 2013
Pilot Trials (1st round)	Consultant firm to perform field work and field supervisor.	Jan 2014
Pilot Trials (2nd round)	Consultant firm to perform field work and field supervisor.	Feb 2014
Training for interviewers and supervisors. Preparation of materials for survey administration.	Consultant firm to perform field work and field supervisor.	Jan-Feb 2014
Follow-up Survey Administration	Consultant firm to perform field work and field supervisor.	March 2014
Data Entry	Consultant firm to perform field work and field supervisor.	April-May 2014
Follow-up Data	Consultant firm to perform field work and field supervisor.	June 2014
Impact Evaluation Analysis for PATCA II	Consultant/BID (SPD; RND)	July-Dec 2014

8. Case Study: Nicaraguan APAGRO Program[13]

This section provides an overview of the evaluation design of the APAGRO (*Programa de Apoyos Productivos Agroalimentarios* loosely translated as Agro-food Support Program), an IDB-funded program, which initiated implementation at the end of 2009 in Nicaragua. The specific objective of this chapter is to use APAGRO as a case study to derive some lessons from designing the impact evaluation procedures to be followed during its implementation.[14]

8.1 The APAGRO Intervention

The *Programa Productivo Alimentario* or PPA seeks to enhance the capitalization of poor farmers so that they can improve their living standards and eventually move towards a more productive and competitive position. The PPA consists of a set of interventions designed to increase the productivity and income of poor farmers throughout Nicaragua with funding from different sources including the national treasury and various international donors. APAGRO, which falls within the broader PPA, consists of two components: i) support to rural families, through the Agro-food Productivity Voucher (*Bono Productivo Alimentario* or BPA), in the adoption of technological improvements to increase food production and farm productivity, and ii) actions designed to improve the managerial capacity of beneficiaries.

APAGRO's targets poor farm families that rely primarily on the production of basic grains for home consumption. The BPA provides funding so that farmers can acquire productive assets to diversify their farm output (e.g., dairy cows, small animals, fruit production, improved vegetative material) and supports the provision of technical assistance so that these assets can be used efficiently. Given the high levels of poverty and food shortages that face many inhabitants of this region, the BPA is an important instrument for promoting an increase in the sustainable production of food, dietary improvements and farm productivity gains among low income rural families.

Under Component I, the establishment costs of eligible technologies is financed through the BPA with a projected aggregate ceiling per farmer of US$1,400, or its equivalent in local

[13] This section was written by Boris E. Bravo-Ureta Professor of Agricultural and Resource Economics at the University of Connecticut in Storrs, CT, USA, and Adjunct Professor of Agricultural Economics at the University of Talca in Talca, Chile.

[14] The discussion concerning APAGRO draws from IDB (2008) and from Bravo-Ureta (2009).

currency, during the life of the program. It is expected that Component I will support at least 11,000 producers, primarily female headed households.

Component II provides financial support to purchase specialized goods and services associated with three activities: 1) Empowering beneficiaries with basic financial education dealing with small businesses provided by the Rural Credit Fund (RCF); 2) Raising awareness and providing information on the BPA so that input and service providers become aware of their potential business opportunities with small producers; and 3) Strengthening the development of associative groups of eligible beneficiaries. The expected results from Component II are: 1) At least 11,000 beneficiaries are provided with basic financial education; and 2) At least 20 new associative projects are successfully carried out.

The technologies considered are incorporated in a Menu of Eligible Options (MEO) and have been selected in accordance with the following technical requirements: (i) contribution to agriculture and food production; (ii) positive impact in the medium term in relation to net income of beneficiary families measured by and Internal rate of Return (IRR) equal to or greater than 12%; (iii) applicability to the ecological conditions where their use is promoted; (iv) neutral or positive environmental impact; (v) ease in monitoring the adoption by beneficiaries; (vi) inputs associated with each technology must meet explicit technical standards established by APAGRO; and (vii) agrochemicals banned in Nicaragua are not funded.

The funding provided by APAGRO is to cover the associated costs of inputs, labor, transport and technical assistance. The cost of each technology is to be reviewed annually and adjustments can be made with consent from the IDB and MAGFOR. A total of 10 specific technologies, shown in Table 8, were analyzed during project preparation. It should be noted that the technological package that a farmer can choose is not fully defined *a priori* and the beneficiary has the option of combining various ingredients from the MEO. In addition, other options were identified that are to be analyzed at the beginning of APAGRO's implementation and include wood utilization, and soil and water management technologies (see Sain, 2008 for details on the technologies).

Table 8. Technologies Included in the Menu of Eligible Options (MEO) for APAGRO

1. Milk production from dairy cows
2. Milk production from goats
3. Production of pork meat
4. Production of eggs and poultry meat
5. Improvements in the productivity of white maize
6. Improvements in the productivity of red beans
7. Improvements in the productivity of vegetables for home consumption
8. Introduction and improvements in the productivity of fruits for home consumption
9. Handling of organic fertilizers
10. Grain handling and storage with metal silos (12QQ)

To be eligible, producers need to fulfill the following conditions: (i) have good faith possession of at least one but no more than 10 manzanas[15] of land, which cannot be located in core protected areas; (ii) have not received the BPA nor have been previous APAGRO beneficiaries; (iii) do not have the production technology of larger livestock species (cattle); and (iv) the fulfillment of these criteria must be confirmed by a signed declaration of the household member that would be registering for the program as a beneficiary. In addition, there can only be one beneficiary per household and women heads of household which fulfill the land possession clause are eligible.

8.2 The Logic of the APAGRO Intervention

It has long been recognized that poor farmers face a number of obstacles that limit and even preclude their ability to adopt production technologies even when the expected profitability of such technologies is high (Feder and Umali, 1993; Feder, Just and Zilberman, 1985). One of these obstacles is that peasant farmers are risk averse and this type of behavior hinders the adoption and diffusion of technologies that could increase the production and income of poor producers (Lee, 2005; Ellis 1988; Hiebert, 1974; de Janvry, 1972).

Another significant barrier that peasant farmers face concerning technology adoption is the inability to secure adequate credit at reasonable cost. Conning and Udry (2007) argue that "…agents in the rural sectors of most developing countries remain cut-off from many

[15] 1 manzana = 0.7 hectares

opportunities for investing, risk-taking and risk spreading that would be available through better financial integration…" (p. 2859). These authors also indicate that there is considerable survey based evidence showing that farmers in developing economies would be willing to borrow more if additional credit was made available at a given interest rate which suggests that credit rationing is indeed an issue. From the lenders perspective, peasant farmers can have high risk of default and any loans that would be made to them would have high costs and thus generate low returns which makes this clientele group not very desirable (Foster and Rosenzweig, 2010). In addition, there are different sources of asymmetric information which further limit the ability of poor producers to qualify or procure credit (Boucher, Carter and Guirkinger, 2008). Moreover, evidence from Guatemala shows that financial constrains limit small farm adoption of technologies that would make it possible to participate in lucrative export markets which can have adverse consequences on competitiveness and ultimately could lead to further concentration in landownership (Barham, Carter and Sigelko, 1995).

A third hurdle that can get on the way of poor farm families interested in adopting desirable innovations concerns lack of information. Information has been found to play a key role in improving agricultural productivity and managerial skills which in turn enhance the awareness of farmers towards technology choice (Solis, Bravo-Ureta and Quiroga, 2009; Anderson and Feder, 2007; Lee, 2005). In a recent survey of microeconomic studies that analyze the technology adoption process, Foster and Rosenzweig (2010) conclude that "…education plays an important role in facilitating the acquisition and processing of new information, which appears to account for the pervasive finding that more educated agents adopt new technologies first…" (p. 421). Coupling this latter conclusion with the Cochrane treadmill effect raises the important concern that early adopters are the only farmers that truly benefit from technological innovations which in turn are likely to be the larger farmers who are less risk averse, and have easier access to credit and information (Sunding and Zilberman, 2001).

The logic underlying APAGRO recognizes the constraints that poor farmers face when contemplating the adoption of beneficial technologies and thus the Program is designed to promote technological improvements through the acquisition of productive assets (e.g., dairy cows, small animals, improved vegetative material) and by supporting technical assistance so that the assets can be used efficiently. APAGRO contributes to the reduction of risk aversion by providing farmers with a menu of proven technologies that are expected to enhance both household nutrition and farm income. In addition, risk aversion and the credit constraint are addressed since the program gives grants to beneficiaries so that they can pay for the assets that are being acquired. The information constraint is also relaxed by making available technical assistance and data on pre-qualified input providers which also contributes to relaxing risk aversion. Moreover, extension services in many developing countries are not well supported although empirical evidence indicates that inefficiency among farmers in developing countries can be significant and that access to extension advice can improve farm performance (Bravo-Ureta et al, 2007; Bravo-Ureta and Pinheiro, 1993). Thus, development projects like APAGRO can also contribute to bridging the gaps left by the weakness of public extension programs.

In sum, APAGRO, and related programs in Nicaragua and other countries, by focusing on serving the poorest farm households, has a significant role to play in addressing a number of market failures while attempting to alleviate extreme poverty and fostering economic development. These interventions can have major societal benefits since they provide poor farmers services and support that have public good characteristics which constitutes a key rationale for government funding (Birkhaeuser, Evenson Feder, 1991; Stiglitz, 1987).

8.2.1 Hypotheses to Test and Indicators

Table 9 contains some of the most salient features of the Results Matrix[16] that was included in the program preparation. It shows that the goal of APAGRO is to contribute to increasing the income of poor farmers and the purpose is to have at least 80% of Year 1 and Year 2 beneficiaries obtain an internal rate of return equal to or greater than 12% attributable to the technologies funded by the program. This is a somewhat unusual indicator of impact, but clearly its measurement is based on changes in the net income of beneficiaries stemming from the technologies supported. Moreover, it is suggested here that the difference between the total value of farm production minus purchased inputs, or what can be called Returns over Purchased Inputs (ROPI), should be used as the indicator of impact and then all the needed information would be available to calculate the IRR.[17]

---

[16] The Results Matrix is an instrument designed to strengthen the achievement of expected results and to assess and report on the performance of individual projects (ProVention Consortium, 2007). The Results Matrix contains the key aspects of the project's roll-out as well as all important indicators to measure the different dimensions of the program presented in a clear but concise fashion. In recent practice at the IDB, the Results Matrix has replaced the Logical Framework.

[17] This implies that the ROPI represents a payment to farm-owned resources including land and family labor.

Table 9. Program Roll-out and Data Collection Plan for APAGRO, Nicaragua

Goal: Contribute to improving the income of poor farm families in Nicaragua.		**Purpose**: At least 80% of Yr 1 & 2 Beneficiaries, based on a representative sample, obtain an IRR =>12%					
COMPONENT 1. Provide Support through the BPA to low-income farm families to introduce technologies to improve their productivity and production.							
Result Indicators	**Baseline**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**TOTAL**
1. # of beneficiaries by gender that have received the full BPA	0	3500	4000	3500			11000
COMPONENT 2. Improving the managerial skills of rural families benefiting from BPA							
Result Indicators	**Baseline**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**TOTAL**
1. # of beneficiaries that received training in financial principles	0	2500	3000	3000	2500		11000
2. # of rural businesses initiated and operating	0	0	5	5	5	5	11000
Program Roll-Out and Evaluation Scheme	**Year 0**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**TOTAL**
COHORT-1							
Interviewed Year 0/Base Line; Year 1, 3 and 5	**C1**			C1		C1	
Beneficiary: Years 1, 2, 3, 4 and 5		3500	3500	3500	3500	3500	**3500**
COHORT-2							
Interviewed Year 0/Base Line; Year 1, 3 and 5	**C2**		0	C2		C2	
Beneficiary: Years 2, 3, 4 and 5			4000	4000	4000	4000	**4000**
COHORT-3							
Interviewed Year 0/Base Line; Year 1, 3 and 5	**C3**		0	C3		C3	
Beneficiary: Years 3, 4 and 5				3500	3500	3500	**3500**
TOTAL INTERVIEWED*	**C1-C2-C3**			**C1-C2-C3**		**C1-C2-C3**	
TOTAL BENEFICIARIES		**3500**	**7500**	**11000**	**11000**	**11000**	**11000**

It is important to indicate that the interest in conducting this evaluation is not just to determine whether the program had an overall impact, but also to consider the mechanisms by which that impact occurred. This analysis allows for a greater understanding of how the program induced changes in farmer behavior and the role the new technologies played in improving farmer welfare. The calculation of the IRR includes a number of intermediate measures that can be analyzed separately. These include the degree to which technologies are adopted, changes in input use, changes in the use of family and hired labor, increases in yields, and differences in output prices, among other indicators. The expectation is that farmers that adopt the technology also increase the use of purchased inputs and labor and this leads to greater production and potentially to higher output prices if product quality is increased. Therefore, the evaluation will test these hypotheses along with the central hypothesis that the technologies promoted by APAGRO and adopted by beneficiaries lead to higher farm income and to an IRR greater than 12%.

8.3 Program Roll Out

The Nicaraguan Ministry of Agriculture and Forestry (MAGFOR) selected the municipalities that will be intervened during the implementation of APAGRO based on the following criteria: (i) highest share of rural poverty; (ii) absolute size of target population; (iii) more isolated municipalities and/or those more prone to drought; and (iv) number of beneficiaries already receiving the BPA. Following these priorities, four municipalities were selected in the Department of Jinotega—Jinotega, Cua-Bocay, Wiwilí and Valí—and an additional seven municipalities were selected in the Department of Matagalpa—Matagalpa, Rancho Grande, La Dalia, Esquipulas, San Dionisio, Terrabona, and Ciudad Darío.

At the beginning of implementation, APAGRO will develop a complete list of eligible beneficiaries followed by an inspection to verify compliance with the established eligibility criteria. Treatment will be available to all eligible families that choose to participate and the implementation will incorporate everyone that is eligible in a given location all at once in a process called *barrido*. Considering that the target population is very poor, and that the program provides a grant for asset acquisition and technology adoption along with technical assistance, it is expected that a very large percentage of those eligible will indeed choose to participate.

The Program Operating Manual or *Reglamento Operativo del Programa* (ROP) indicates that, based on a representative sample of beneficiaries who have received support from the program, an independent audit will be made to verify if the support is delivered effectively to the beneficiaries and whether the acquired technologies meet APAGRO's technical standards. This verification process will be incorporated into a management, monitoring and evaluation system that must be developed by APAGRO. The ROP also indicates that baseline[18] data is to be

[18] The baseline study is an analysis -OECD/DAC, 2002- describing the situation prior to a development intervention, against which progress can be assessed or comparisons made.

collected to serve as the foundation for a study to establish the starting point of the indicators included in the Results Matrix of the program. This study will also serve as a reference point for the final performance report and for any possible ex-post evaluation conducted by the IDB and/or the Government of Nicaragua. Thus, the basic structure to undertake an adequate impact evaluation at the end of the program has been established from the beginning, which is a necessary condition to perform such work.

The ROP (MAGFOR, 2009) states that communities served during the first year of the program are chosen randomly through a lottery and the number of beneficiaries per year cannot exceed the number of bonds allocated for that year and location. Communities that are selected for treatment in year 1 but that do not get treated will receive priority in the second year. In case the number of eligible households is higher than planned, which is not anticipated to be the case, those enrolled will be again chosen through a lottery. Although the initial implementation of the program was planned over a five year period, the ROP does not include the enrollment rules beyond year 1, and this is an issue that is addressed below.

The original intention of the program was to proceed with a relatively slow implementation schedule during which different components of the technological packages would be introduced gradually to ensure full and sustainable adoption. Consistent with this intention and in the spirit of compromise with some officials that were interested in 100% implementation in Year 1, the program roll-out proposed is divided into three cohorts (see Table 9):

- Cohort 1: 3,500 beneficiaries which begin treatment in Year 1
- Cohort 2: 4,000 beneficiaries which begin treatment in Year 2
- Cohort 3: 3,500 beneficiaries which begin treatment in Year 3

This schedule allows sufficient time to incorporate all the technologies included in the program and to provide adequate technical support, after integrating the third cohort during Year 3. One must take into account that some of the technologies may have a fairly long implementation period. Of particular interest on this regard are dairy cows, of great importance in the program, which have an 18 month average cycle. Therefore, the schedule proposed in Table 9 will allow the last beneficiaries of the program to complete this cycle prior to the close out of APAGRO.

To make the selection of beneficiaries in accordance to the ROP, while at the same time allowing for the evaluation of the program as stipulated in various documents, the selection will begin at the *comarca* level, which is an official geographical grouping of adjacent communities. This would facilitate data collection logistics and thus reduce associated costs. Therefore, early on the planning process a geographically referenced list of all *comarcas* within APAGRO's intervention area is needed for each of the 11 municipalities already selected (four in Jinotega Department and seven in Matagalpa Department).

The *comarcas* can be organized into 6 groups for each municipality, according to their geographical proximity to facilitate the eventual *barrido* process. This will yield a total of 66 groups of *comarcas* and each group gets a distinct label (e.g., Jinotega1, ..., Jinotega6; San Dionisio1….; etc.). Then, a random selection is made where the first label chosen gets the number 1 and so on until reaching the last group which gets the number 66. The order of incorporation to the project, that is to say the sequence of the *barrido*, will follow the selected number, from lowest to highest. The aim is to get three groups of *comarcas*, one for each of the three cohorts already defined.[19]

Now that the order of treatment has been determined, the list of all eligible farmers in each *comarca* needs to be obtained, which will come from the lists that will be prepared by the executing unit in the eligibility verification process mentioned earlier. This will make it possible to have a count of all eligible families by *comarca* and group of *comarcas,* which in turn makes it possible to make a final determination concerning the number of *comarcas* needed to arrive at the total to be treated in each cohort. With this information the sample size can be determined, i.e., the total number of farms to be surveyed, which will then be allocated proportionally to the stipulated size of each cohort. These calculations will follow standard procedures but should be done in consultation with a statistician with experience on sampling design (World Bank, 2007; Bartlett, Kotrlik and Higgins, 2001).

In summary, to select the individual families to be sampled and knowing which of the 66 groups fall in each cohort, a sample of *comarca* groups will be selected randomly and then for each selected group, individual households will be randomly selected. In this fashion, and as

[19] It appears that the promotion of the program has already started and the municipality of Jinotega and Matagalpa have been targeted by the government for early treatment. In this case, it seems reasonable that the groups of *comarcas* for these two municipalities are incorporated randomly, half in Year 1 and the other half in Year 2. For this to be accomplished, the random order obtained initially would have to be modified, and this reallocation should also be made randomly.

alluded earlier, the clustering of adjacent *comarcas* from where individual farmers are selected should reduce costs of the field work while generating a reasonable distribution of households to be surveyed. Again, the executing unit needs to seek out the advice of a qualified statistician in order to set up the specific procedures to be used in sample selection.

8.4 Evaluation Methodology

To assess the project's impact, it is necessary to identify a control group to serve as the counterfactual situation, which must have the same characteristics on average as the group that will be studied. It is common in agricultural development projects to use quasi-experimental methods where the counterfactual group is constructed using *Propensity Score Matching* approach (e.g., Cerdán-Infantes, Maffioli and Ubfal, 2008; Bravo-Ureta et al., 2011; Cavatassi et al., 2010; Lopez and Maffioli, 2008; Rodriguez, Rejesus and Aragon, 2007). However, randomized experiments are gaining considerable support as a preferred option (e.g., Angrist and Pischke, 2009; Duflo, Glennerster and Kremer, 2008; Ravallion, 2008).

The proposed experimental procedure, which seeks to generate the counterfactual group for APAGRO, is called Randomized Order of Phase-in (Duflo, Glennerster and Kremer, 2008). This procedure allows for an experimental approach to be applied which is particularly appropriate in situations where all eligible units (e.g., individuals, households, farms) will be treated, but not at the same time, and incorporation to the program overtime is randomly assigned, which coincides with the process contemplated for APAGRO. One of the advantages of this procedure is that randomization is a fair and transparent way of prioritizing the incorporation of those eligible since it is not possible, and most likely not desirable, to treat all eligible units at once. An added advantage is that it potentially promotes cooperation from the control group in responding to surveys, since its members know that they will be treated within the life of the program, as is expected with APAGRO (Duflo, Glennerster and Kremer, 2008). On the other hand, there is a concern it may induce strategic behavior on the part of the control group either leading them not to adopt technologies as they wait for government support or to answer questions in ways they think will induce more support in the future. Another possible disadvantage of this procedure can arise in situations where the incorporation occurs very rapidly in relation to the time it takes the intervention to have its effects, making it difficult to detect, and therefore, evaluate the results of the treatment. However, within the proposed schedule for

APAGRO the difference between the incorporation of the first and last beneficiaries could be as much as 34 months, which should allow the effect of the intervention to be identified.

Several recent studies have used the Randomized Order of Phase-in including the work of Miguel and Kremer (2004) on a deworming program in groups of school children in Kenya, an analysis of the Social Safety Network Program (PRS) in Nicaragua (Maluccio and Flores, 2004), a study of PROGRESA in Mexico (Skoufias, 2005) and the Family Allowance Program (PRAF) in Honduras (Stecklov et al, 2007). The procedure used by Maluccio and Flores (2004) in their analysis of the PRS in Nicaragua is very similar to what is recommended in this report, and therefore, we rely on that study to illustrate how the methodology could be applied to evaluate the impact of APAGRO.

To illustrate the evaluation methodology to be used, let us define the indicator $\boldsymbol{I}$ (Intervention) for a sample of treated individuals and $\boldsymbol{C}$ (Control) for a sample of untreated farmers. Before the program, it is expected that both randomly chosen groups exhibit the same values for the indicator of interest (e.g., farm net returns or total value of farm production). Thus, the difference $\boldsymbol{I_0 - C_0}$ should be very close to zero, where $\boldsymbol{0}$ indicates the program's baseline year. However, after the program, we expect that any differences between the two groups can be attributed to the intervention, given that the timing of the treatment for each group was assigned at random (Angrist and Pischke, 2009). Then, the difference $\boldsymbol{I_1 - C_1}$ would be a valid measure of the average effect, which under this design is called a first difference. A more robust measure would take into account any preexisting difference, from either observable or unobservable variables, between the two randomly selected groups, and this would be a double difference (***DD***) or difference-in-differences as explained in section 4.1. The ***DD*** estimator is given by:

$$\boldsymbol{DD} = (I_1 - I_0) - (C_1 - C_0) = (I_1 - C_1) - (I_0 - C_0).$$

Given that we have three different cohorts in APAGRO, it is easier to visualize the measurement of the impact of the program by using the following equation: [20]

$$\boldsymbol{AP_{ict} = \alpha_0 + \beta_1 Y_3 + \beta_2 Y_5 + \beta_3 B_c + \gamma_1 Y_3 B_c + \gamma_2 Y_5 B_c + \varepsilon_{ict}} \quad (12)$$

[20] This is an adjustment to equation (2) for this particular case.

where:

AP_{ict}**:** APAGRO's indicator for individual I, in comarca group c in year t;

Y_3**:** binary variable equal to 1 for year 3 (year 0 is the excluded category);

Y_5**:** binary variable equal to 1 for year 5;

B_c**:** binary variable equal to 1 for households that are APAGRO beneficiaries in comarca group c;

ε_{ict}: error term; and

α, β, γ**:** parameters to be estimated.

Equation (2) is fitted econometrically in order to get estimates of the various parameters. The parameters of particular interest are γ_1 and γ_2 where γ_1 is the ***DD*** estimator of the average effect of the program for Year 3 with respect to Year 0, the baseline period, and γ_2 is the ***DD*** estimator for Year 5 also with respect to Year 0. In the actual estimation of these types of models it is customary to include household, farm and/or location variables in order to increase the precision of the estimated parameters (Maluccio and Flores, 2004).

A point, noted briefly above, that can be a matter of concern is the strategic behavior that future beneficiaries might exhibit as they await their turn to be integrated into the program and start receiving support. In the APAGRO environment, beneficiaries awaiting their turn are not in a position to adopt these technologies, given that they are very poor and the technologies are costly making it unlikely they will adopt without some support. It is not possible *a priori* to know if the expectation of receiving this support will change the behavior of future beneficiaries concerning adoption, but it is safe to argue that such change is highly unlikely given that these farmers do not have the resources to allocate and adopt these technologies without external support from a program such as APAGRO. In other words, those farmers not participating initially will hold out on adopting the featured technologies until they are incorporated into the program not because of strategic behavior but because they have little or no choice to do otherwise. And this of course is the logic of an APAGRO-type intervention as argued earlier.

Another source of concern has to do with spillover effects. Interventions have a target group, the intended direct beneficiaries or program participants, but often individuals not in the target group are also affected by the intervention and this unintended effect is called spillover (Angelucci and Di Maro, 2010). In order to fully account for the impact of a project such spillovers need to be incorporated in the analysis. In APAGRO, spillover effects are not expected to be significant since all eligible farmers in the selected regions will be incorporated and the participation of eligible households is expected to be very high. The likelihood of spillover effects in neighboring communities should not be a problem either for several reasons: the *barrido* approach implies that all eligible individuals in the area of intervention will be treated if they choose to participate and participation should be near 100%; communities in APAGRO's area of influence tend to be fairly isolated; non-participants do not have the resources needed to acquire the assets; and the extension support required to fully implement the technologies is not available outside of the program.

Before concluding this section it is necessary to indicate that by the end of the implementation of APAGRO in Year 5, beneficiaries that adopt technologies with longer gestation periods, such as fruit trees and dairy cows, will not have received significant revenue streams. Since the IRR is an indicator of interest in this program, it will be necessary to examine other intermediate indicators noted above and make projections of the expected revenue stream to be realized during the five to ten years following the closing of the project. These projected flows can be discounted to Year 5 in order to have an estimate of potential future return on these technologies. The present value of these flows in Year 5 could also be seen as the residual value which would then allow the calculation of the IRR without punishing the longer-term technologies.

Despite the fact that this technical adjustment is possible and is often done in the evaluation of projects, it should be asked if the adoption of technologies with expected benefits that take several years to be realized (e.g., trees for fruit and/or wood production) should be part of a menu offered to low income, risk averse and financially constrained households. The merits of including technologies that take several years to generate positive net income flows in programs that are primarily oriented to the goal of poverty alleviation is a broader issue that merits further analysis and discussion. An additional complication is that tree-related technologies require assistance (technical and marketing) for a period that goes beyond the life of the project which most likely will not be available and this could present major challenges to low income households. Technologies incorporating trees might be very desirable from an environmental perspective, particularly in hilly terrains, which are predominant in much of Central America, but that kind of intervention needs a different implementation scheme than the one contemplated for APAGRO.

8.5 Data Collection Strategy

8.5.1 Timing and Survey Organization

The random assignment of incorporation into APAGRO by group of *comarcas* has already been explained. The proposed plan is to conduct a baseline survey at the beginning of Year 1 for a random sample of all three cohorts. Treatment of the 1st cohort begins in Year 1, of the 2nd Cohort in Year 2 and of the 3rd Cohort in Year 3. At the beginning of Year 4, a new survey is undertaken to collect information for agricultural Year 3 for everyone surveyed in the first round. All data collected at this point constitutes the basis for the midterm evaluation, which should be conducted as soon as Year 3 data are ready for analysis.

At the end of the implementation of the program (beginning of year 6) a new survey is conducted for the samples of all three cohorts in order to collect data for agricultural year 5. At this point a panel data set covering three years will be available for the final evaluation of the program. Table 9 details the sequence of data collection for all three cohorts. It is important to stress that untreated farmers that eventually will be treated serve as control groups and whenever a survey is conducted, the data collected corresponds to the preceding agricultural year. Therefore, the best time for fieldwork in Nicaragua is the month of May for each round of surveys.

A final consideration has to do with how to undertake the data collection. In general, there are two modalities: (i) to conduct the data collection and analysis using personnel from within the executing unit; or (ii) to outsource all aspects of the work. A major advantage of the first option is the continuity that can be given to the work considering that it needs to be performed in several steps over a number of years spanning the design of the monitoring and evaluation system, the baseline study, ending with the final report containing the impact evaluation per se with various other steps along the way. However, using in-house personnel to conduct the evaluations is subject to criticism because it lacks the necessary impartiality and distance from the daily details in order to assure all stakeholders that the work is being done in an unbiased fashion. Option 2 avoids the bias issue, but makes the continuity of the work a challenge since the diverse phases typically need to be put out to bid separately and most likely there will be an array of individuals and consulting firms that carry out different pieces of the work.

A third approach, and the one recommended here, is to combine options 1 and 2. The executing unit would assign an internal person to coordinate and manage the data collection process which would be done by project staff perhaps with the help of outside enumerators. In addition, a small team of consultants would be hired to supervise all phases of the work and to conduct the analysis of the information and the preparation of the various reports. In this fashion, continuity and familiarity with the ins and outs of the program can be insured while the issue of potential bias is mitigated. There are several examples where this type of approach has been used which can serve as guidelines for APAGRO.

8.5.2 Questionnaire

As discussed it is important to recognize at the outset that the survey instrument required to collect the baseline and subsequent data needs to be constructed with great care. Due to the complexity of family farming operations, the development and implementation of the survey will entail a substantial effort. The questions that need to be included are not only directly related with the project indicators, but also with variables that are required to incorporate as controls in the various econometric models and analyses that need to be performed.

The firm/consultant that will be hired to collect the baseline data will have the responsibility of developing a detailed questionnaire. To do so, it is essential to review the Results Matrix and all relevant project documentation to insure that every pertinent aspect of the project is incorporated into the questionnaire. In addition, the questionnaire needs to be developed with the participation of APAGRO's executing unit and coordinated with the IDB project team.

The questionnaire needs to be broken down into several modules including the following:

a) Demographic information of the household (e.g., age, gender, schooling, relationship with household head, etc.);

b) Geographical position of the farm and a graphical sketch of the distribution of the various plots and location of the farm house and additional structures;

c) Assets owned/controlled by the household (e.g., land, machinery, draft animals, major tools, etc);

d) Quantity of inputs used (e.g., family labor, hired labor, purchased inputs, etc.) and of farm outputs produced;

e) Expenditures on hired labor, hired machinery services and draft power, and purchased inputs (e.g., fertilizer, seeds, pesticides, etc);

f) Quantity, price, value and disposition of farm outputs;

g) Environmental practices used in the farm (e.g., soil conservation practices);

h) On farm value addition activities;

i) Type of support received from APAGRO;

j) Support received from other projects and institutions;

k) Participation in extension activities sponsored by APAGRO and by other agencies;

l) Access to and use of credit;

m) Social capital and off-farm physical infrastructure in the area where the farm is located (e.g., participation in farm groups, distance to major town, type of access road, access/distance to health care facilities and schools, distance to market, availability of public transportation, etc); and

n) Other sources of income including off-farm activities.

It is important to reiterate that in each round of the survey the data to be collected needs to be for the previous agricultural year. Therefore, the best time to carry out this work is during May, which is right after the end of the agricultural year in Nicaragua and thus farmers are in the best position to have good recollection of the information and hence provide accurate responses.

8.6 Budget and Implementation

As noted, one activity that needs to be completed before the baseline data is collected, is the determination of the sample size, which should be done using the methodologies discussed elsewhere in this publication. Clearly, the number of households interviewed will have a direct incidence on the budget required to undertake the evaluation.

Drawing from two recent IDB activities in Nicaragua, which are closely related to APAGRO (the ex-post Evaluation of POSAF II and the ex-ante evaluation of *Programa Ambiental de Gestión de Riesgos de Desastres y Cambio Climático*), the expectation is that the average cost of a questionnaire is in the neighborhood of US$75 per survey in each round. It is important to keep in mind that three rounds of surveys are anticipated (Baseline, Mid-term

Evaluation and Final Evaluation) and in each round the same households will be interviewed in order to generate a panel data set. Thus, the cost per household for the three rounds is expected to be around US$225.

The products that need to be produced by the firm implementing the survey work for the baseline includes: i) elaboration of the survey form; ii) sample size determinations and selection of individuals households to be interviewed; iii) training of enumerators; iv) field testing of the survey; v) preparation of the final survey form incorporating the feedback received during field tests; vi) full implementation of the survey; vii) data coding and cleaning; and viii) the generation of a final electronic data set.

The Mid-term and Final Evaluations have a similar set of products with the exception of the first two items. Nevertheless, in each of these last two rounds it is necessary to review the list of farmers in the sample and make any updates that might be necessary due to attrition, and also review the survey form to consider any corrections, additions or deletions that might be needed based on evidence acquired in the application of the survey in the preceding round. It is important, however, to make every effort to keep both attrition and any survey modifications to a minimum.

The executing unit clearly has a major role to play in the data collection process and subsequent analysis. In addition, the data set produced by the consulting firm should be imbedded in the overall Monitoring and Evaluation System that will be developed for APAGRO and that will be managed by the executing unit.

Finally, in addition to the cost per survey mentioned above, it is necessary to add the cost for data analysis and report preparation. In the two related projects mentioned, these activities will be performed by a separate international consultant, which is also recommended here in the case of APAGRO. The cost that has been estimated for this work is US$20,000 for the Baseline Study, US$20,000 for the Mid-term Evaluation, and US$40,000 for the Final Evaluation.

9. Case study: Evaluating the Impact of Index Insurance on Cotton Farmers in Peru[21]

This section describes the research design put in place to evaluate the impact of index insurance on the welfare of small-holder cotton farmers in Peru. In contrast to the previous two case studies, this intervention is implemented by private sector actors, namely a local insurance company and bank, instead of either the government or donors. This research project is an example of a number of recent efforts by researchers to evaluate technological and contractual innovations in rural financial markets in developing countries (McKenzie, 2009).

From an academic point of view, carrying out research with private sector actors is attractive for several reasons. First, it generates direct insights into the performance of contracts and markets, thereby providing important contextualization and feedback regarding economic theory. Second, given the recent retreat of government from direct intervention in rural financial markets, the deepening and extension of rural financial markets to low income households will, to a large degree, depend on the decisions of micro-finance institutions, rural banks, and insurance companies. On the other hand, designing impact evaluations for interventions implemented by private sector actors presents a number of challenges. First, in contrast to many government interventions, the products offered by the private sector are not free. This may lead to relatively low take-up rates (demand) by the population of interest to the researcher. Second, private actors may be reluctant or unwilling to deny access to a product that they are offering in order to create the type of control group required by researchers to identify causal impact. A potential research design that addresses these two challenges, and indeed the one used in the Peru project, is the randomized encouragement design.

This case study is organized as follows. First, how index insurance works and why there is significant excitement around it as a poverty alleviation tool is explained. Next, the pilot insurance project in Peru is described. The following section then turns to the encouragement design. After laying out the general logic of randomized encouragements and the specific impact evaluation questions that it can (and cannot) answer, the specific methodology used in Peru is explained. The last section concludes with some general (and hard learned) lessons from this specific research design experience.

---

[21] This section was written by Steve Boucher, Associate Professor of Agricultural and Resource Economics at the University of California at Davis and Conner Mullally a PhD candidate in the Department of Agricultural and Resource Economics at the University of California at Davis.

9.1 A Primer on Index Insurance

In order to understand what index insurance is, it is perhaps easiest to start with what it is not. Conventional, multi-peril crop insurance provides an indemnity when the insured farmer suffers damage to his or her crop. Prior to making an indemnity payment, the insurance company typically sends an adjustor to verify that the damages actually occurred. Two immediate challenges of conventional insurance are apparent. First, verification of damages can be quite costly, especially in developing countries where rural infrastructure is poor. Second, the insurer faces significant informational problems in the form of moral hazard (did the farmer do everything he could to avoid damages?) and adverse selection (did higher risk farmers disproportionately purchase insurance?). The end result is that markets for conventional crop insurance tend to by very thin or completely missing unless they are accompanied by significant subsidies, a luxury developing countries can ill afford.

In contrast, index insurance provides an indemnity to insured farmers when the value of an external index crosses a critical threshold, or strike-point. An ideal index should be correlated with farmers' yields but have a probability distribution that is not affected by either the actions or composition of insured farmers. For example, an index insurance contract based on rainfall might pay out if accumulated precipitation falls below the strike-point, with payouts increasing in the size of the shortfall. Other examples of indices include alternative meteorological variables, such as wind and temperature; satellite-based indices such as the NDVI, which measures deviations in plant mass from historical means; and directly measured average yields in a specified region.[22] In addition to circumventing moral hazard and adverse selection, index insurance can (at least in theory) be offered at significantly lower cost, because the insurer does not have to verify damages on insured farmers' plots.

One of the primary challenges facing the development of index insurance markets is that it offers less protection than conventional insurance. The reason is that a farmer may suffer a loss on his or her own farm but will not receive an indemnity if the index does not fall below the strike-point. The size of this so-called "basis risk" will depend on the degree to which an individual farmer's yields co-vary with the index. In general, basis risk will be reduced, and thus the value of index insurance greatest, in contexts where covariate shocks – such as drought --

[22] Barnet, Barrett, and Skees (2008) provide an overview of recent experiences with index insurance in developing countries.

account for a large fraction of total risk and when the index closely moves with the source of the covariate shock. Other challenges include insufficient quantity and quality of historical data on indices to price contracts, lack of familiarity by farmers with insurance, and lack of trust by farmers in insurance providers.

Index insurance represents a potentially powerful poverty alleviation tool. There are two primary channels of impact. First, the availability of index insurance may permit small farmers to abandon costly self-insurance strategies; for example inducing them to move away from traditional low yield but safe crops and techniques towards more productive, although riskier, alternatives. Second, if farmers are insured, lenders may be willing to offer credit to farmers previously deemed un-creditworthy because of risk or to offer larger loans to finance larger investments. In this sense, a vibrant index insurance market may "crowd-in" credit supply.

While there is significant excitement in the academic and policy world regarding index insurance, there is little empirical evidence about its performance and impact on small farmers. Cole et al. (2010) study demand for rainfall insurance in India. Their research design includes four separate treatments hypothesized to affect demand, each randomized at the household level.[23] The treatments are: price, liquidity, informational visits, and an endorsement from a trusted microfinance institution. Each treatment is found to have a significant impact on demand in the expected direction. Cai et al. (2009) use an innovative instrumental variable approach to estimate the impact of a program to insure sows in rural China. The authors create exogenous variation in insurance demand by randomizing the incentives to marketing agents that sell the insurance. The authors find positive and significant impacts of buying insurance on the number of sows owned. As described in detail below, assuming a number of conditions hold, their estimate is a Local Average Treatment Effect; an estimate of the average impact on those farmers induced to buy insurance by receiving a sales pitch from a more highly incentivized marketing agent. Finally, Gine and Yang (2008) examine the impact of rainfall insurance on the demand for formal credit in Malawi. Their research design randomizes whether loans made by a microfinance institution are bundled with the insurance. The authors find the counter-intuitive result that insurance reduces the demand for credit. They attribute this result to the fact that the credit contract already provides some insurance via limited liability so that bundling insurance with the loan implies an increase in the interest rate without any additional insurance.

[23] The authors do not evaluate the impact of insurance on farmer welfare.

The research project described below is an attempt to provide additional evidence on the performance and impact of index insurance in developing countries.

9.2 The Intervention and Primary Research Hypotheses

In 2008, a research team from the University of California at Davis and the Instituto de Estudios Peruanos, in conjunction with a local insurance company and micro-finance institution, launched a pilot program to examine the impacts of index insurance on cotton farmers in the valley of Pisco, on Peru's south coast. The valley contains approximately 22,000 irrigable hectares and is dominated by small-holders who work less than 10 hectares. Cotton is the dominant crop in the valley, accounting for between 50–75% of planted area over the last decade. On average, approximately 60% of the 5,000 farmers in the Pisco valley plant cotton each year.

The insurance product developed for this project, called Agro-Positiva, is an area yield insurance contract. The index is the average yield of planted cotton area in the valley. The first step in designing the contract was to estimate the probability density function of area yields in the valley. This was done using the 25 year time series of annual yield estimates published by the Ministry of Agriculture between 1982 and 2006.[24] The estimated distribution function was then used to calculate the actuarially fair premium. The contract established a strike-point of 1,412 kilograms per hectare, or 85% of average valley yield. If the average valley yield falls below the strike-point, policy holders are paid an indemnity of US$0.63 per kilogram below the strike-point. The final premium was set at US$48 per insured hectare. The actuarially fair premium was US$22 per hectare. The final premium reflects loading costs (administrative costs, taxes, plus profits) of the insurance company minus a 30% premium subsidy provided by the Ministry of Agriculture.

While the research team conducted the statistical analysis underlying the insurance contract, the contract is offered by a triangular arrangement involving three private institutions. The insurance contract is formally offered and registered with the Superintendence of Banks and Insurance by the Peruvian insurance company La Positiva. The insurance is sold, however, by the Caja Rural Señor de Luren, a locally-based rural bank and the largest formal lender to agriculture in the region. The Caja is essentially an agent of the insurance company; collecting a small commission for each policy sold. Finally, HanoverRe provides reinsurance to La Positiva.

[24] Using maximum likelihood, the de-trended yield data were used to fit a Weibul distribution.

9.2.1 Hypotheses to be Tested

The primary objective of the research program is to evaluate the impact of this insurance contract on three types of farmer-level outcome variables. First, as suggested above, index insurance may relax credit constraints faced by farmers, leading to an increase in farmers' participation in the formal credit market. On the supply side, quantity rationing may be reduced if insurance induces lenders to extend credit to farmers that were previously deemed too risky. The bundling of insurance with credit in the contract was designed, in part, to maximize this supply side impact. Indemnity payments are made directly to the bank in order to pay down the loan. After canceling the loan, any remaining indemnity payment goes to the farmer. On the demand side, insurance may reduce risk rationing, which occurs when farmers qualify for loans but choose not to borrow because they are afraid of losing the assets (typically land) required as collateral (Boucher et. al., 2008). As it should reduce the risk of default and collateral loss, index insurance is expected to reduce the incidence of risk rationing. In the survey, a module to directly elicit farmers' credit rationing status in the formal sector is therefore included.[25]

Second, index insurance is expected to increase the intensity of input use and investment on cotton plots. On one hand, insurance should increase a risk averse farmer's input demand, moving the farmer towards a risk neutral, profit maximizing level. On the other hand, insurance – by relaxing credit constraints – will allow this increased input demand to be met. In the case of cotton in Pisco, yield is significantly affected by the farmer's per-hectare expenditure in land preparation and fertilization. In the survey, detailed information is collected on these costs.

The first two types of hypothesized impacts can be considered intermediate outcomes. Both credit access and the intensity of input use are mechanisms through which insurance may affect farmers' welfare. The research project is also concerned with measuring impacts on welfare. As such, the household survey also measures cotton yield and net income from cotton production.

9.3 Research Design: Randomized Price Encouragement

In order to test the hypotheses and examine the impact of index insurance on the outcome variables described above, a randomized encouragement design was implemented. The section is divided into two parts. First, a simple model that will illustrate the basic idea underlying the

[25] See Boucher, Guirkinger, and Trivelli (2009) for a description of the direct elicitation methodology.

encouragement design is outlined. Then, the specific types of average treatment effects that can be answered are defined.

9.3.1 A Basic Econometric Model

The following simple model, based on Moffitt (2008), illustrates the basic ideas underlying the encouragement design. It consists of the following 2 equations:

$$y_i = \beta_i + \alpha_i d_i \quad (13)$$

$$d_i = 1\{k(z_i, \delta_i) \geq 0\}a \quad (14)$$

In equation (13), the variables y_i y_i and d_i d_i are, respectively, the observed outcome variable (yield in this example) and a binary indicator taking value 1 if the farmer buys insurance and 0 if not. β_i is the individual specific parameter giving the farmer's yield without insurance; $\alpha_i + \beta_i$ is the farmer's yield with insurance; and thus the parameter α_i is the change in yield due to purchasing insurance. Equation (14) states that the farmer decides to purchase insurance if the function $k(z_i, \delta_i)$ is positive. This decision is a function of the observed value of z_i and the unobserved parameter, δ_i. The variable z_i is an *instrument* which obeys certain assumptions, which are discussed below in greater detail.

This potential outcome framework illustrates the fundamental evaluation problem. At any given point in time, either yield with insurance, $\alpha_i + \beta_i$, or yield without insurance, β_i, can be observed but not both. Thus, the impact of insurance, α_i, for an individual cannot directly be observed. One might be tempted to compare mean yields of insurance purchasers versus non-purchasers. Since farmers self-select into the insurance market, however, this comparison would likely give a biased impact estimate. Specifically, the unobserved determinants of demand, δ_i, are likely to be correlated with the unobserved impact of insurance, α_i.[26]

So how can an unbiased average treatment effect be estimated? One possibility would be to directly randomize farmers into and out of the index insurance program. In other words, use a randomized control trial with perfect compliance, or in which everybody offered insurance purchases it, while nobody who is not offered insurance purchases it. Perfect compliance,

[26] For example, farmers with lower unobserved basis risk – who are thus be more likely to purchase insurance -- may tend to have more productive land. In this case, the average yields of non-purchasers would underestimate what the average yields of purchasers would be in the absence of insurance. The result would be an over-statement (positive bias) of the average treatment effect.

however, is unlikely for several reasons. First and foremost, it is unfeasible (and unethical) to compel farmers to purchase insurance.

An alternative strategy is to use an instrumental variable strategy in which the researcher randomizes a variable that affects the probability that a farmer purchases insurance but does not directly affect yield. Conceptually, the simplest strategy is to randomize farmers' *eligibility* to purchase insurance. This research design would require a mechanism by which randomly chosen farmers are not allowed to purchase insurance (ineligibles) while others are allowed to purchase it if they want it (eligibles). In some contexts this will be a viable strategy and, if followed rigorously (i.e., the ineligible group is truly randomly selected and effectively denied access), can yield consistent estimates of the average treatment effect. This is the strategy pursued by Giné and Yang (2008) in their study of the impact of rainfall insurance on credit demand in Malawi. In the Peruvian context, it was not feasible to randomize eligibility because neither the insurance company nor the lender was willing to deny access to a random group of farmers.

Given the infeasibility of randomizing eligibility, the option of a randomized encouragement design was used. In this case, while all farmers are eligible to purchase insurance, a subset of farmers is randomly selected to receive an additional "encouragement" to purchase insurance in the form of discount coupons that lower the price of the premium. To see how this random encouragement can generate consistent estimates of average treatment effects, return to the basic equations. Now let z_i indicate the coupon value. For simplicity, assume there is a single coupon value so that z_i takes value 1 if the farmer receives a coupon and 0 if not.

Taking conditional expectations of the model with respect to z_i, yields:

$$E\left(y_i \mid z_i = z\right) = E\left(\beta_i \mid z_i = z\right) + E\left(\alpha_i \mid d_i = 1, z_i = z\right) P\left(d_i = 1 \mid z_i = z\right) \quad (15)$$

$$E\left(d_i \mid z_i = z\right) = P\left(d_i = 1 \mid z_i = z\right) = P\left(k\left(z_i, \delta_i\right) \geq 0\right) \quad (16)$$

where $P(.)$ denotes probability and (15) follows because d_i is a [0,1] binary variable.

In order to identify average treatment effects, the instrument (our coupon) must satisfy the following four criteria (Moffitt 2008). First, z_i must satisfy *mean independence*:

$$E\left(\beta_i \mid z_i = z\right) = \beta \quad (17)$$

Independence implies that coupon assignment does not depend on what individuals' average outcomes (farmers' yields) would be without insurance. Since the evaluation will be comparing average outcomes across farmers with and without coupons, this criterion ensures that the non-recipients are a good counterfactual to the recipients. By randomly assigning who gets coupons, it is more likely that the independence criterion is satisfied.[27]

Second, z_i must satisfy an *exclusion restriction*:

$$E\left(\alpha_i \mid d_i = 1, z_i = z\right) = g\left(P\left(d_i = 1 \mid z_i = z\right)\right) \quad (18)$$

The left hand side is the average gain in the outcome among those assigned $z_i = z$. The right hand side rewrites this expectation as the function *g*(), which in turn is a function of $P\left(d_i = 1 \mid z_i = z\right)$, the probability that a person assigned $z_i = z$ buys insurance. This criterion requires that the instrument (coupon) has no direct effect on the outcome of interest (yields); instead the instrument only affects the outcome through its impact on the probability that individuals purchase insurance. Given that α_i is likely to vary across farmers, the impact of insurance that is measured in an encouragement design depends crucially on the composition of farmers that are induced to purchase insurance by the instrument (coupon). This point will be returned to shortly. For now, note that a coupon that lowers the price of insurance should have no direct effect on farmers' investment decision and is thus likely to satisfy the exclusion restriction.

Third, z_i must be *relevant*:

$$Cov\left(z_i, d_i\right) \neq 0 \quad (19)$$

This criterion implies that the coupon has some predictive power with respect to whether or not farmers buy insurance. The stronger is the covariance, the easier it will be to detect impacts of a given size.

These first three criteria are similar to those underlying a conventional instrumental variable approach. In the case of an encouragement design, one additional criterion is required that will permit an interpretation of the estimates as an average treatment effect. This fourth assumption is *monotonicity*:

$$d_i = 1 \mid z_i = z^j \rightarrow d_i = 1 \mid z_i = z^k \text{ and } d_i = 0 \mid z_i = z^k \rightarrow d_i = 0 \mid z_i = z^j \;\; \forall i, \forall z^k > z^j. \quad (20)$$

[27] As in the case of RCT's, randomization of the instrument does not necessarily imply that independence is met, especially in small samples. The researcher should always check the quality of the randomization by comparing means of variables hypothesized to be related to the outcome variable across groups assigned different instrument values.

This criterion, introduced by Imbens and Angrist (1994), says that the values of the instrument z_i can be ordered in such a way that moving from z_i^j to z_i^k, the sign of the impact on the decision to participate must be the same for everyone. In this case, giving farmers a coupon (or a larger coupon in the case of multiple coupon values) should either induce them to buy insurance or have no effect, but not push some people to buy insurance while dissuading others.[28]

With these assumptions in hand, equations (13) and (14) can be rewritten in an estimable form as:

$$y_i = \beta + g\left(P\left(d_i = 1 \mid z_i = z\right)\right)P\left(d_i = 1 \mid z_i = z\right) + e_i \quad (21)$$

$$d_i = P\left(d_i = 1 \mid z_i = z\right) + u_i \quad (22)$$

Equation (21) states that the value of y_i for everyone assigned $z_i = z$ is equal to the mean outcome without insurance, β, plus the average impact of insurance given the share of this sub-population that buys insurance, $g(P(.))$, weighted by the share of this same sub-population buying insurance, $P(.)$, plus white noise. The error terms e_i and u_i are random variables with expected values equal to 0, conditional on the probability of buying insurance.

9.3.2 What treatment effects can be estimated?

Theory suggests that the impact of insurance is likely to be quite different across individuals depending on several difficult to observe factors including risk aversion, credit constraints, farm aptitude, etc. For this reason, the theoretical model is written to allow α_i, and thus also the function *g(.),* to vary across farmers. This has important implications for the type of treatment effect that can be estimated.

The first treatment effect that one might want to estimate is the marginal treatment effect (MTE):

$$\frac{\partial E\left(y_i\right)}{\partial P\left(d_i = 1 \mid z_i = z\right)} = g\left(P\left(d_i = 1 \mid z_i = z\right)\right) + \frac{\partial g\left(P\left(d_i = 1 \mid z_i = z\right)\right)}{\partial P\left(d_i = 1 \mid z_i = z\right)}P\left(d_i = 1 \mid z_i = z\right) \quad (23)$$

[28] If providing a coupon had opposing effects on farmers, then the estimated expected gain from buying insurance would include the expected gain among those induced to buy insurance by the instrument, minus the expected gain for those induced to not buy the insurance. In general, the direction of the inequalities in equation 20 is not important, as they could be reversed and the assumption would still serve its purpose.

If the outcome variable of interest is yield, the MTE would give the change in the average yield due to an arbitrarily small change in the share of the farmers purchasing insurance. If the impact of insurance is heterogeneous across farmers, the MTE will vary depending on who the marginal farmers are that are induced to participate. To estimate MTEs, first it is necessary to estimate insurance demand (equation (22)) using a model, such as logit or probit, that generates a continuous range for the probability of purchasing insurance.[29] Once predicted probabilities of buying insurance have been generated for the sample, the MTE can be estimated by picking a continuous flexible functional form for $g(.)$, and estimating it along with its derivative with respect to the probability of buying insurance.

Conceptually, the MTE is the building block for all other treatment effects, as the latter can be expressed as weighted averages of the former (Heckman and Vytlacil, 2007). In practice, however, the MTE is not commonly estimated because the data requirements are steep; it requires a sample with sufficient observations at multiple values of the instrument, or a continuous instrument.[30]

A much more commonly estimated treatment effect with encouragement designs is the Local Average Treatment Effect (LATE). The LATE, introduced by Imbens and Angrist (1994), is the discrete version of the MTE. The LATE for any pair of instrument values z^1 and z^2, is:

$$\frac{E\left(y_i \mid z_i = z^2\right) - E\left(y_i \mid z_i = z^1\right)}{P\left(d_i = 1 \mid z_i = z^2\right) - P\left(d_i = 1 \mid z_i = z^1\right)} \quad (24)$$

The LATE is quite intuitive. Suppose, for example, that z is the price of insurance, and half of the sample has been randomly chosen to receive a discounted price z^2 should they elect to buy insurance, while the other half must pay the market price z^1. The numerator in equation (24) is the difference in expected yields between farmers offered the discounted premium and those offered the market price, while the denominator is the difference in the fraction of farmers purchasing insurance across the two groups. The LATE is thus the average impact of buying insurance on the outcome of interest for "compliers," i.e., individuals who would buy insurance if assigned the lower price, z^2 , but not buy insurance if assigned the higher price, z^1. Impacts on

[29] Semi-parametric estimators offer more flexible alternatives to the logit and probit models. See the reviews of these methods by Ichimura and Todd (2007) and Chen (2007).

[30] For applications of MTE estimation, see Heckman et. al. (2006), Carneiro et al (2006), and Carneiro and Lee (2009).

individuals who would always buy the insurance (the "always-takers") or would never buy insurance ("never-takers") given the values of z are not represented in the estimated LATE (Angrist and Imbens, 1995). The inequalities in the monotonicity condition (equation (20)) rule out the possibility of "defiers," i.e., individuals who would only buy the insurance if not assigned the lower price.

Finally, note that estimation of the LATE is straightforward. It can be estimated non-parametrically by replacing the components of (24) with sample averages, or via two-stage least squares.

9.4 Implementation in Pisco

In this section, the operationalization of the encouragement design in the research setting in Pisco, Peru is described.

9.4.1 Choice of Instruments

The primary instrument used in the encouragement design is randomly distributed coupons which lower the price of the insurance premium. Four different coupon values were distributed: 15 S/., 35 S/., 65 S/., and 90 S/. per insured hectare.[31] The 90 S/. coupon lowered the effective price of the insurance to just below the actuarially fair rate. Five hundred and forty-three cotton farmers were randomly selected to receive the coupons, with one-fourth receiving each value.[32] Randomly distributed coupons were expected to serve as an ideal instrument for the encouragement design as it would likely meet the four criteria listed above. First, by randomizing the distribution of the coupons, the value of the coupon received should be independent of farmers' yields and other outcome variables. Second, the exclusion restriction should be satisfied as variation in the price of insurance may impact the composition of insurance purchasers but should have no direct impact on farmers' outcome variables. Third, given their monetary value, the coupons were expected to significantly affect farmers' demand for insurance. Finally, monotonicity is likely to be satisfied as a lower price should induce some farmers to purchase, but would never dissuade a farmer who would purchase without the coupon from buying with a coupon.

---

[31] The Peruvian currency is the Nuevo Sol and \$1 = S/.3.2.

[32] The cost of the redeemed coupons is assumed by the research team.

In addition to the coupons, invitations to insurance education sessions were randomly distributed. In the sessions, farmers participated in a game, based on experimental economics, which simulated farmers' credit, technology and insurance purchase decisions. Participating farmers earned between US$2 and US$15, depending on their choices and random draws simulating covariate and idiosyncratic shocks.[33] In addition, a member of the research team made a brief presentation describing index insurance. The distribution of invitations had two levels of randomization. First, 16 of the 40 irrigation sub-sectors in which the valley is divided were randomly chosen. In each of the sub-sectors, a single educational session was carried out. Within each sub-sector, invitations to the session were delivered to 60 randomly selected cotton farmers. Only invited individuals were allowed to participate. Approximately 75% of invitation recipients attended the sessions. While the original intention was that the invitations to the education sessions would serve as a second instrument, eventually it was concluded that the sessions may violate two of the criteria. First, because significant information was provided about yield risk, it was possible that participation in the sessions could alter farmers' perceptions of risk which, in turn, may have a direct impact on production and credit decisions. Second, for related reasons, monotonicity may not hold. While the sessions likely induce some farmers to purchase insurance, they may dissuade other farmers who, in the absence of the education sessions, would have purchased insurance for example because of heightened sensitivity to risk. While invitations to the education sessions could (and should) still be used as a control variable for both insurance demand and impact, it was decided not to rely on this variable as an instrument.

9.4.2 Sample Size Calculations

In order to choose the sample size, a slightly modified version of the power calculation given by equation (7) in section 5 was used. Specifically, at 80% power and a significance level of 5%, the minimum detectable effect size, α_{MDE}, is given by:

$$\alpha_{MDE} = 2.49 * \sqrt{\frac{\sigma^2(1-R^2)}{NP(1-P)}\frac{1}{c-s}} \quad (25)$$

where σ^2 is the variance of yield or an alternative outcome variable. R^2 is the share of the variance of yields explained by the model. N is the total sample size, and P is the share of the

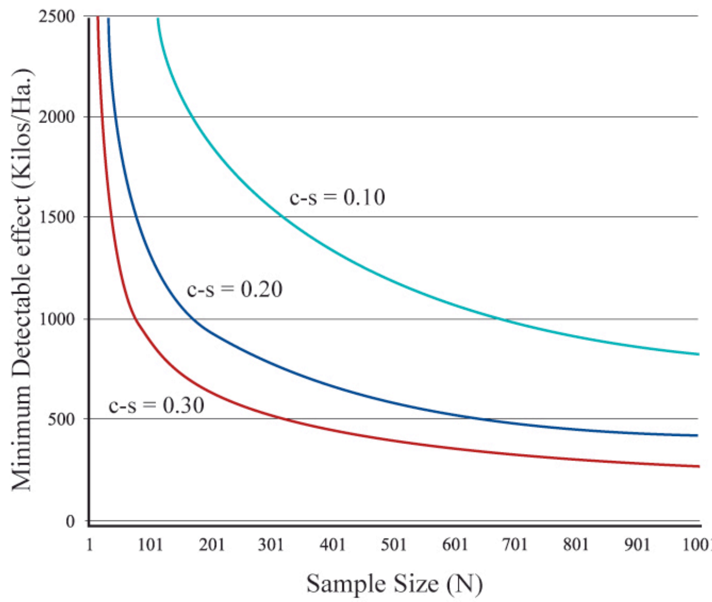
[33] The daily wage in agriculture is approximately $7.

sample that receives the encouragement (a coupon of any size). Finally, *c-s* is the difference in compliance rates between encouraged farmers (coupon recipients) and non-encouraged farmers (coupon non-recipients). Specifically, c is the share of coupon recipients that purchase the insurance while s is the share of non-recipients who buy the insurance. Equation (25) demonstrates an important risk of the encouragement design. If the instrument is weak, then compliance rates between encouraged and un-encouraged farmers will be low. The term $1/(c\text{-}s)$ will be large and, as a result, the precision of the estimate will be low, thereby requiring the true impact to be quite large in order to be detected for a given sample size. If, on the other hand, there is perfect compliance, the term $1/(c\text{-}s)$ takes its minimum value equal to 1.

Determining the sample size requires making some educated guesses, since neither σ^2, R^2, nor the compliance rates, c and s, are known. To make the calculations, R^2 is set to 0.3, which is a typical R^2 value for yield regressions. Taking advantage of a panel data set of farm yields conducted by the Ministry of Agriculture, the variance of cotton yields, σ^2, was estimated. This data set, called ENAPROVE, was applied annually to approximately 250 cotton farmers between 2002 and 2006 in Pisco. Given the potentially strong impacts on the precision of the estimates to compliance rates, equation (25) is used to calculate the necessary sample sizes for different compliance rates. Figure 6 graphs α_{MDE} as a function of the sample size for several different compliance rates. In all cases, sample is equally split between encouraged and un-encouraged farmers.

Figure 6 shows the sensitivity of the minimum detectable effect size to compliance rates. At a sample size of 800, insurance would have to raise yields by just under 1,000 kilograms per hectare, which is over 60% of mean yields, in order to be detected when the coupons raise demand by just 10%. This drops to just under 500 kilos per hectare when coupons raise demand by 20% and to 311 kilos per hectare, or about 18% of mean yields, when coupons raise demand by 30%. Based on these calculations, a sample size of 800 farmers was chosen, with an equal split between coupon recipients and non-recipients.

Figure 6: Minimum Detectable Effect as a Function of Sample Size



9.4.3 Sample Frame and Sample Selection

In order to draw the sample of cotton farmers, the research team worked closely with the Irrigation Commission (*Junta de Regantes*). To gain access to irrigation water, each farmer must file with the Commission a production and irrigation plan (*plan de cultivo y riego*). The Commission thus maintains a list of the population of farmers and the areas planted to each crop. Using this list, farmers throughout the valley were randomly selected for inclusion. A letter describing the research and notification of an enumerator visit along with the coupon was distributed to each farmer selected in the sample by the Commissions field staff (*sectoristas*).[34]

9.4.4 Household Surveys

A baseline survey was applied to the 800 farm households in August of 2008, prior to the first year of insurance availability. The survey collected data on household demographics, land holdings, cropping patterns, detailed production and cost data for cotton, and a credit market module designed to explore terms of credit access and identify households' credit rationing status. The survey was repeated in August 2009, with a final survey planned for August 2011.

[34] The research team paid a small honorarium plus gasoline costs to each sectorista and contributed a printer to the central office in recognition of their support.

Cuanto, a local firm with significant experience carrying out agricultural surveys, was hired to administer the survey. Field supervision was provided both by Cuanto as well as a team of graduate students from the Instituto de Estudios Peruanos and UC-Davis.

9.4.5 Budget of Survey Administration

As described above, the research team hired a local survey firm to administer the survey, enter the data, and deliver a cleaned data set in STATA format.[35] The survey firm charged US$85 per survey, for a total cost of ($85/survey)*(800 surveys) = US$68,000 per survey round. The research team also assumed the cost of measuring average valley yield to determine whether or not an indemnity payment would be made. This required applying a separate, but much shorter, yield survey to 650 households. The same survey firm was hired to carry out the yield survey at a cost of US$11 per survey, or ($11/survey)*(650 surveys) = US$7,150 per round. This gives a total cost of US$75,150 per survey round. With three annual survey rounds, the total research cost is approximately ($75,150/round)*(3 rounds) = US$214,500.

9.5 Pisco sour?[36] Lessons from a Discouraging Encouragement Design

In spite of the best intentions of the research team, the demand for insurance among farmers in Pisco has been disappointingly low. In the first year, a total of 51 farmers purchased the insurance, with coverage of just over 140 hectares. This rose to 75 farmers with 220 hectares in the second year and 150 farmers with just over 400 hectares in the third year. While the increase in demand raises hope for the continuing development of the market, the initially low take-up rates greatly reduce the ability of the team to conduct impact evaluation of the insurance program. Most importantly, the coupon turned out to be a surprisingly weak instrument, with take-up rates essentially independent of whether or not farmers received a coupon and, conditional on receiving a coupon, on the size of the coupon. The fraction of coupon recipients purchasing insurance across all three years is less than 5% greater than non-recipients. Referring back to Figure 6, with a sample size of 800, it would be necessary for insurance to more than double farmers' yields, which is clearly unreasonable, in order to detect any significant effect.

[35] The research team developed and field tested the questionnaire, created the sample frame and drew the sample. The research team also trained the survey firm project managers and enumerators. Neither these costs, nor the time of UC-Davis faculty and graduate students, are included in the budget figures reported in the text.

[36] Thanks to Alan DeBrauw for suggesting this sarcastic, but creative, title.

So what went wrong? Why has take-up been so low and farmers so unresponsive to price reductions? Several hypotheses have been explored which are briefly outline here.

First, farmers' understanding of index insurance may remain low. While the experimental game has high potential for teaching farmers about index insurance, it appears that the games were less effective than initially hoped at conveying the basics of the insurance contracts.[37] Based on an exit survey from the games and focus group discussions, many farmers did not understand that the payout depends on average instead of individual yields. In addition, many farmers whose yields were consistently higher than the valley average were quite skeptical about buying the insurance because their *own* yields were very unlikely to fall below the strike-point. Additional work is needed in improving farmer education, in particular to convey the notion that the value of the insurance depends on the degree of co-movement between an individual farmer's yields and the index (in this case valley average yield).[38]

A second problem that emerged rather unexpectedly was a resistance to the project from the manager of the local branch of the Caja Rural. In hindsight, the research team failed to develop sufficient understanding of the incentives in the management system in the Caja. At the outset of the project, the research team negotiated a participation agreement with the board of directors of the Caja. The board of directors, who meet in the departmental capital in the city of Ica, then passed the decision down to the manager of the Caja branch in Pisco. For a number of reasons, this manager was not enthusiastic with the project. First, the manager likely resented the process; the order to implement an experimental pilot program came down from above without any input from the branch itself. Not only was the order to participate very vertical, but it also implied costs in terms of time and training for the loan officers who would be the face of the insurance product. Second, the board of directors ordered an interest rate cut on loans for farmers who also purchased the insurance. It was later learned that the manager resented this because he felt it reduced the branch's earnings. Although in the long run, insurance would likely reduce

[37] At the end of the game sessions, a brief survey was administered to farmers. One question asked if the indemnity payout depended on the farmer's draw from the idiosyncratic shock bag. Just over a quarter of the farmers incorrectly said that it did.

[38] An entirely different, although perhaps no less important, challenge is the notion of "average". Many farmers seem to equate "average yield" (*rendimiento promedio*) with the parcel's potential yield or the yield they would expect to get in a good year. Similarly, farmers seem to discount exceptionally bad years from their mental calculation of average yields. This factor tend to make farmers' perceptions of average yields (both their own and in the valley) quite a bit higher than the statistical mean. Equivalently, farmers' subjective pdf's of average valley yields may be shifted significantly to the right relative to the pdf generated from official statistics. If this is the case, then farmers would underestimate the value of the insurance and be less likely to buy it.

default rates and thus offset the interest loss, the manager was understandably concerned with the short run earnings position of his branch.

Third, communication between the insurance company and the Caja was less than optimal. For example, there was confusion about who would lead the marketing campaign for the insurance product. By the time the confusion was cleared up, the insurance sales season had already begun.[39]

9.6 Concluding Remarks

Encouragement designs offer a potentially viable research design in contexts where randomized control trials are not feasible. This is likely to be the case for the types of innovative contract design in financial markets, such as the index insurance example described here, because private actors—insurance companies and lenders—may be reluctant to deny access to new products to maintain a strict "control" group. Encouragement designs can also be valuable when participation rates are likely to be low, thus reducing the precision of impact estimates. Mullally (2010) suggests that encouragement designs may be especially valuable in contexts of new products or markets. Providing additional incentives, such as below market prices, can induce farmers who would otherwise not participate for lack of understanding or familiarity to participate. Viewed this way, under certain circumstances an encouragement design could be viewed as a public policy to get nascent markets off the ground.

As evidenced from this experience with index insurance in Peru, encouragement designs face a number of challenges. First and foremost, identification of average treatment effects requires an effective instrument that induces participation in the project or, in this case, take-up of the contract. If instead the instrument turns out to be weak, then the overall research agenda may be compromised. In retrospect, the research team would have been well served carrying out more preliminary research, perhaps via focus groups, to evaluate the likely effectiveness of the coupons.

[39] Attempts have been made to address these problems by increasing communications flow between all parties and, in particular, by attempting to create incentives for the manager to fully get behind the project. In part this has taken the form of including the manager and the loan officers in discussions about how to improve the product, making it more valuable to their clients. Monetary incentives have been provided for loan officers and the manager based on the number of policies sold.

A second caveat to encouragement designs, and indeed to any instrumental variable approach, is that the researcher must be cautious in interpreting the estimated treatment effect. Typically, encouragement designs permit estimation of local average treatment effects, which are the average impact on compliers, those induced to participate by the encouragement. If theory suggests that impacts vary across the population, in particular across difficult to observe characteristics, then the composition of the compliers must be considered. Are the compliers the most interesting group for policy purposes? In the Peru example, it is important to ask how useful it would be to learn the impact of insurance on farmers that are only induced to purchase insurance if they receive a very high level of price discount. Given the fiscal constraints of governments, it is unlikely that these farmers would participate in any insurance program that would exist in the real world.

Finally, the experience suggests that, while there are a unique set of challenges in designing impact evaluations around private sector interventions, collaboration with the private sector can be both feasible and productive. This is particularly encouraging given that the types of financial market innovations most likely to benefit until-now excluded segments are likely to be driven by these actors. An increasingly rich set of innovative and rigorous collaborations across academia, the public sector and the private sector is anticipated as a means of pushing poverty alleviation policies forward in the future.

References

Abadi-Ghadim, A.K., D.J. Pannell, and M.P. Burton. 2005. “Risk, Uncertainty, and Learning in Adoption of a Crop Innovation”. *Agricultural Economics* 33(1):1-9.

Adesina, A.A., and J. Baidu-Forson. 1995. “Farmers′ perceptions and Adoption of New Agricultural Technology: Evidence from Analysis in Burkina Faso and Guinea, West Africa”. *Agricultural Economics* 13(1):1-9.

Adesina, A.A., and M. Zinnah. 1993. “Impact of Modern Mangrove Swamp Rice Varieties in Sierra Leone and Guinea”. *International Rice Research Notes* 18(4):36.

Anderson, J., and G. Feder. 2007. “Agricultural Extension”. In: R. Evenson and P. Pingali, editors. *Handbook of Agricultural Economics*, vol. 3, Agricultural Development: Farmers, Farm Production and Farm Markets. Amsterdam, The Netherlands: Elsevier, North-Holland.

Angelucci, M., and V. Di Maro. 2010. “Project Evaluation and Spillover Effects”. Impact-Evaluation Guidelines, Strategy Development Division, Technical Notes No. IDB-TN-136. Inter-American Development Bank, Washington, D.C.

Angelucci, M., and G. De Giorgi. 2009. “Indirect Effects of an Aid Program: How do Cash Injections Affect Ineligibles’ Consumption?”. *American Economic Review* 99(1):486-508.

Angrist, J. 2001. “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice”. *Journal of Business & Economic Statistics* 19(1):2-28.

Angrist, J., and J.S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. New Jersey, United States: Princeton University Press.

Angrist, J., and V. Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”. *The Quarterly Journal of Economics* 114:533-7.

Avitabile, C., and V. Di Maro. 2007. “Spillover Effects in Healthcare Programs: Evidence of Social Norms and Information Sharing”. Paper presented at the 2009 Latin America and Caribbean Economic Association. Buenos Aires, Argentina.

Banerjee, A. 2007. *Making Aid Work.* Cambridge: MIT Press.

Barham, B.L., M.C. Carter, and W. Sigelko. 1995. "Agro-Export Production and Peasant Land Access; Examining the Dynamic between Adoption and Accumulation". *Journal of Development Economics* 46:85-107.

Barham, B.L., S. Boucher, and M.C. Carter. 1996. "Credit Constraints, Credit Unions, and Small-Scale Producers in Guatemala". *World Development* 24(5):793-806.

Barnett, B., C. Barrett, and J. Skees. 2008. "Poverty Traps and Index Based Risk Transfer Products". *World Development*, 36:1766-1785.

Bartlett, J., J. Kotrlik, and C. Higgins. 2001. "Organizational Research: Determining Appropriate Sample Size in Survey Research". *Information Technology, Learning, and Performance Journal* 19(1):43-50.

Besley, T., and A. Case. 1994. "Diffusion as a Learning Process: Evidence from HYV Cotton", Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies, No 228.

http://econpapers.repec.org/RePEc:pri:rpdevs:228

Birkhaeuser, D., R.E. Evenson, and G. Feder. 1991. "The Economic Impact of Agricultural Extension: A Review". *Economic Development and Cultural Change* 39:607-650.

Boucher, S.R., M.C. Carter, and C. Guirkinger. 2008. "Risk Rationing and Wealth Effects in Credit Markets: Theory and Implications for Agricultural Development". *American Journal of Agricultural Economics* 90(2):409-423.

Boucher, S.R., C. Guirkinger, and C. Trivelli. 2009. "Direct Elicitation of Credit Constraints: Conceptual and Practical Issues with an Application to Peruvian Agriculture". *Economic Development and Cultural Change*, 57(4):609-640.

Bravo-Ureta, B.E., and A. E. Pinheiro. 1993. "Efficiency Analysis of Developing Country Agriculture: A Review of the Frontier Function Literature". *Agricultural and Resource Economics Review* 22:88-101.

Bravo-Ureta, B.E. 2009. "Capacitación y Evaluación de Impacto de APAGRO, Nicaragua". Informe de Consultoría.

Bravo-Ureta, B.E., D. Solís, V. Moreira, V. Maripani, A. Thiam, and T. Rivas. 2007. "Technical Efficiency in Farming: A Meta-Regression Analysis". *Journal of Productivity Analysis* 27(1):57-72.

Bravo-Ureta, B.E., A. Almeida, D. Solís, and A. Inestroza. 2011. "The Economic Impact of MARENA's Investments on Sustainable Agricultural Systems in Honduras". Forthcoming. *Journal of Agricultural Economics*.

Cai, H., Y. Chen, H. Fang, and L. Zhoi. 2009. "Microinsurance, Trust, and Economic Development: Evidence from a Randomized Natural Field Experiment". NBER Working Paper 15396. Accessed October 15, 2009.

http://www.nber.org/papers/w15396.

Caliendo, M., and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching". *Journal of Economic Surveys* 22(1): 31-72.

Carneiro, P., J. Heckman, and E. Vytlaci. 2006. "Estimating Marginal and Average

Returns to Education". Unpublished working paper. Accessed September 15, 2009.

http://www.ucl.ac.uk/~uctppca/school_all_2006-10-30b_mms.pdf.

Carneiro, P., and S. Lee. 2009. "Estimating Distributions of Potential Outcomes Using Local Instrumental Variables with an Application to Changes in College Enrollment and Wage Inequality". *Journal of Econometrics*, 149(2):191-208.

Carter, M.C. 1989. "The Impact of Credit on Peasant Productivity and Differentiation in Nicaragua". *Journal of Development Economics* 31(1):13-36.

Cavatassi, R., M. Gonzalez, P. Winters, J. Andrade-Piedra, P. Espinosa, and G. Thiele. 2010. "Linking Smallholders to the New Agricultural Economy: the Case of the Plataformas de Concertación in Ecuador ". Forthcoming. *Journal of Development Studies*.

Cerdán-Infantes, P., A. Maffioli, and D. Ubfal. 2008. "The Impact of Agricultural Extension Services: The Case of Grape Production in Argentina". Office of Evaluation and Oversight (OVE). Inter-American Development Bank. Washington, D.C.

Chakravarty, S. 2009. "The Impact of Fertilizer Provision on Agricultural Production in HIV-Affected Households". Paper presented at the workshop *Evaluating the Impact of*

Agricultural Projects in Developing Countries held at the Inter-American Development Bank, Washington, D.C. October 19-20, 2009.

Chen, S. 2007. “Large Sample Sieve Estimation of Semi-nonparametric Models”. In Heckman, J., Leamer, E. (Eds.), *Handbook of Econometrics*, vol. 6B, 5549-5623. Elsevier, Amsterdam.

Cole, Shawn A., Xavier Gine, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery. (Revised November 2010.) "Barriers to Household Risk Management: Evidence from India." Harvard Business School Working Paper, No. 09-116, April 2009.

Conley, T.G., and C.R. Udry. 2010. “Learning About New Technology: Pineapple in Ghana”. *The American Economic Review* 100(1):35-69.

Conning, J., and C. R.Udry. 2007. “Rural Financial Markets in Developing Countries”. In: R. Evenson and P. Pingali, editors. *Handbook of Agricultural Economics*. vol. 3, Agricultural Development: Farmers, Farm Production and Farm Markets. Elsevier, North-Holland.

de Janvry, A. 1972. “Optimal Levels of Fertilization under Risk: The Potential for Corn and Wheat Fertilization under Price Policies in Argentina”. *American Journal of Agricultural Economics* 54:1-10.

Del Carpio, X., N. Loayza, and G. Datar. 2010. “Is Irrigation Rehabilitation Good for Poor Farmers? A Non-experimental Impact Evaluation of a Peru Irrigation Project”. Forthcoming. *Journal of Agricultural Economics*.

Del Carpio, X., and M. Maredia. 2009. “Measuring the Impacts of Agricultural Projects: A Meta-Analysis of the Evidence”. Working Paper, Independent Evaluation Group, World Bank. Washington, D.C.:WB.

Dercon, S., and L. Christiaensen. 2008. “Consumption, Risk, Technology Adoption and Poverty Traps: Evidence from Ethiopia”. The Centre for the Study of African Economies Working Paper Series, Centre for the Study of African Economies, Paper 265.

Duflo, E., and M. Kremer. 2005. “Use of Randomization in the Evaluation of Development Effectiveness,” In: G. Pitman, O. Feinstein and G. Ingram, editors. *Evaluating Development Effectiveness*. New Brunswick: Transaction Publishers.

Duflo, E., Glennerster, R. and M. Kremer. 2008. “Using Randomization in Development Economics: A Toolkit”. In: T.P Schultz, editor. *Handbook of Development Economics*, vol. 4.

Ellis, F. 1988. *Peasant Economics: Farm Households and Agrarian Development*. Cambridge University Press.

Feder, G. 1980. “Farm Size, Risk Aversion and the Adoption of New Technology under Uncertainty “ *Oxford Economic Papers*, New Series 32(2):263-283.

Feder, G. and D. Umali. 1993. “The Adoption of Agricultural Innovations: A Review”. *Technological Forecasting and Social Change* 43:215-239.

Feder, G., R.E. Just, and D. Zilberman. 1985. “Adoption of agricultural innovations in Developing Countries: A Survey”. *Economic Development and Cultural Change* 33(2):255-298.

Foster, A., and M. Rosenzweig. 1995. “Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture”. *The Journal of Political Economy* 103(6): 1176- 1209.

Foster, A., and M. Rosenzweig. 2010. “Microeconomics of Technology Adoption”. *Annual Review of Economics* 2: 395-424.

Gine, X., and D. Yang. 2008. “Insurance Credit, and Technology Adoption: Field Experimental Evidence from Malawi”. *Journal of Development Economics*, 89(1):1-11.

Gine, Xavier & Yang, Dean, 2009. "Insurance, credit, and technology adoption: Field experimental evidence from Malawi," *Journal of Development Economics*, 89(1):1-11.

González, V., P. Ibarrarán, A. Maffioli, and S. Rozo. 2009. “The Impact of Technology Adoption on Agricultural Productivity: The Case of the Dominican Republic”. OVE Working Papers 0509. Office of Evaluation and Oversight (OVE). Inter-American Development Bank, Washington, D.C.

Heckman, J., U. Urzua, and E. Vytlacil. 2006. “Understanding Instrumental Variables in Models with Essential Heterogeneity”. *The Review of Economics and Statistics*, 88(3):389-432.

Heinrich, C., A. Maffioli, and G. Vázquez. 2010. "A Primer for Applying Propensity-Score Matching". Impact Evaluation Guidelines, Strategy Development Division, Technical Notes No.IDB-TN-161. Inter-American Development Bank, Washington, D.C.

Hiebert, L. D. 1974. "Risk, Learning and the Adoption of Fertilizer Responsive Seed Varieties". *American Journal of Agricultural Economics* 56:764-768.

Ichimura, H., and P. Todd. 2007. "Implementing Nonparametric and Semiparametric Estimators". In Heckman, J., Leamer, E. (Eds.), *Handbook of Econometrics*, vol. 6B, 5369-5468. Elsevier, Amsterdam.

IDB (Inter-American Development Bank). 2010. *Development Effectiveness Overview, Special Topic, Assessing the Effectiveness of Agricultural Interventions*. Inter-American Development Bank, Washington, D.C.

IDB. 2008. Nicaragua-Programa de Apoyos Productivos Agroalimentarios (NI-L1020): Propuesta de Préstamo.

Imbens, G., and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects". *Econometrica*, 62(2):467-475.

Imbens, G., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice". *Journal of Econometrics* 142(2):615-635.

Jack, K. 2009. "Barriers to agricultural technology adoption: Market failures". White paper prepared for the Agricultural Technology Adoption Initiative, JPAL (MIT) / CEGA (Berkeley).

Jalan, J., and M. Ravallion. 1999. "Are the poor less well insured? Evidence on vulnerability to income risk in rural China". *Journal of Development Economics.* 58(1):61-81.

Joshi, G., and S. Pandey. 2005. "Effects of Farmers' Perceptions on the Adoption of Modern Rice Varieties in Nepal". Conference on International Agricultural Research for Development. Stuttgart-Hohenheim.

Lee, E. 2005. "Egalitarian Peasant Farming and Rural Development: The Case of South Korea", In: D. Ghai et al. editors. *Agrarian Systems and Rural Development*.

Levy, D. and Cossens. (2011). "Designing Impact-Evaluation". Impact-Evaluation Guidelines, Strategy Development Division, Technical Notes. Forthcoming. Inter-American Development Bank, Washington, D.C.

Lopez, F., and A. Maffioli. 2008. "Technology Adoption, Productivity and Specialization of Uruguayan Breeders: Evidence from Impact Evaluation". OVE Working Papers OVE/WP-07/08. Office of Evaluation and Oversight (OVE). Inter-American Development Bank, Washington, D.C.

MAGFOR. 2009. Ministerio Agropecuario y Forestal, Republica de Nicaragua. Programa de Apoyos Productivos Agroalimentarios. APAGRO/2055/BL-NI, Operación NI-L1020: Reglamento Operativo del Programa (ROP).

Maluccio, J., and R. Flores. 2004. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social". Discussion Paper 184. International Food Policy Research Institute, Washington, D.C.

McKenzie, D. 2009. "Impact assessments in finance and private sector development: What have we learned and what should we learn?" *World Bank Research Observer*, 25(2):209-233.

Miguel, E., and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities". *Econometrica* 24: 159-217

Moffitt, R. 2008. "Estimating Marginal Treatment Effects in Heterogeneous Populations". Unpublished working paper. Accessed August 26, 2009.

http://www.econ.jhu.edu/People/Moffitt/welfls0_v4b.pdf.

Moser, C., and C. Barrett. 2003. "The Complex Dynamics of Smallholder Technology Adoption: the Case of SRI in Madagascar". Working Paper WP2003-20. Department of Applied Economics and Management, Cornell University.

Mullally, C. 2010. "Perceptions and Participation: Randomized Encouragement Designs and the Demand for Index Insurance". Unpublished working paper. University of California-Davis.

ProVention Consortium. 2007. *Marco Lógico y Matriz de Resultados: Nota de Orientación 6*. In Herramientas para la Integración de la Reducción del Riesgo de Desastres, Geneva, Switzerland.

Ravallion, M. 2008. "Evaluating Anti-Poverty Programs". In: T. P. Schultz and J. Strauss, editors. *Handbook of Development Economics*, vol. 4. Amsterdam, The Netherlands: North-Holland.

Ravallion, M. 2009. "Should the Randomistas Rule?" *Economists' Voice*.

www.bepress.com/ev.

Rodriguez, D., R. Rejesus, and M. Aragon. 2007. "Impacts of an agricultural development program for poor coconut producers in the Philippines: an approach using Panel Data and Propensity Score Matching Techniques". *Journal of Agricultural and Resource Economics* 32:534-557.

Sain, G. 2008. "Evaluación de los Componentes Tecnológicos de los Bonos Agroalimentarios: Antecedentes, Métodos, y Resultados". Final Report.

SIAP (Servicio de Información Agroalimentaria y Pesquera). 2010. Production Data. Retrieved October 20, 2010.

http://www.siap.gob.mx/index.php?option=com_content&view=article&id=10&Itemid15

Simtowe, F., and M. Zeller. 2006. "The Impact of Access to Credit on the Adoption of Hybrid Maize in Malawi: An empirical Test of an Agricultural Household Model under Credit Market Failure". MPRA Paper N.45.

Singh, I.J., L. Squire, and J. Strauss. 1986. *Agricultural household models: Extension, application and policy*. Baltimore, MD: Johns Hopkins University Press.

Skoufias, E. 2005. "PROGRESA and its Impacts on the Welfare of Rural Households in Mexico". IFPRI Research Report 139, Washington, D.C.:WB.

Solís, D., B.E. Bravo-Ureta, and R.E. Quiroga. 2009. "Technical Efficiency among Peasant Farmers Participating in Natural Resource Management Programs in Central America". *Journal of Agricultural Economics* 60(1):202–219.

Stecklov, G., P. Winters, J. Todd, and F. Regalia. 2007. “Unintended Effects of Poverty Programmes on Childbearing in Less Developed Countries: Experimental Evidence from Latin America”. *Population Studies* 61:125-140.

Stecklov, G., and A. Weinreb. 2010. “Improving the Quality of Data and Impact-Evaluation Studies in Developing Countries”. Impact-Evaluation Guidelines, Strategy Development Division, Technical Notes No. IDB-TN-123. Inter-American Development Bank, Washington, D.C.

Stiglitz, J.E. 1987. “Some Theoretical Aspects of Agricultural Policies”. *The World Bank Research Observer* 2:43-60.

Sunding, D and D. Zilberman. 2001. “The Agricultural Innovation Process”. In: B. Gardner and G. Rausser, editors. *Handbook of Agricultural Economics*, vol. 1A. Agricultural Production. Amsterdam, The Netherlands: Elsevier, North-Holland.

Uaiene, R.N. 2008. “Determinants of Agricultural Technical Efficiency and Technology Adoption in Mozambique”. ETD Collection for Perdue University. Dissertation.

World Bank. 2007. *Data for Impact Evaluation*. Doing Impact Evaluation N.6, Washington, D.C.: WB.

World Bank. 2010. LSMS-Integrated Surveys on Agriculture. In: Living Standards Measurement Study. Washington, D.C.: WB. www.worldbank.org/lsms