

MILLENNIUM CHALLENGE CORPORATION
PRINCIPLES *into* **PRACTICE**

Impact Evaluations of Agriculture Projects

By Katherine Farley, Senior Director, Agriculture and Land, Department of Compact Operations
Sarah Lucas, Senior Policy Advisor, Department of Policy and Evaluation
Jack Molyneaux, Director, Impact Evaluation, Department of Policy and Evaluation
Kristin Penn, Senior Policy Advisor, Department of Compact Operations

In collaboration with:

Emily Hogue, Technical Team Leader for Monitoring and Evaluation, Bureau for Food Security, USAID
Bureau for Food Security, USAID
Agriculture Practice Group, MCC
Monitoring and Evaluation Practice Group, MCC



MILLENNIUM
CHALLENGE CORPORATION
UNITED STATES OF AMERICA

MILLENNIUM CHALLENGE CORPORATION
PRINCIPLES *into* **PRACTICE**

The Millennium Challenge Corporation was founded with a focused mandate to reduce poverty through economic growth. MCC's model is based on a set of core principles essential for development to take place and for development assistance to be effective—good governance, country ownership, focus on results, and transparency.

The MCC Principles into Practice series offers a frank look at what it takes to make these principles operational. The experiences captured in this series will inform MCC's ongoing efforts to refine and strengthen its own model. In implementation of the U.S. Global Development Policy, which emphasizes many of the principles at the core of MCC's model, MCC hopes this series will allow others to benefit from and build on MCC's lessons.

The series also offers insights into MCC's experience with the technical and operational approaches it uses to support poverty reduction through economic growth. Consistent with MCC's focus on results, MCC is committed to learning from experience and applying lessons learned to its programs. Consistent with MCC's commitment to transparency, and in the spirit of ongoing collaboration with other development practitioners, MCC will share this learning publicly, even when it reflects challenges that MCC and partner countries have faced. The full Principles into Practice series is available at www.mcc.gov/principlesintopractice.

This paper reflects a collaborative effort between the Millennium Challenge Corporation and the U.S. Agency for International Development's Bureau of Food Security, both of which have made strong commitments to impact evaluation in the agriculture sector. The learning reflected in this paper directly contributes to Feed the Future, the U.S. Government's global hunger and food security initiative.

The lessons summarized in this paper emerged from a one-day roundtable, "Agriculture Impact Evaluations: Learning What Works," co-hosted by MCC and USAID in July 2011, and attended by a mix of agricultural development practitioners and impact evaluation experts. While the lessons in this paper are based primarily on MCC's experience with program and evaluation implementation, the analysis of these challenges and development of lessons were done in close partnership with colleagues from USAID.

In Principle: Impact Evaluations of Agriculture Projects

The U.S. Government is committed to learning what works to reduce poverty and increase food security through agriculture investments. There are many tools for learning from program experience. However, impact evaluations are central to a learning agenda that moves beyond business-as-usual assumptions about what works. There is no more rigorous tool than impact evaluation to answer difficult questions such as: Did we achieve the impact we intended? Did we make the right assumptions about how project interventions would translate into poverty reduction? What is the impact of our investment as compared to what would have happened without it? What impacts are directly attributable to this investment, and not to other factors? In other words, what can the investment really take credit for? Did the benefits of this investment outweigh the costs? What can we learn from what went right and what went wrong? When impact evaluations are rigorous and independent, and coupled with a commitment to transparency, they are a powerful tool for accountability, learning and making investments based on evidence about what works.

This paper describes why learning through impact evaluation matters and describes both challenges and lessons to getting it right in the agriculture sector. The lessons in this paper are based primarily on MCC's experience—with a focus on program and evaluation implementation rather than impact evaluation findings¹—though they reflect joint learning between MCC and USAID.

In Practice: Impact Evaluations of Agriculture Projects

MCC has made public commitments to rigorous learning through impact evaluation.

MCC's evaluation policy states:

“MCC's results framework reflects a commitment to technically rigorous, systematic and transparent methods of projecting, tracking and evaluating the impacts of its programs. Coupled with transparency, this approach is a cornerstone of MCC's commitment to accountability and learning. MCC's focus on results is motivated by some of the basic questions of aid effectiveness: Do the expected results of this program justify the allocation of scarce aid dollars? Has the investment achieved its goals? What can we learn from the experience to inform future programs and international best practice?”

“MCC is committed to making its evaluations as rigorous as warranted in order to understand the causal impacts of its programs on the expected outcomes and to assess cost effectiveness. Evaluations support two objectives derived from MCC's core principles: accountability and learning. Accountability refers

¹ A forthcoming paper will summarize project and evaluation lessons from the findings of MCC's first round of agriculture impact evaluations.

to MCC's and MCAs' obligations to report on their activities and attributable outcomes, accept responsibility for them and disclose these findings in a public and transparent manner. Learning refers to improving the understanding of the causal relationships between interventions and changes in poverty and incomes."

MCC has about a dozen impact evaluations underway in agriculture and rural development projects across its compact portfolio designed to measure changes in farming practices, agriculture productivity and sales and household income and consumption that are attributable to MCC investments. MCC also has almost 100 independent evaluations across other sectors, over a third of which are rigorous impact evaluations.

USAID has recently renewed its commitment to rigorous evaluation. In its evaluation policy released in 2011, Administrator Rajiv Shah explains the motivation for evaluation this way:

"In an increasingly complex operating environment, the discipline of development demands a strong practice and use of evaluation as a crucial tool to inform our global development efforts, and to enable us to make hard choices based on the best available evidence. I have great expectations for the work of USAID. I expect us to succeed in some of our efforts, and to fall short in others. I expect a strong evaluation function and feedback loop that enables us to be accountable in both cases, and to learn from each so that we can make continuous improvements. We can do this only with evidence and data to inform our decisions, and with unprecedented transparency about what we have learned and where."

USAID, in the context of Feed the Future, plans over 20 impact evaluations over the next five years to contribute to the body of knowledge on food security. Data and findings from those evaluations will be incorporated into the programming cycle as they become available to improve the design and management of interventions in the agriculture and nutrition sectors. In the context of Feed the Future, impact evaluations have a two-fold purpose: (1) to strengthen accountability to stakeholders and (2) to foster learning that will improve the effectiveness of all U.S. Government-funded food security programs. This paper offers lessons that can strengthen Feed the Future's ability to meet these two aims.

Both MCC's and USAID's evaluation policies highlight principles of independence, learning, accountability, and transparency. Both are based on the understanding that development work is filled with challenges, and resources are scarce. Not all interventions will achieve their intended aims, but it is essential to know what has and has not worked so that future investment decisions can be based on this evidence. It is in this spirit that MCC and USAID are working together to improve accountability and learning through impact evaluation in the agriculture sector. Both agencies are still learning how to strike an effective balance between program and impact evaluation implementation. The five lessons identified in this paper are an important and substantive start.

Box 1: Defining Evaluations

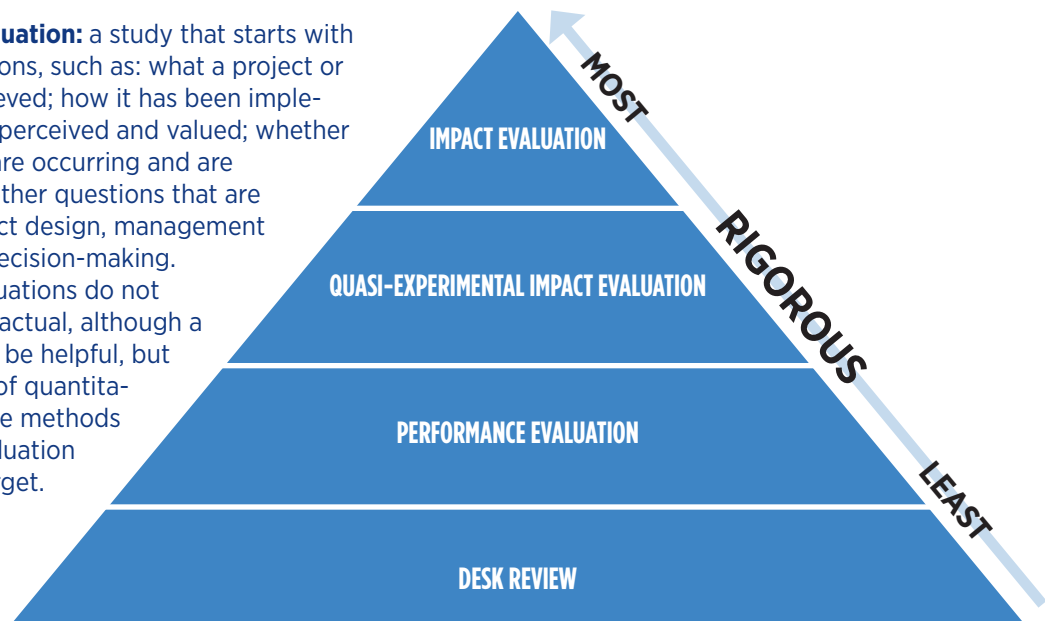
Both MCC and USAID make a distinction between impact and performance evaluations in the following way.

Impact Evaluation: an independent study that measures the changes in income and/or other program objectives that are attributable to a defined intervention. Impact evaluations require a credible and rigorously defined counterfactual that estimates what would have happened to the beneficiaries absent the project. Estimated impacts, when contrasted with total related costs, provide an assessment of the intervention’s cost-effectiveness, what is commonly referred to as the project’s “bang for the buck”. Randomized control trials (that are referred to as RCTs or experimental designs), with appropriate use of mixed methods, are the impact evaluation methodology that generally provides the greatest opportunity for learning and for structuring a strong counterfactual. However, when an RCT is not feasible or desirable, quasi-experimental impact evaluations that use methodologies such as propensity score matching and regression discontinuity design, combined with mixed methods, are other means that facilitate learning and allow for attribution of impact.

Impact evaluations serve two key purposes – accountability and learning. Accountability compares costs and impacts on final outcomes such as income and poverty that are attributed to MCC investments. Learning tests development hypotheses and explores how well or poorly a particular development approach works. Learning relates to better understanding the causal chains expected to link MCC investments to these income changes. For example, trained farmers should: 1) learn why improved soil management practices increase yields; 2) adopt these practices; 3) improve their yields; 4) increase farm income; and 5) ultimately raise their household incomes. Learning requires understanding *how and why* these causal linkages do or don’t happen and is essential to testing the assumptions behind program design.

Impact evaluations are an essential tool for learning and for accountability, but are not the right tool for every project. They should be used selectively, with a special focus on where the potential for learning is greatest. (See Box 11.)

Performance Evaluation: a study that starts with descriptive questions, such as: what a project or program has achieved; how it has been implemented; how it is perceived and valued; whether expected results are occurring and are sustainable; and other questions that are pertinent to project design, management and operational decision-making. Performance evaluations do not require a counterfactual, although a control group can be helpful, but can use a variety of quantitative and qualitative methods to answer the evaluation questions they target.



Challenges Putting Impact Evaluation into Practice in the Agriculture Sector

Impact evaluations in the agriculture sector are growing in number in this exciting era of aid effectiveness and accountability. However, most, if not all, evaluations that have employed experimental methods have been conducted on a relatively small-scale as pilot projects, with a primary goal of learning whether certain interventions work.² By planning and conducting numerous impact evaluations of large-scale, donor-funded agriculture projects, MCC and USAID are putting evaluation methodologies to the test. Likewise, implementing large-scale projects with a limited amount of time and budget puts a great deal of pressure on implementers to demonstrate quick results, adapt as necessary and sometimes work with a complex array of actors. At this scale, the challenges inherent to both project implementation and impact evaluation are magnified. Managing these challenges requires very close, early and ongoing coordination between project implementers and impact evaluators.

Several factors make evaluation in the agriculture sector especially challenging. These factors fall into two broad categories – how agriculture projects can pose challenges for evaluation and how evaluation approaches can cause challenges for the implementation of agriculture projects. Combined, these factors can make implementers, partner countries and sector specialists hesitant about rigorous impact evaluation. This paper is based on a frank acknowledgement that these challenges exist and a desire to develop practical solutions for managing impact evaluations in this context.

How agriculture projects pose challenges for evaluation: Several of the factors below are specific to agriculture, such as crop cycles and weather variations. Others, such as self-selection, are relevant to a variety of sectors, but can be especially magnified in agriculture.

★ **Crop cycles and seasonality:** The seasonality of agriculture creates two challenges related to timing. The first is the timing relative to crop cycles. Agriculture production cycles pose strict windows for when training and related activities, including fielding evaluation surveys, can occur. If these windows are missed because of delays in program implementation or evaluation planning, a full crop cycle can be lost. This has implications for the program's ability to achieve and to evaluate impact. The second is the expected time between an intervention and the expected results. Agriculture projects often require several crop cycles to yield benefit, as farmers become proficient in new techniques, expand their application and learn from one season to the next. In addition, with some projects, the difference in outcomes between treatment and control farmers should substantially grow over time. This creates challenges for evaluation done very soon after programs complete because they can underestimate impact. Randomized rollout methods (see Box 2) particularly are susceptible to both these risks. If, by design, they allow for only one or two crop cycles between treatment and control groups, and this gap is further compressed by delays,

² See a summary of farmer field school evaluations in Waddington, et al (2010) and Vandenberg (2004).

the project impacts can be underestimated, especially if they were expected to accrue slowly over time (or over-estimated if, for example, the indicator of interest is income and the evaluation years capture short-term price increases in targeted products).

- ★ **Context variables:** Agriculture production is highly affected by weather, mainly temperature and rainfall. These conditions or variables change from year to year. In theory, a valid control group should, on average, face the same weather shocks as the treatment group. But agriculture interventions may influence the magnitude of weather effects on outcomes³, implying that weather shocks can influence impact estimates. And while changes in prices for agriculture products affect everyone, not just the treatment or control group, they can change the returns of the intervention that the treatment group receives. Therefore, like weather effects, price changes can influence impact estimates.
- ★ **Spillover effects:** Spillover or demonstration effects, such as when people outside the primary targeted beneficiary groups adopt techniques supported by a given intervention, are sometimes expected in agriculture projects and often a desired outcome in design. If a project indirectly affects the control group outcomes even though the control group itself did not participate in the program, this will bias estimates of impact. Spillover effects in agriculture can be quite large and are often an explicit component of program logic, particularly in the case of agriculture projects where technology, such as planting in rows and weeding instead of broadcast seeding, is easily transferred.
- ★ **Implementation changes:** Even when project design is set, agriculture implementation approaches can evolve significantly over the course of a project in response to changing market conditions or more detailed implementation planning. While adjusting implementation approaches may seem to make interventions more effective or better target participants, it can challenge the validity of the evaluations and reduce the potential for learning what really works.
- ★ **Sequencing of interventions in integrated projects:** For projects that include a variety of interventions, such as farmer training, access to credit, and infrastructure investments, sequencing is important for achieving the desired outcomes. For example, if the rehabilitation of irrigation canals is delayed, newly trained farmers will not yet have access to water resources and therefore can lose a planting season. Again, if a randomized rollout is used, these delays for the first round of interventions will shrink the differences between treatment and control groups, thereby limiting the evaluation's learning potential.

Box 2: Randomized Rollout Impact Evaluations

Randomized rollout is a type of randomized control trial in which control groups are treated after a lag. For example, in the case of farmer training, the first round of treatment farmers is compared to a control group of farmers that receive training at a later date. The first round control farmers become second round treatment farmers. The usefulness of a randomized rollout evaluation depends on there being sufficient time between the first and second round of trainings to detect the program impacts.

³ For example, high-yield seed varieties may be less drought resistant; conversely, adoption of more efficient irrigation systems may increase a farmer's resilience to droughts.

- ★ **National-level interventions:** Projects that target agriculture-related national-level policy reform or institutional change, such as new water laws, credit regulations, improved phytosanitary and inspection services, and improved linkages to export channels for targeted value chains, are not generally conducive to identifying a within-country counterfactual.
- ★ **Self-selection:** An important constraint for the evaluation of agriculture projects is that they often include some degree of self-selection. This implies that only certain types of farmers may choose to participate in a given project, requiring the evaluation to isolate the impacts of a given intervention from the influence of unobservable characteristics, such as personal motivation, of the individuals who chose to participate.

How evaluation approaches can cause challenges for the implementation of agriculture projects:

- ★ **Implications for scope:** Requiring a control group that is comparable to the target population can reduce the overall reach and scope of projects. If the demand for project interventions exceeds the scope allowed by the project budget, timeline or willingness to invest in an unproven intervention, it is easier to construct a control group. However, in cases where donors and implementers are confident of the intervention's cost-effectiveness, and have the ability to reach all potential beneficiaries, there are inherent tradeoffs between project scope and learning potential.
- ★ **Implications for flexibility:** Adherence to an impact evaluation methodology may limit implementers' ability to adapt implementation approaches in response to changing conditions or new information. While impact evaluations can be designed to be remarkably robust to planned ranges of implementer adaptation, an intervention with no planned structure poses a serious evaluation risk. It also may pose a significant investment risk for a donor.
- ★ **Implementers often make strong assumptions about what works and what doesn't work:** A recent study by a team of Ghanaian and American evaluators⁴ offers a good example of evidence overriding strong assumptions about what works. The team set out to understand why many Ghanaian farmers do not invest in inputs, such as fertilizer or hybrid seeds, despite evidence that they can yield significant increases in crop income. Focus groups with Ghanaian farmers suggested that lack of capital and high risk from weather or disease were key obstacles to investment. The study aimed to understand which mattered more. The surprise was that addressing the credit constraint, which is universally seen as important, through cash grants (akin to starter packs) had little effect on farmer investment. Instead, rainfall index insurance through grants or for purchase was associated with a very large increase in investment. In this case, the less common and less costly intervention yielded a much better impact. Development professionals, donors, implementers, and partner countries may think they know what works. But there is strikingly little evidence

⁴ See Karlan, Osei, Osei-Akato and Udry (2012). "Agricultural Decisions after Relaxing Credit and Risk Constraints."

available to test prevailing assumptions. Consider the fact that while thousands of farmer training projects have taken place over the last decade, a recent World Bank study identified only three impact evaluations using rigorous experimental designs in farmer training anywhere in the world between 2000 and 2009⁵.

- ★ **Political sensitivities:** Randomized selection of communities or beneficiaries may mean that the project may not work with *all* of the most promising beneficiaries because some will be part of the control group. This may be politically challenging for program implementers or country counterparts who have to explain why some potential beneficiaries will not be able to participate in the program.
- ★ **Unclear value or costs:** Implementers are sometimes given little opportunity to see the value of impact evaluations. Focused primarily on how evaluation methodologies might restrict scope or operational flexibility, agriculture implementers (and agriculture specialists in donor organizations) get little information regarding the design of impact evaluations or few opportunities to provide input into what they aim to learn. In addition, the level of effort required for the implementer to coordinate with the evaluator during the evaluation design and implementation can be underestimated.
- ★ **Conflicting incentives:** Implementers often face incentives to meet targets, regardless of how this affects evaluation methodology, while evaluators have incentives to adhere to evaluation approaches, regardless of how this may limit the flexibility of implementers to adapt to changing conditions and new information.

⁵ IEG (Independent Evaluation Group). 2011. Impact Evaluations in Agriculture: An Assessment of the Evidence. Washington, DC: World Bank.

Five Lessons on Putting Impact Evaluation in Agriculture into Practice

While these challenges are real and can be difficult to manage, they should not prevent the pursuit of rigorous impact evaluation in the agriculture sector. Given the critical role of agriculture for development, and tightening development budgets globally, it is essential that the development community deepen its understanding about what approaches work best to reach desired outcomes in a cost-effective way.

Many of the following lessons identify approaches that facilitate use of a counterfactual, which compare the changes that occur both with and without a given intervention through use of treatment and control groups. These solutions also adhere to evaluation methodologies that maintain the integrity of counterfactuals so that impact evaluations can identify attributable impacts.

The lessons are designed to support agriculture donors, partner countries, implementers, and evaluators in striking a healthy balance between achieving impact, measuring results and learning what works in agriculture investments. While these challenges and lessons emerge from discussion about impact evaluation in the agriculture sector, many are broadly applicable to other development sectors as well.

Box 3: The Five Lessons

Lesson 1: Define early the program logic and objectives of the evaluation, and how to integrate the two.

The most important first step—both for successful implementation and evaluation—is to have a clear picture of what a program aims to achieve, and how planned interventions are expected to lead to that outcome.

Lesson 2: Engage early and communicate often. Coordinated planning and ongoing communication are the essential ingredients for minimizing and managing tradeoffs between implementation approaches and evaluation methodologies.

Lesson 3: Foster joint ownership by aligning incentives. Everyone involved must feel ownership over both the program implementation and evaluation so incentives must be aligned for donors/sponsors, partner countries, project implementers, and evaluators.

Lesson 4: Match evaluation methodology and program design. The most rigorous method for measuring attributable project impacts is through randomized control trials (RCTs), but when they are not feasible, there are other rigorous methods for evaluation to consider.

Lesson 5: Focus on long-term impacts but be prepared to show early results. Impact evaluations are often not carried out for a year or two after project completion. In planning to be accountable for progress and communicate early results, one must take care not to undermine the ability to measure long-term impact.

Lesson 1:

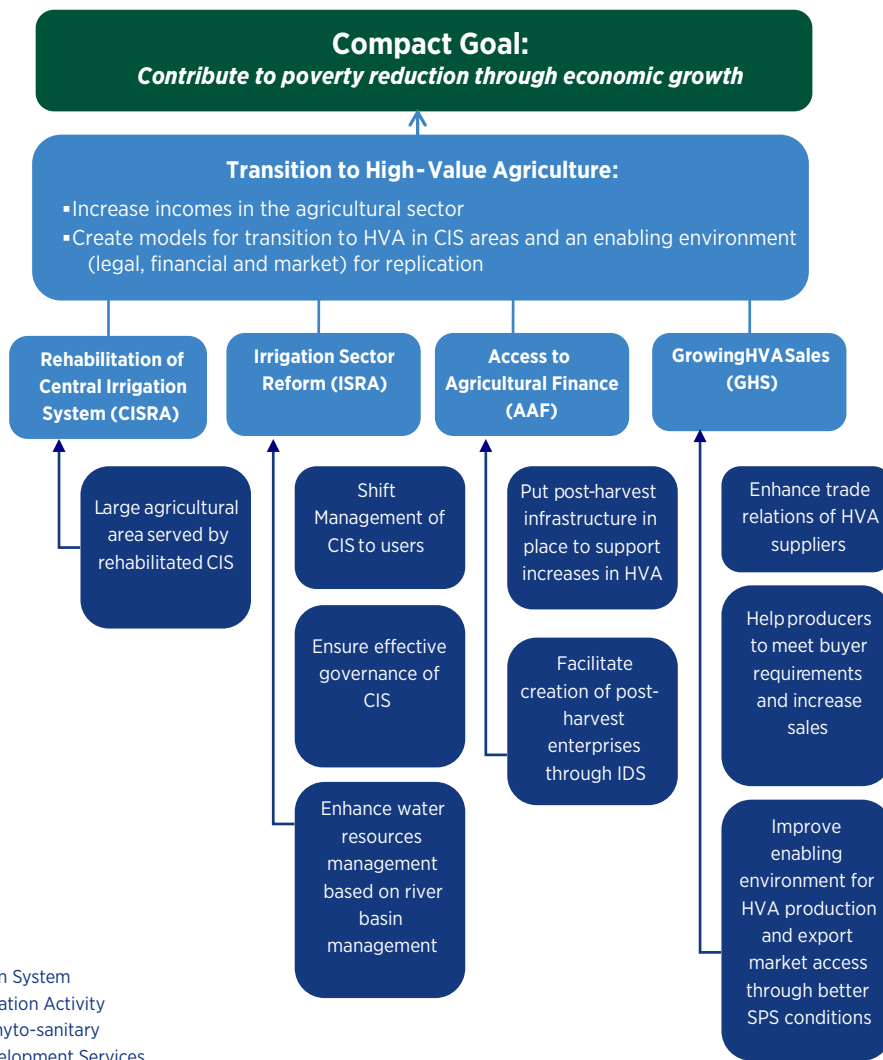
Define early the program logic and objectives of the evaluation, and how to integrate the two

The most important first step—both for successful implementation and evaluation—is to have a clear picture of what a program aims to achieve and how planned interventions are expected to lead to that outcome. This program logic, or theory of change, is the starting point for designing both the implementation approach and its evaluation. See Box 4 for a sample program logic. Once this is in place, the following considerations can help integrate the implementation and evaluation so that both are manageable and can maximize impact and learning.

- ★ **Focus the evaluation on the most important learning goals.** If the central program objective is to raise the incomes of poor farmers, an evaluation must assess those impacts. Alternatively, if a project’s essential objective includes increasing the value-added of locally produced export crops, an evaluation must be designed to measure that outcome. Evaluations of a broad range of potential benefits (poverty alleviation, enhanced exports, reduced environmental deterioration, etc.) are risky: the more outcomes one tries to measure, the greater the risk grows of false positive and negative results. This may be warranted for experimental projects but typically only at a small scale.
- ★ **Define early the stable components of the project that are essential to the project’s theory of change.** To integrate program and evaluation planning and maximize learning, it is important to distinguish early the essential components that must be defined up front from those that can be developed as part of project implementation. In agriculture projects, an essential component often includes concerted efforts to change the behavior of large numbers of farmers. If additional specifics are essential to the project, they should be specified in advance as well, such as demonstration plots and targeted training of “lead farmers” intended to trigger a demonstration effect on other farmers. Others may need to be developed as part of the early stages of the project, such as specific value chain selection and identification of the farming methods to be improved. A project budget can help discipline the selection of essential components versus elements that can be added later if time and budget allow.
- ★ **Ensure the causal pathway is clear and reflect it in the evaluation.** It is critical to ensure that the evaluations are well-articulated for testing development hypotheses and the causal pathway is well-defined. For instance, assuming that increased training and adoption of a particular technology will necessarily lead to increased income does not allow for learning about the “why.” Evaluation designs should consider all the factors related to adoption that will get households to increased income, such as improved productivity and increased sales. When considering evaluation approaches, sector experts and implementers may want to jump straight-away to the final intended

Box 4: Program Logic for MCC’s Transition to High Value Agriculture (THVA) Project in Moldova

Moldova’s THVA project is a set of four reinforcing and integrated activities that, when implemented together, aim to address the key constraints facing Moldovan producers: lack of reliable water, lack of financing, lack of access to markets and technologies, and lack of know-how. Development of the program logic by the MCC and Moldovan teams during project development was a critical early step in project design, especially in understanding the interrelated nature of the four project activities. In order to understand the overall impact on incomes in the agriculture sector, MCC has several evaluations planned for this project, including an RCT for the farming training component of the Growing High Value Sales Activity (GHS) complemented by a process evaluation and value chain network analysis and a comparison group design for the Rehabilitation of Central Irrigation System Activity (CISRA) and the Irrigation Sector Reform Activity (ISRA). The evaluation approach for the Access to Agricultural Finance Activity (AAF) is currently under review but a requirement under the compact.



impact, but evaluations that do not thoroughly track changes in the causal pathway will not produce useful findings to foster learning on the approach tested.

- ★ **Determine what components of the program logic are evaluable.** Specifically, identify components of the program logic for which a counterfactual can be credibly established and measured. Some important program elements may be difficult to evaluate. For example, projects that support national level policy change or awareness-raising campaigns are not conducive to identifying a counterfactual. In some cases, one might not choose to evaluate all pieces of a multi-faceted intervention, but look for opportunities to do smaller-scale evaluations within a larger project to increase learning on the effectiveness of one intervention relative to another. This might be an evaluation designed to measure what intervention is most effective at fostering farmer adoption of new technologies rather than to measure ultimate impacts like income gains.
- ★ **Explicitly assess how non-evaluable components can affect outcomes.** While some project components may not be easily evaluated, their expected contribution to overall project impact should be understood and captured in the evaluation. This requires clarity about the interrelationships between the evaluated and non-evaluated components. For example, policy interventions are often supported for multiple, justifiable reasons. First, they are supported simply based on compelling logical arguments: e.g. improved phytosanitary standards may be a political imperative or may be broadly accepted as worth their cost through their effects on improved consumer confidence, even without country-specific evidence. While it is hard to conduct an impact evaluation on a policy change that affects the whole population (because there is no counterfactual), learning about how these factors might affect impacts is important. This learning can take place through interviews with implementers and other stakeholders and, in some cases, impact evaluations can include explicit modeling to capture the interactions between program components. This is essential to properly measure impacts of an evaluated project component, since the interacted effects may be independent, conditional or multiplicative.
- ★ **Keep in mind the challenges magnified in agriculture projects.** As discussed earlier, there are some challenges for impact evaluation that are magnified in agriculture projects. Evaluators and implementers must consider these factors when designing evaluations. For example:
 - * **Crop cycles and seasonality:** Two distinct factors to consider are timing—when the agricultural year begins and ends— and periodicity—the time between when an intervention takes place and when the resulting outcome is expected. These factors should directly inform when in each year the baseline and endline surveys are collected, and how long after an intervention occurs the endline surveys should be conducted.
 - * **Self-selection:** For impact evaluations to minimize self-selection bias, they require both ensuring adequate sample sizes and using appropriate evaluation methods to account for observable and unobservable differences between participants and non-participants.

- * **Spillover effects:** While spillover effects may reduce the observable impact of the project by blurring differences between treatment and control groups, good evaluation models can capture such secondary impacts if planned appropriately from the outset.⁶
- * **Weather variations:** It is important to ensure that evaluation sampling errors properly account for effects of location-specific micro-climates. Otherwise, weather-related variations can be incorrectly interpreted as impact signals.
- ★ **Think carefully about scale and complexity.** Many agriculture development projects are multi-faceted, large scale projects with multiple integrated activities that can be hard both to implement and to evaluate. To achieve its desired impact, a project must be implementable. To achieve an impact evaluation's accountability objectives, a project needs clear and consistent program logic and objectives. And to achieve its learning objectives, a project should be replicable. Unmanageably complex projects compromise all of these objectives. The added management requirements of an impact evaluation (respecting control and treatment distinctions, adhering to the proposed objectives and program logic, etc.) threaten both the project and the evaluation of a management-challenged project. If the program logic supports a multi-faceted package of interventions, consider limiting the size of the intervention, and building in the capacity to scale-up through subsequent projects or with other partners.

Lesson 2:

Engage early and communicate often

Coordinated planning and ongoing communication are the essential ingredients for minimizing and managing tradeoffs between implementation approaches and evaluation methodologies. Building a culture of partnership, mutual respect and open dialogue among parties through early engagement and ongoing communication is a critical factor of successful impact evaluations. (See Box 5.) This includes:

- ★ **Early engagement during design:** Early engagement lays the foundation for establishing a strong professional relationship and fosters a deeper understanding of the implementation approach as well as the impact evaluation design methodology. Impact evaluations go more smoothly when evaluators involve implementers in key aspects of methodology design and continue this engagement throughout implementation. This may often require unpacking the broad program logic, distinguishing components that can and cannot be rigorously evaluated, clearly assessing what can and cannot be learned this way, and ensuring that there are benefits to both the implementers and the evaluators. Ideally, the implementers have included ideas for how to integrate implementation planning and impact evaluation design in their proposal.

⁶ See Conley & Udry (2010) for a good example of explicit modeling of farmer training spillover effects.

- ★ **Harnessing implementers' knowledge:** Implementers are tremendous resources of project, sector and local knowledge that are essential contributions to grounding impact evaluation methodologies in the reality of local circumstances. Including implementers in the process of developing survey instruments and, where possible, other aspects of methodology design, enables impact evaluators to harness the knowledge of implementers and forms a stronger and enduring partnership. Evaluation teams, in order to be most effective, should include technical specialists in agriculture to inform evaluation approaches, including the selection of treatment and control groups.
- ★ **Training:** The more implementers and evaluators know about and understand each others' methods, the more likely they are to collaborate in managing tradeoffs during implementation. Donors should support (and even require) basic training of implementers on evaluation methodologies and learning objectives so they recognize how implementation changes might affect the integrity of the evaluations. Likewise, training of evaluation data collectors on the specifics of implementation can increase the quality of data derived from surveys and thereby strengthen the evaluation.
- ★ **Anticipate and plan for change:** It is common for agriculture project approaches to change and evolve in response to changing local and market conditions or to monitoring information gathered during implementation. It is important to acknowledge that these changes may occur and explicitly plan to discuss their impact on evaluation methodologies before decisions are made about changing implementation approaches. Ideally, these changes would be made as a result of learning from impact evaluation interim findings and at least on the basis of monitoring data.
- ★ **Explicitly structure ongoing communication:** Knowledge is power in terms of finding mutually agreeable solutions to challenges. Though communication can be time-consuming, the integrity of the project implementation as well as the impact evaluation design will benefit from ongoing communication. Expectations and structure for this communication can be formally built into contract agreements and implementation plans, such as through quarterly or semi-annual workshops in which project implementers, evaluators, partner countries, and donors meet to discuss project design challenges or changes and how they might affect evaluation methodologies. All participants should be encouraged to raise concerns early and often so that solutions can be found that best manage tradeoffs between implementation flexibility and evaluation validity.
- ★ **Be present:** It helps for the impact evaluation team to have representation in the country to respond rapidly to changes in the project, visit the project sites, understand the implementation approach first-hand, and communicate directly with implementers about their challenges and concerns. While a full-time presence may not be feasible or cost-effective, a part-time presence with mutually-agreed frequency can be very useful.

Box 5: Early Engagement and Ongoing Communication (or not)—Experiences in Armenia and Burkina Faso Regarding the Design and Implementation of Impact Evaluations**The projects:**

Under MCC's compact in Armenia, the Irrigated Agriculture Project aimed to increase incomes by rehabilitating and constructing irrigation infrastructure, providing farmer training, expanding access to credit, strengthening post-harvest processing and marketing, and strengthening management capacity of local and national water supply entities.

Under MCC's compact in Burkina Faso, the Agriculture Development Project also provides support for water management and irrigation and for on-farm production and related activities throughout the agricultural value chain.

The evaluations:

The impact evaluations for the Armenia project were designed to measure the impact of on-farm water management and higher-value agriculture training on farming practices, agriculture productivity and the income of rural farming households as well as the impact of rehabilitating irrigation infrastructure (tertiary canals) on value-added agriculture and rural household income.

In Burkina Faso, the impact evaluation is designed similarly to measure the impact of training and extension services activities in newly-irrigated areas on farmers adopting new technologies and techniques and the impact of training/technical assistance on farmer yields, sales and incomes. Despite the similarity in the projects and evaluations, the experiences were very different.

The lessons:

In these cases, MCC has learned the importance of good coordination and communication between program implementers and evaluators. In the Armenia case, the evaluation team worked closely with the implementing organization from the very beginning, establishing a relationship based on mutual respect, good communication and common goals. As a result of this close communication,

when challenges arose, they were handled quickly and effectively in a collegial manner. Their close coordination allowed the teams to avoid problems seen in other countries of too many independent surveys, interviews and data collection by different teams interfering with the actual training or making the farmers nervous. The Armenia implementing team and evaluating team worked together to create survey questions and ensure that they weren't duplicating efforts and overburdening farmers.

In contrast, the impact evaluation implementation process in Burkina Faso has suffered due to limited early engagement of the impact evaluators with the implementers and constrained communications over time. In this case, the evaluators were interested in measuring the impact of projects on farmers' income, while the implementers preferred to be accountable to outputs, such as the number of farmers trained. There was a fundamental difference of opinion between the teams as to whether the evaluating team and associated surveys would be used for *monitoring* or *evaluating*. In addition, the implementers rejected the evaluators' preference of having randomly selected project participants, wanting to choose instead participant villages that they knew would be successful. As project implementation progressed, the implementers instituted a series of project design changes without always informing the impact evaluation team. Consequently, communication between the parties eroded and the most rigorous impact evaluation methodology was compromised.



A tertiary irrigation canal in Armenia

Box 6: Changing Implementation Affects Evaluation—Lessons from Nicaragua

The project:

For MCC's compact in Nicaragua, the Rural Business Development Project aimed to increase beneficiary incomes by providing business development services, disseminating market information, developing improved production techniques, and providing financial assistance to small and medium farms and agribusinesses to transition to higher profit activities.

The evaluation:

The impact evaluation is designed to measure the impact of receiving Rural Business Development services on beneficiary household income and other measures of well-being. The evaluation uses randomized sequencing of beneficiaries, in which half of the geographic regions were randomly selected to receive services before 2009 and the other half were randomly selected to receive services after 2009.

The lessons:

This case is an illustration of how decisions to modify project implementation approaches can have implications for evaluation methodology. In their work with bean producers, the project implementers and MCA-Nicaragua decided to change the implementation approach after the first rounds of assistance, primarily to increase sustainability. In the first rounds, the implementer delivered a package of inputs and technical assistance directly to each farmer. In the later rounds, the inputs were delivered to a cooperative that, in turn, gave these inputs to farmers on credit, repayable upon delivery of farmers' harvest. This approach allowed the cooperative to maintain some revolving funds, one step toward improving the sustainability of the approach. This change, however, actually made it more difficult to evaluate impact. The impact evaluation was designed to assess the difference in impacts between those receiving assistance for different periods of time. This calls for a comparison between the early and later rounds of farmers, but since the implementation approach changed, it's more difficult to accurately make this kind of comparison.



Rural Business Development Project beneficiaries

Lesson 3:

Foster joint ownership by aligning incentives

To strike a healthy balance between achieving impact, measuring results and learning what works in agriculture investments, everyone involved must feel ownership over both the program implementation and evaluation. This means incentives must be aligned for donors/sponsors, partner countries, project implementers, and evaluators. As a starting point, the onus is on the donor, in collaboration with partner countries, to clearly set expectations for both implementers and evaluators and hold them accountable to these expectations. Aligning incentives also entails the following:

- ★ **Plan together:** The starting point to align incentives is the planning process. Whenever possible, design of both the project and evaluation plans should include both implementers and evaluators. This takes time, so the project timelines should reflect this process.
- ★ **Create clarity and incentives through contracts:** All contractors—both implementers and evaluators—look to requests for proposals (RFPs) and contract language to set expectations and create incentives for performance. Therefore, RFPs for both implementers and evaluators should include clear expectations for joint planning and ongoing collaboration to manage tradeoffs between implementation approaches and evaluation methodology. Project implementer RFPs should include as much information as possible about potential evaluation methodologies, and can invite bidders to articulate how these methods may affect project design in order to demonstrate an understanding and commitment to rigorous evaluation whenever possible or appropriate. RFPs could also request input from bidders on what they would like to learn from the impact evaluation so that it can be tailored to address their learning needs. In order to best address and anticipate the challenges in impact evaluations in the agriculture sector, signing of implementer and evaluator contracts should coincide to maximize opportunities for joint planning. Ideally, the implementer, in the original proposal, is able to clearly articulate how rigorous evaluation will be integrated with project design and implementation and this demonstrated ability and understanding of the importance of impact evaluations will be reflected in the evaluation criteria in the RFP. See Boxes 7 and 8 for some challenges in aligning incentives and Box 9 for potential solutions.
- ★ **Get the timing right:** Joint planning and other steps to align interests in incentives depends on implementers and evaluators being on similar timelines. Contracts for evaluators should be established before or at the same time as the contracts for implementers so that their planning and startup activities coincide.

Box 7: Getting the Incentives Wrong in Honduras**The project:**

MCC's Rural Development Project in Honduras aimed to increase farmer incomes by providing training in the production and marketing of high-value horticultural crops, improving access to credit, constructing or improving feeder roads to connect farms to market, and providing grants to support adapting global technological advances in agriculture to local conditions.

The evaluation:

The evaluation for the farmer training component was designed to measure the impact on incomes of training in increased productivity and business skills. It was designed as a randomized rollout, in which the first round of treatment villages would receive program interventions 18 months before the second round (control) villages.

The lessons:

This case is an illustration of how misaligned incentives, poor donor planning and limited coordination between implementers and evaluators can hinder learning and accountability. MCC and its Honduran MCA counterparts had not yet identified an evaluation methodology or contracted an evaluator before signing a contract with the implementer. Therefore, the implementer's contract did not include information about the impact evaluation, incentives to design project plans to support impact evaluation or requirements to work with the evaluator to carefully identify treatment and control groups. The evaluator started after the implementer was already working with farmers across the country. This, together with changing criteria for selecting treatment and control sites and reduced time to work with treatment farmers, resulted in a control group that may not be sufficiently comparable to serve as a counterfactual. In addition, the incentives in its contract actually pushed the implementer away from adhering to the evaluation methodology. For example, the implementer was required to meet annual targets with regard to the number of farmers receiving training and other interventions. When reaching this many farmers in the planned area proved difficult, the implementer was motivated to seek farmers in other areas, some of whom were part of the control group for the impact evaluation. This situation created tradeoffs between the goals of responding to and reaching monitoring and evaluation targets, for which the implementer had significant incentives in the contract, and adhering to an evaluation methodology, about which the contract was silent. This created the risk that the impact evaluation would be unable to detect the attributable impacts of the program, a situation that disadvantages all stakeholders, including the implementer.

Experiences like this have helped MCC learn the value of incorporating impact evaluation approaches into implementation plans and implementer contracts, and the necessity of anticipating and planning for managing these tradeoffs.

- ★ **Help implementers find value in impact evaluations:** In addition to creating incentives for implementers to adhere to evaluation methodologies, implementers can benefit from impact evaluations in tangible ways. Donors and partner countries committed to impact evaluation can help implementers recognize value in:
 - * **Baseline surveys:** Baseline surveys can be designed to gather information that is valuable for project design and targeting beneficiaries.
 - * **Informing the learning agenda:** Implementers can be given the opportunity to define questions that the evaluation answers and to recommend interventions for which experimentation is built into the evaluations, like testing assumptions about standard approaches.
 - * **Capturing demonstration effects:** Implementers are often interested in a project's spillover or demonstration effects. Spillover effects can be included in impact evaluations, typically through explicit statistical modeling both of the targeted first-learners as well as the expected spillover transmission effects. The most informative of these evaluations have helped to understand what types of first-learners effectively transmit learning and through what networks.

Box 8: **Trying to Get the Incentives Right in Moldova**

The project:

MCC's compact in Moldova includes a Transition to High-Value Agriculture (THVA) Project that includes a Growing High Value Sales Activity (GHS). This GHS activity aims to increase farm incomes by improving market linkages and providing training in HVA production. This activity is designed to benefit from other THVA project activities, including irrigation rehabilitation and policy and institutional reform of the irrigation sector.

The evaluation:

The impact evaluation for this activity is designed to measure the impact of farmer training on adoption of new practices, production, sales, farm income, and how these affects vary by value chain.

The lessons:

Attempting to learn from past experiences (including Honduras), MCC tried to work with Moldovan partners and implementers to align incentives starting with the RFP for the GHS activity. The RFP clearly stated there would be an impact evaluation of some of the components of the GHS activity, outlined anticipated data requirements and required implementer cooperation and involvement in design and implementation of the impact evaluation. Despite this attempt to get the incentives right and inform the implementer in advance about impact evaluation requirements, the lack of specific instructions about the degree of this cooperation in the RFP and subsequent contract led to misunderstandings at the impact evaluation design phase in terms of the implementer's role. It took several months and long negotiations for MCC and the project implementers to reach a common understanding. Had the implementers been required to think more about the design of the impact evaluation during the project planning phases, their incentives would have been more in line with those of MCC and the evaluator.

- * **Showcasing results:** The power of impact evaluations to identify results that are directly and credibly attributable to projects provides the opportunity for implementers' projects to be recognized as a success and source of learning.
- * **Measuring long-term results:** The benefits of agriculture projects often accrue over many years. Measuring results (by any method) only upon project completion can underestimate the overall impact of the project. Therefore, if donors make provisions for evaluating impact farther out than just two years, implementers have the opportunity to learn about and be recognized for the longer-term impacts of their projects.
- * **Credibility:** Evaluation methodologies can be used to build awareness, buy-in and credibility of projects. For example, transparent lottery selection of participants can build confidence that a project is avoiding political influence and corruption in selecting participants.
- ★ **Be honest about costs and level of effort:** Building in time for joint planning, designing implementation approaches to accommodate evaluation methodologies and staffing evaluation teams with sector expertise all have implications for costs and level of effort. Impact evaluations themselves can be costly. Donors should be honest about these implications and make conscious decisions about the value for money of the learning that evaluations will yield. Where there is a lack of concrete evidence about what works, there is no more rigorous or valuable tool than impact evaluations to help build an evidence base. But the tool is only useful and cost-effective under certain circumstances (see Box 12). Projects to be evaluated should be selected deliberately, with an eye toward the value of the potential learning, buy-in of key stakeholders of the learning agenda and budgets.

Box 9: Aligning Incentives through Requests for Proposals and Contracts

Successful integration of implementation and evaluation priorities requires aligning incentives for project implementers and evaluators. MCC, its partner countries and USAID are still learning how to best do this in requests for proposals (RFPs) and contracts. Below are some possible approaches. *In the case of MCC, partner countries lead program implementation. Partner country governments establish accountable entities to manage program implementation. These MCAs are signatories to program implementation contracts, while MCC is signatory to impact evaluation contracts.

Request for Proposals (RFPs)	
Implementers	Evaluators
<p>RFPs should provide:</p> <ul style="list-style-type: none"> • Clear expectation that/if an impact evaluation with treatment and control groups will be used and basic information about possible evaluation methodologies (explored during project design/development phase) • Information about specific ways in which the implementer could be expected to cooperate with the evaluation, e.g. generate longer lists of potential beneficiaries to establish control and treatment groups • Evaluation criteria for the proposal that clearly outlines the priority of impact evaluation and expertise and experience with successful impact evaluations <p>RFPs should require that bidders:</p> <ul style="list-style-type: none"> • Articulate how possible evaluation methodologies will affect implementation and be successfully integrated into program design and implementation • Identify how they will obtain evaluation expertise for their team • Propose questions for learning to be addressed by the evaluation 	<p>RFPs should provide:</p> <ul style="list-style-type: none"> • Basic information about possible implementation approach and evaluation approaches considered during project design and development <p>RFPs should require that bidders:</p> <ul style="list-style-type: none"> • Include specific sector expertise in key personnel • Identify how they will obtain local sector expertise for their team • Include a plan for field presence to facilitate coordination with implementers • Include joint planning with implementers early and often in the budget and implementation approach • Provide description of past experiences effectively coordinating with implementers and working in the sector on a similar evaluation
<p>RFPs for both implementers and evaluators should provide:</p> <ul style="list-style-type: none"> • Expectation of integrated planning and specific measures for ongoing coordination for project implementation and evaluation • A preliminary project schedule that includes a juncture for joint planning prior to program implementation <p>RFPs for both implementers and evaluators should require:</p> <ul style="list-style-type: none"> • Proposal for how to integrate implementation and evaluation planning and execution <p>Donors and partner countries should:</p> <ul style="list-style-type: none"> • Include a person with impact evaluation experience on the technical evaluation panel for reviewing proposals from possible implementers and vice versa 	
Contracts	
Implementers	Evaluators
<ul style="list-style-type: none"> • Assign payments to deliverables in the contract that are related to coordination with the impact evaluation (like an annual statement from the implementer on the evaluation) • Consider ways to tie a payment or bonus for the implementer to the specific measures of successful cooperation with the evaluation 	<ul style="list-style-type: none"> • Include incentives for quick start-up of activities, adequate staffing of technical specialists and accurate and timely identification of treatment and control groups
<ul style="list-style-type: none"> • Both program implementer and evaluator contracts require an integrated program and evaluation plan before detailed contractors' delivery and payment schedules can be finalized. If these are not detailed in an integrated proposal, this should be developed in the initial phase of program implementation and agreed to by the funder(s), the implementer and the evaluator before actual implementation commences. 	

Lesson 4:

Match evaluation methodology and program design

The most rigorous method for measuring attributable project impacts and for learning is through randomized control trials (RCTs). Because RCTs identify similar groups of individuals (control groups) that will and will not be exposed to project interventions, evaluators can compare the groups to measure their impacts, potentially over a long period of time. This use of a statistically identical control group creates the greatest opportunity for learning what works and for measuring project impacts, including those that accrue over time. Therefore, it is often useful to start by exploring whether an RCT is a good fit for a proposed project.

However, challenges to this approach are common and have led to a range of alternative valid evaluation methods. One such challenge can arise when a project must target an entire eligible population, making it hard to find a comparable control group. This problem can yield a simple, valid solution if the program budget is inadequate to serve the same population: allow those willing to participate in a random assignment lottery to sign up for it. With over-subscription, an RCT may still be valid. But if there is under-subscription, randomizing some potential participants out of the program may not be justifiable. Alternative methods for identifying adequately similar populations should be considered.

In addition, to ensure that the results of RCTs are valid, it is important to know about the risks of spill-overs (i.e. benefits from the implementation that affect the control population), spontaneous adoption of program methods (i.e. adoption uninfluenced by the program) or other participant self-selection into or out of the implemented population. These risks generally can be accommodated into valid experimental designs, but the risks must inform the appropriate evaluation methods. This requires that implementers and evaluators have significant amounts of information about all potential beneficiaries (both treatment and control groups) from the outset, which can be challenging particularly when project design is under tight timelines. That said, the information about potential participants needed for defining treatment and control groups is often the same information needed to plan implementation approaches. Implementers can be asked to identify and gather information about a larger number of potential participants who they plan to work with, and it can be used for both implementation and evaluation planning, if the two are coordinated. The extra time and cost to gather this additional information can be considered part of the cost of the evaluation. If these marginal information costs are driven by the evaluation needs, these costs should be considered part of the costs of learning, rather than part of the program. If program budget constraints are binding, there should be explicit consideration of whether the learning benefits warrant their costs.

When it is not acceptable or possible to exclude control groups from program exposure over an extended evaluation period, another kind of RCT—randomized rollout (see Box 2) — may be an alternative. It has the benefit of allowing for a clear counterfactual over a

limited evaluation period if done well, while eventually treating the initial controls. MCC has taken this approach in a number of cases when other RCT approaches were deemed unfeasible or undesirable by implementers or partner countries (see Box 11). However, MCC is learning that this methodology has significant vulnerabilities in farmer training programs. Randomized rollouts generally offer project interventions to the control group a year or two after the first treatment group. This allows for a very short period to compare the differences in change between the two groups, which can be further limited by delays in project start-up. This approach risks underestimating the impact of project interventions because it simply does not allow enough time for benefits to accrue. This is particularly limiting for agriculture projects with a gestational period of multiple years before the primary benefits can be observed. And even projects with short gestations often will benefit from payback periods that accrue over many years.

Even when a RCT is not a suitable methodology, there are other ways or methodologies to measure project impact that may be appropriate. These quasi-experimental approaches⁷ have well-defined limitations but, when appropriate, are useful and practical alternatives that provide critical learning opportunities if randomization is not possible.

In quasi-experimental impact evaluations, instead of creating treatment and control groups by random assignment ex ante or prior to the beginning of the program, these groups can be created ex post or once the program has begun or possibly even ended. This is done by using observed socio-cultural, economic, ecological, and geographic characteristics to ensure that the comparison groups are sufficiently similar, at least in observable characteristics, so that it can be argued that the observed impact is due to the program as opposed to confounding factors. Several quasi-experimental methods may be used to generate a control group. Such methods include parametric and semi-parametric regression-type analysis, non-parametric matching-type analysis or a combination of both. In some cases, such methods may be mixed with difference-in-difference techniques to enhance the precision of the impact measures.

In a quasi-experimental evaluation, the program is non-randomly placed across units (individuals, households, villages, etc.). The various non-experimental assessment methods used to solve the problem of a missing counterfactual can be classified in two groups. The first group assumes that non-observable characteristics of beneficiary units and control group units are independent of the decision to participate in the program. Single-difference methods and double-difference methods are in this group. The second group comprised of matching (including propensity-scores) methods, discontinuity design methods and instrumental variables address the possibility that participation in the program is non-random, even after controlling for observable characteristics.

One strategy may be to compare the changes—instead of the level—of a given indicator between the group of beneficiaries and the control group. Assuming that the change in the indicator in the control group is a good representation of what the change in the

⁷ See, for example discussions in Duflo, Glennerster and Kremer (2007), and Gertler et al., (2011).

indicator would have been among the beneficiaries, this difference in differences estimate may provide a valid way to neutralize the *external selection bias* and hence provide an unbiased assessment of the program's effect. The idea behind this approach is to compare a group of beneficiary units with a group of control units before and after the intervention.

Although quasi-experimental methods (including propensity score matching) may offer a viable alternative when randomization is not possible, there may be conditions under which they do not represent a valid counterfactual. Both matching and regression methodologies are susceptible to selection biases. These biases are most problematic if program participation decisions will, on average, select higher (or lower) performers than the evaluator should expect, compared with the control group. If highly motivated candidates, for example, are over-represented in the treatment group, the evaluator cannot generally distinguish the motivation effects on outcomes from the program effects.

This problem extends beyond the self-selection of highly motivated individuals into programs. It must also be considered when program administrators select communities for farmer training programs or implementers choose suitable program participants. If participant selection puts more (or fewer) high performers in the treatment group than in the control group, then to measure the effect of the program, the evaluation must have a credible approach to distinguish the effect of the selection.

In light of these factors, it is essential that evaluators, implementers, partner countries, and donors convene as early as possible to identify what they seek to achieve and to learn and what evaluation methodology is the best fit, given that a cost-effective solution typically requires a range of adaptations by all parties. (In some cases, an impact evaluation might not be feasible or desirable. See Box 12.) Such discussions are also valuable for designing impact evaluations to address questions that donors, partner countries or implementers are particularly interested in learning about.

Box 10: Managing the Politics of Evaluation in Mongolia and Georgia**The project:**

As part of MCC's five-year, \$285 million compact with Mongolia,* local governments are leasing pastureland to groups of herders. The project is supplying these herder groups with wells and materials to build fences and animal shelters. Herders also receive training on livestock management, rangeland productivity, business skills, and improved dairy production. This leasing program operates in five areas and aims to improve herder-group productivity and incomes through better range-management practices and the reduction of the livestock level to allow for the sustainable use of the range. The program is considered a pilot effort which, if successful, could be replicated in other parts of Mongolia. A rigorous evaluation is essential for this learning.

The evaluation:

The question of how to choose who would receive land leases, training and infrastructure was crucial during project design. A lottery was chosen to promote fairness and because it allows for an effective statistical evaluation of the leasing program's impact. By randomly assigning a sufficiently large number of the target population into control and treatment groups, each group will have similar characteristics, whether or not they are identified, and the observed results can then be attributed to the leasing program.

The lessons:

The Mongolia case illustrates how an evaluation methodology can enhance project implementation if the two are well coordinated. On the day of the lottery, almost 120 herder groups gathered in a public auditorium, understanding that only half of them would be selected at random to receive leases, training and infrastructure. The lottery numbers were drawn as the audience watched. Half of the audience went home disappointed, but they accepted the results. Many congratulated the winners. This acceptance is a testament to close collaboration between the implementer and evaluator, effective organization and implementation of the lottery and several months of advance work to explain the selection system to the herders. This work included managing expectations so everyone who entered the auditorium knew that they—just like everyone else in attendance—had a 50-50 chance of their number being called.

This lesson is similar to what MCC learned in the Georgia compact's Agriculture Development Activity (ADA). The ADA provided technical assistance and targeted matching grants to farmers and businesses in critical value chains in order to support the development of Georgia's agriculture and agribusiness sectors. A lottery was also used to select projects to receive grants from a pool of projects that had already been screened and reviewed as acceptable in meeting project selection criteria. The very public (and televised) lottery helped address issues of transparency in a project where demand for grants far exceeded the supply of project funds to support such grants. While initially very resistant, MCC's Georgian counterparts over time grew to appreciate the value of this approach for both impact evaluation and addressing issues of excess demand versus the supply/ability to fund.

* For more information, see "An Innovative Approach to Selecting Program Beneficiaries: Lottery Day in Eastern Mongolia" at <http://www.mcc.gov/documents/press/headway-2012002104301-mongolia-lottery.pdf>

Box 11: The Challenges of Randomized Rollout—Lessons from Morocco**The project:**

MCC's Fruit Tree Productivity Project in Morocco aims to increase beneficiary income by supporting improvements to increase irrigation efficiency and productivity of olive and date trees, rehabilitate existing olive trees and expand production of olive and date trees, and move small farms from high water-use, low-value cereal grains to low water-use, high-value and drought-resistant commercial fruit tree species. The project is extremely complex and ambitious, working in four value chains in both irrigated and rain-fed areas in 25 provinces.

The evaluation:

The impact evaluation for this project is designed to measure the impact on farmer incomes of training and technical assistance for the rehabilitation of olive trees in rain-fed areas, and the impact of technical assistance on farming practices and on the quality of olive oil processing. The evaluation methodology was designed to compare changes in income of farmers who received the training (treatment group) to farmers who did not receive training but possessed similar characteristics (control group). However, problems arose in the execution of the methodology. In order to identify both treatment and control groups (organized by communities called perimeters), both groups needed to be identified to permit the random selection of the treatment group. The impact evaluation consultant began its engagement after the project implementers had already identified most, if not all, of the rain-fed perimeters that would benefit from the rehabilitation and begun outreach efforts to these communities, raising their expectations. Going back to these communities and refusing training after raising expectations created a difficult choice.

The compromise reached during the design stage was to pursue a “rolling” evaluation: the perimeters scheduled for rehabilitation in year 3 would serve as the control group for those rehabilitated in year 1; and the perimeters scheduled for rehabilitation in year 5 would serve as the control group for the year 3 perimeters. However, significant delays in project implementation, including in the construction of irrigation rehabilitation, thwarted this approach.

The lessons:

In retrospect, this problem could have been avoided had the discussions about the evaluation begun earlier in project appraisal and design. It might have been possible to identify a universe of perimeters eligible for project intervention and then randomly select treatment and control perimeters before project personnel engaged with the farmers. The use of a transparent, public lottery system to select beneficiaries also could have been possible.

Box 12: When an impact evaluation Is and Isn't the Right Answer

Impact evaluations are an essential tool for learning and for accountability but are not the right tool for every project. They should be used selectively, with a special focus on where the potential for learning is greatest. The decision about whether, and how, to conduct an impact evaluation for a given investment depends on several factors:

What is the potential for learning? For programs where the assumptions underlying the program logic are based on limited evidence, there is a strong case for an impact evaluation. A rigorous impact evaluation tests assumptions about a project's effectiveness and contributes substantially to future decision-making as well as the global evidence base.

Can a counterfactual be identified? An impact evaluation requires a control group that accurately represents what would happen without the project. This is easier to do for investments that target individual communities or participants, like agriculture training, education or land titling projects. In these cases, evaluators might be able to identify comparison groups that either do not receive project interventions or that receive them at a later date than a first group. For projects that are national in scope, such as major policy reform, there may be no easily identified control group. In these cases, independent evaluators might use alternative methods of evaluation. These may estimate benefits based on changes in specific indicators that are identified in the program logic and design as the key links between project activities and intended outcomes.

Can the evaluation approach be worked into project implementation? Some impact evaluation methodologies, especially randomized control trials, must be built into implementation design. For some projects, this would require implementation modifications that are just too complicated, time consuming, or unfeasible.

Does strong stakeholder commitment exist? Identifying a control group and ensuring adherence to an impact evaluation design requires significant commitment and collaboration by sector staff, program implementers and evaluators, both within donor institutions and among partner countries. Program designers, implementers and evaluators must work together to understand and define the program logic, estimate how long expected impacts are likely to take to accrue and identify what is most important to learn about how the program works.

Are resources available? Some evaluation methodologies are quite costly. A rigorous study of a five-year agricultural training project might cost several million dollars. MCC and USAID both estimate that the average impact evaluation for agriculture projects is between \$1 million and \$2 million. Both agencies are committed to evaluating their projects but also recognize that resources are scarce, and the value of the information has to be considered with regard to the cost of the evaluation.

Lesson 5:

Focus on long-term impacts but be prepared to show early results

Impact evaluations are generally intended to learn about long-term program benefits. Consequently, post implementation surveys should be designed to reflect the lags in impacts implied by the program logic and estimate the impacts of completed projects, requiring that they are often carried out for multiple years after project completion. These may not inform attributable project impacts immediately upon project completion and often must wait for several crop cycles to detect real, sustained impacts of farmers adopting new technologies. That said, stakeholders naturally demand results as soon as projects end. Donors, partner countries and implementers should plan accordingly and anticipate the need to be accountable for progress and communicate early results. But in so doing, they must take care not to undermine their ability to measure long-term impact. Some tips for managing these concerns include:

- ★ **Honesty about what you'll know and when:** It is essential to message to key stakeholders that donors and implementers may not have data about impacts attributable to project interventions immediately upon project completion. With good monitoring and evaluation (M&E) systems, information about outputs should be available almost immediately, but impact evaluations generally yield information about outcomes, such as income or consumer expenditure improvements, that may accrue several years after project completion. Even these should identify conditions under which benefits may be expected to accrue over the longer term and how they may be efficiently assessed. Making evaluation design documents that include a timeline or research protocols available publically is a good way to demonstrate transparency about data and information while managing expectations.
- ★ **Monitoring and evaluation plans:** M&E plans that track progress against targets for indicators that are drivers of project outcomes help show whether projects are on track and communicate with constituents about interim results. Donors and partner countries should identify a number of indicators that are essential for monitoring project progress and communicating interim results, noting that indicators for monitoring can be quite different from those for communicating, and taking care not to overburden implementers with an unwieldy number of indicators for which they must collect regular data.
- ★ **Pilot projects:** Some projects lend themselves to scaling or tranches. In such cases, results from short pilot projects can help satisfy demand for immediate results.
- ★ **Conviction about taking the time to evaluate:** Even with excellent messaging about what you will know when, there will be inevitable pressure to report outcome information quickly. This may tempt donors, partner countries or implementers to push evaluators to conduct endline surveys earlier than planned. While this will yield some data more quickly, it will probably not give project interventions sufficient time

to take hold and create the change donors and partner countries set out to achieve. Therefore, by rushing to measure results, one risks under or over-estimating project impact.

- ★ **Ability to monitor and evaluate some projects for longer periods:** For projects with the highest potential for rich learning about long-term impacts, it is useful for donors to have the funds, flexibility and authority to continue monitoring key indicators and to evaluate several years after project completion. This may be particularly important for agriculture development projects because of the pace of change. For example, farmers may be slow to adopt changes or may start at a small scale and then expand as approaches prove promising. Planting of new tree crops can take years before the maximum yields are seen, and can often have a two to three year gestation period before any yield is seen. Even for quicker crops it can take multiple crop cycles to see the full effects of interventions.

Looking Ahead

The U.S. Government is committed to implementing impact evaluations of development programs for purposes of strengthening accountability to stakeholders and to foster learning that will improve the effectiveness of future investments. Development projects and investments such as those supporting agriculture and food security represent an opportunity to deliver on this commitment.

A focus on learning: MCC has been a leader of rigorous impact evaluation since its inception in 2004 and now has an extensive pipeline of impact evaluations associated with its investments in agriculture and rural development. These evaluations aim to measure project impacts on farming practices, agricultural productivity, the income of farming households, and other measures of well-being. In addition, they will offer insight into the effectiveness of various program designs and implementation approaches and inform future project design and compact implementation. MCC expects to publish a first set of agriculture-related impact evaluations in October 2012.

Feed the Future has accountability and learning as part of its core principles. To that end and similar to MCC's approach, a robust monitoring and evaluation agenda is being implemented. This includes the development of a learning agenda that focuses on six themes or key areas of learning, such as agriculture productivity, R&D, improved market access, dietary diversity and nutrition, resilience of vulnerable populations, and women's empowerment and gender equality. The Feed the Future learning agenda includes a set of strategic questions under each theme for which Feed the Future and its implementers intend to produce evidence, findings and answers through impact evaluations and other methods, such as performance evaluations and policy analysis. The findings and answers to these strategic questions will be disseminated broadly through web platforms, annual progress reports and knowledge sharing events called Global Learning and Evidence Exchanges (GLEEs).

Courage to be transparent: In the spirit of learning and accountability, both MCC and USAID will publish the findings of their evaluations, even when they reveal that the agencies and partner countries have fallen short of program targets or when interventions have been less cost-effective than expected. There is risk associated with this because it will be easier than ever to point to development investments that have not worked well. However, with independent and transparent evaluation in place, it will also be easier to have confidence in claims of successful impact.

Shared interest and capacity to learn: There is much to learn about which interventions have the greatest impact in a given context, which interventions are most cost effective and what combination and or sequence of interventions/investments have the greatest impact on the objectives of improving agricultural growth, reducing poverty and decreasing malnutrition. Together, USAID and MCC will launch and finalize a minimum of 40 impact evaluations of food-security related investments over the next five years and contribute to the body of knowledge on food security to improve the design and management of interventions in the agriculture and nutrition sectors.

Technical Resources

This paper complements MCC's Monitoring and Evaluation Policy and USAID's Evaluation Policy and Feed the Future M&E Guidance Series, Volume 4: Impact Evaluation and Feed the Future. It builds on MCC's *Principles into Practice Focus on Results* paper that offers 10 lessons associated with putting MCC's results focus into practice, including lessons associated with conducting experimental impact evaluations. It also builds on the body of knowledge of best practices in impact evaluations. For further reading on impact evaluations in the agriculture sector, please consider:

- ★ 3ie's Systematic Reviews: <http://www.3ieimpact.org/systematicreviews/>
- ★ Campbell Collaboration Library of Systematic Reviews: <http://www.campbellcollaboration.org/library.php>
- ★ Systematic reviews found on the DFID R4D website: <http://www.dfid.gov.uk/r4d/systematicreviews.aspx#Agriculture%20and%20Rural%20Development>
- ★ Conley, T; and C. Udry, 2010. "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, American Economic Association, vol. 100(1), pages 35-69, March. [An un-gated earlier version is also available: Conley, T; and C. Udry, 2000. "Learning About a New Technology: Pineapple in Ghana," Working Papers 817, Economic Growth Center, Yale University, revised May 2004. Web version: <http://ideas.repec.org/p/egc/wpaper/817.html>
- ★ Duflo, E.; R. Glennerster; and M. Kremer, 2007. "Using Randomization in Development Economics Research: A Toolkit". Discussion Paper No. 6059 January 2007, Centre for Economic Policy Research, 90-98 Goswell Rd, London EC1V 7RR, UK. Web version: <http://www.aniket.co.uk/teaching/devt2009/duflo2006.pdf>
- ★ Feder, Murgai & Quizon(2004) "Sending Farmers Back to School: The Impact of Farmer Field Schools in Indonesia" *Review of Agricultural Economics* 26 (1), pp. 45-62.
- ★ Gertler, P.; S. Martinez; P. Premand; L. Rawlings; C. Vermeesch, 2011. *Impact Evaluation in Practice*. The World Bank, Washington, DC. Permanent URL: <http://go.worldbank.org/0FUDDHCD20>
- ★ J-PAL/CEGA Agriculture Technology Adoption Initiative: <http://atai-research.org/>
- ★ Karlan, D.; R. Osei; I. Osei-Akato; and C. Udry, 2012. "Agricultural Decisions after Relaxing Credit and Risk Constraints", working paper – Web version: <http://www.econ.yale.edu/~cru2/pdf/eui.pdf>.
- ★ USAID's Development Experience Clearinghouse: <http://dec.usaid.gov/index.cfm>
- ★ USDA Food and Nutrition Service, "Nutrition Education: Principles of Sound Impact Evaluation. <http://www.fns.usda.gov/ora/menu/published/NutritionEducation/Files/EvaluationPrinciples.pdf>

- ★ Waddington, H., Snilstveit, B., White, H., and Anderson, J. (2010) The impact of agricultural extension services: Study protocol, 3ie Synthetic Reviews SR009, International Initiative for Impact Evaluation, New Delhi. http://www.3ieimpact.org/admin/pdfs_synthetic/009%20Protocol.pdf
- ★ Van den Berg, H. (2004), 'IPM Farmer Field Schools: A synthesis of 25 impact evaluations,' Prepared for the Global IPM Facility, Wageningen University, the Netherlands. <ftp://ftp.fao.org/docrep/fao/006/ad487E/ad487E00.pdf>
- ★ World Bank (2007) World Development Report 2008: Agriculture for Development, World Bank, Washington, D.C. http://siteresources.worldbank.org/INTWDR2008/Resources/WDR_00_book.pdf