A guide for monitoring and evaluating the impact of vegetable research and development at AVRDC – The World Vegetable Center

Pepijn Schreinemachers

Victor Afari-Sefa

Greg Luther

Annelie Öberg

AVRDC The World Vegetable Center

AVRDC - The World Vegetable Center is an international nonprofit institute committed to alleviating poverty and malnutrition through the increased production and consumption of nutritious, health-promoting vegetables. The Center's improved vegetable lines and technologies help small-scale farmers boost yields, increase their incomes, and safely grow nutritious vegetables for their families and communities.

Published by
AVRDC – The World Vegetable Center
P.O. Box 42 Shanhua, Tainan 74199
Taiwan
T +886 6 583 7801
F +886 6 583 0009
E info@worldveg.org

avrdc.org

AVRDC Publication: 14-780

Summary

Results-based management has become more important as donor agencies are increasingly held responsible to show results for the money they are investing into research and development (R&D), and also because scientists themselves want to demonstrate the impact of their work. This guide offers a framework for institutionalizing project monitoring and outcome evaluation in project planning and implementation at AVRDC – The World Vegetable Center. It was written primarily for the Center's staff and management and its project partners in developing countries.

Projects at the Center are positioned at different points along the research-development continuum, which affects the way project objectives are defined and how monitoring and evaluation are to be conducted. The framework defines three strategies. First, projects need to make their impact pathways explicit by using a Logical Framework Approach (LFA) and Theory of Change (ToC), as these methods are largely complementary. The LFA gives a more systematized and linear view of how inputs and activities will lead to outcomes, while the ToC highlights the complexities of system change and interactions between project components.

Second, projects oriented toward development must have an explicit scaling strategy to show how the project will disseminate an innovation within a well-defined target population. It is also important to have a strategy to identify and address possible dissemination constraints.

Third, outcome evaluation must be able to attribute observed changes in outcome indicators to project interventions using scientifically sound methodologies. It must be kept in mind that higher-level outcome indicators such as health and nutrition are affected by a larger number of confounding factors than lower-level indicators such as crop output and farm profitability and are therefore more difficult to prove. Outcome indicators must be quantifiable, and any measure of impact needs to be statistically tested. For attributing impact, the use of control groups is essential. Experimental methods are preferred over quasi- and non-experimental methods, as the random selection of study subjects together with the random allocation of project benefits gives the highest level of internal and external validity. However, qualitative methods of evaluation are also important, as these can provide insight on the conditions that constrain or promote impact. Institutionalizing methodological rigor in project planning, monitoring and evaluation will require close working relationships between project scientists, the Center's monitoring and evaluation (M&E) team, and external collaborators. This document presents the current state of thinking and will be updated regularly as and when necessary to reflect dynamic changes in the Center's R&D efforts.

[intentionally blank]

Table of contents

[intentionally blank]

1 Introduction

The mission of AVRDC – The World Vegetable Center is "to overcome malnutrition and poverty and facilitate good health for both the rural and urban poor by increasing the production, quality, consumption and utilization of nutritious and health-promoting vegetables" (AVRDC - The World Vegetable Center, 2010). The mission defines three development objectives and three strategies to accomplish them. But how do we know how much progress we are making in accomplishing the Center's mission?

Until fairly recently, we did not ask this question explicitly. Like other research and development organizations, monitoring and evaluation at AVRDC was based on various ad-hoc activities but was not guided by an explicit strategy to monitor performance or evaluate outcomes using scientific rigor. This has changed as donors and other stakeholders as well as we ourselves are increasingly interested to demonstrate the impact of what we do, and to learn from successes as well as failures. As a decision-making tool, monitoring and evaluation (M&E) can improve the transparency, accountability and effectiveness of projects and policies and has become an indispensable component of most projects.

This document outlines a strategic framework for monitoring progress and evaluating outcomes and impacts by providing guidelines on how to integrate robust monitoring and evaluation protocols into projects. It was written primarily for AVRDC staff and management and the Center's project partners in developing countries, but also to share with donors, colleagues in academia and other research centers, project beneficiaries and the general public.[1]

The framework can be seen as a toolbox containing available methods and approaches for use in project planning and management. The framework can help researchers select the appropriate tools for monitoring progress in project implementation, for ensuring that project outputs are taken up by target populations and for scaling-out the uptake in non-intervention areas, and for quantifying the outcomes and impact of the project interventions. We note that the framework was developed for project management and does not apply to the monitoring and evaluation of the Center as a whole.

The strategic framework is divided into three sets of strategies, each discussed in a separate section:

> **Strategy 1** aims to ensure the highest standards of project management through the continuous monitoring of project performance. It describes tools such as problem tree and objective tree analysis to formulate problem statements and project objectives, and then describes the use of a logical framework model to monitor how a project transforms inputs and activities into outputs and outcomes.
>
> **Strategy 2** describes our strategy for optimizing the uptake and use of project outputs in project intervention areas, and scaling-out strategies to promote the uptake of Center's interventions and capacity building projects in non-intervention areas. It deals with technology dissemination, public-private partnerships and how to identify and address adoption constraints.
>
> **Strategy 3** describes project designs and methodologies for ex-post project evaluation in terms of attributing outcomes and impacts to project interventions. It discusses the strengths and weaknesses of various methods such as the difference-in-differences method, matching techniques and randomized controlled trials. After describing each of these strategies, the document concludes with suggested guidelines and directives for implementing the M&E strategy at the Center.

[1] This framework is specific to the work of AVRDC - The World Vegetable Center. For a more general discussion we refer to frameworks published by various aid agencies (ADB 1998, Norad 1999, SIDA 2003, The World Bank 2012, UNDP 2009). Several recent papers also give concrete guidelines on best practices in impact assessment (e.g. de Janvry et al 2011 and Walker et al. 2008).

2 Optimizing project management through monitoring and evaluation

2.1 Role of monitoring and evaluation

Project monitoring is about tracking and documenting the continuous flow of inputs, outputs and possible project results (Table 1). It is done on a day-to-day basis by a project manager to observe whether a project is on course to delivering its expected results and to detect possible bottlenecks in project implementation. Project evaluation, on the other hand, is concerned with uptake and use of project results by stakeholders and target populations (e.g. poor smallholder farm households). The uptake of project results is a voluntary decision by stakeholders and therefore outside the direct control of a project manager. Evaluation concerns with whether the project's objectives were met. Figure 1 gives an overview of M&E systems, showing most of the issues to be covered in this framework.

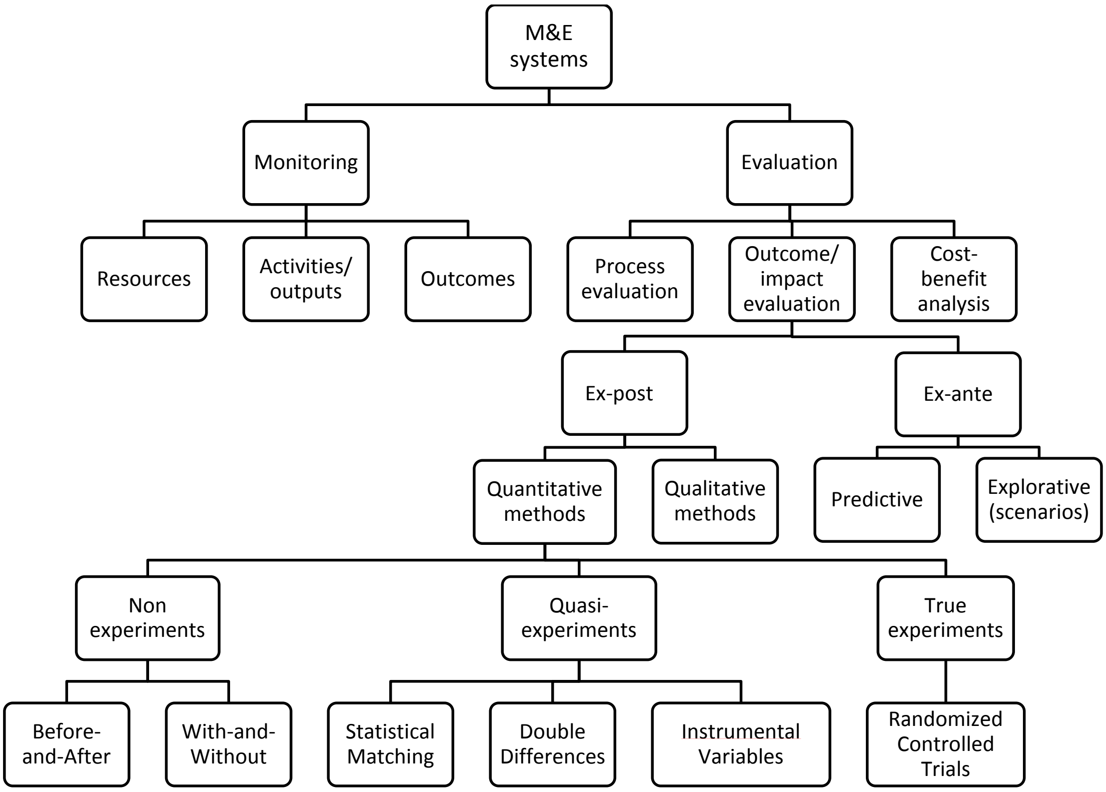

Figure 1 Monitoring and evaluation systems

Table 1 Monitoring versus evaluation

Monitoring	Evaluation
Continuous: day-to-day	Periodic: important milestones
Documents progress	In-depth analysis of achievements
Focuses on inputs and outputs	Focuses on outcomes and impacts
Alerts managers to problems	Provides managers with strategy and policy options
Self-assessment	Internal or external

The work of AVRDC comprises both research- and development-oriented projects. The distinction between the two is relevant because it affects the way projects are defined in terms of problem statements, objectives, outputs and outcomes, and the M&E approaches required. A development project aims to change the conditions or behavior of a group of people, called the target population. Progress towards this end should be observable within the lifespan of a project. A research project, on the other hand, typically aims to fill a knowledge gap or to develop a method or technology; although the results can contribute to development, such impacts are usually not observable within a project's lifespan.

Most projects at AVRDC combine research and development objectives (Figure 2). As a result, most projects have short as well as long-term objectives. Typically, most projects focus more on research in the first years of the implementation—designing, developing or adapting technologies—and more on development in the final year or years. The short-term outcomes of such projects may be limited, particularly since most projects only have a period of three or four years, yet long-term impacts might be substantial.

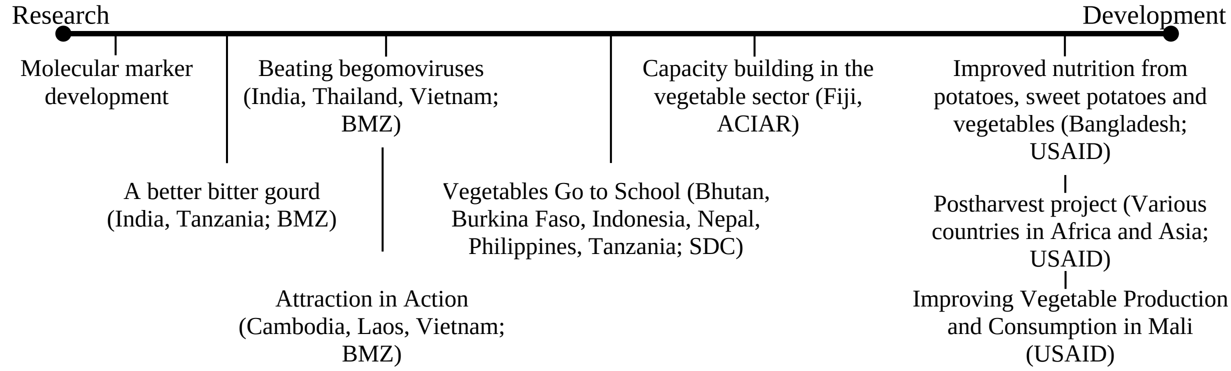

Figure 2 The research-development continuum, with a selection of ongoing projects to illustrate the Center's span of activities.

2.2 Designing projects that can make impact

The first step in project design is to identify a problem that needs to be solved or abated. In a research project, problem identification is based on identifying knowledge gaps from previous research or identified development problems. In a development project, it is ideally based on the needs of a target population such as poor smallholder farm families.

Identifying the current needs of farm households and other stakeholders is essential to ensure project outputs will be used and that the project will contribute to development. Care must be given that people's needs are not assumed for them—for instance, drawn from the Millennium Development Goals (MDGs)—but truly represent the actual development priorities of the target population.

Some projects deal with this by using Research for Development (R4D) platforms, Innovation Platforms or Learning Alliances. These are networks of individuals with different backgrounds and interests who come together to diagnose problems and to identify opportunities to achieve their goals. Network members, representing various stakeholders, might have different problems and different goals. Apart from farmers, network members may include representatives from input suppliers, vegetable seed producers, government officers, output traders and local researchers. Regular network meetings serve to develop a common understanding of the challenges and promote collaboration. The network facilitator has an important role to sustain the network, to mediate between opposing interests, and to ensure that all members have a voice in the meeting. The participation of stakeholders in diagnosing problems and finding solutions creates a sense of ownership over the project and raises interest in the project outputs.

Participatory appraisal (PA) tools can be used to create dialogue between networks, project team members and stakeholders to identify needs and find potential solutions. PA tools are suitable for capturing differences in opinions and attitudes among stakeholders and for understanding spatial and temporal heterogeneity in resource constraints. Table 2 lists a selection of PA methods that can be used to conduct a situational analysis to identify the development needs of a community.

For various reasons, development-oriented projects should not be based solely on the needs identified by local communities themselves. First, local communities might have special interests that conflict with other people's interests, or are not representative for the population as a whole. Second, stakeholders are likely to focus on their short-term needs and might not be fully aware of trends or future events that pose a threat or provide an opportunity for development. For instance, local farmers might not fully appreciate the impact of climate change, resource degradation, the long-term detrimental effects of pesticides on their health, or upcoming changes in laws and regulations—all of which might have a dramatic effect on their future livelihoods. Third, the project team must further filter stakeholder needs based on what is realistic to achieve given the available expertise within the team, the available resources, and the expectations and interest of a donor willing to provide these resources. Farmers are likely to identify low output prices and high input prices as a problem that they would like to address. Yet, prices are driven by the supply and demand in regional, national or even global markets, and a project usually cannot influence prices directly; however, projects can intervene to reduce on-farm losses, improve efficiency, or add value to farm produce and thereby improve farm incomes.

The result of such an analysis is a clearly defined **problem statement**, which provides information about the situation that needs changing, who it affects and what the presumed causes are. A typical research project would then test the validity of these causes through hypothesis testing, whereas a development project tries to formulate interventions that could make the change actually happen. We note, however, that as the success of a development project hinges on a correct identification of the causes of a problem, there are important synergies between research and development within a project and many donors explicitly recognize this.

Table 2 Selection of participatory appraisal tools to identify the development needs of local communities

Tool	Main strengths (+) / challenges (−)
DISCUSSION TOOLS	
1. **Focus group discussions:** Dialogue session with 6-10 participants to understand people' opinions, beliefs, attitudes. Facilitator prepares a few questions to stimulate discussion.	+ Can create many new ideas − It can be difficult for the group to stay focused on the topic and to give a voice to all participants.
2. **Semi-structured interviews:** Usually a face-to-face interview with farmers or other stakeholders in which interviewer has prepared a list of topics to talk about (but preferably no questions).	+ Is a flexible and more natural way of communication than a structured interview − Because of its unstructured nature, the results of different interviews might not be comparable.
VISUALIZATION TOOLS	
3. **Crop calendar:** Table is drawn on a flipchart of months in columns and various activities in rows. Farmers discuss monthly farm activities and relate these to resource use, pest problems, food supplies and income.	+ Can identify specific crop constraints. − Diversity in opinions or practices is difficult to capture on flipchart.
4. **Vignettes:** This tool is valuable in exploring sensitive topics with individuals because it de-personalizes the issue by asking the respondent to think about a situation that is occurring to someone else and provide their own judgment of how this person should respond. Vignettes can be used alone or anchored in household surveys.	+ Allows actions in context to be explored to clarify people's judgment; provide a less personal and therefore less threatening way of exploring sensitive topics. − The indeterminate relationship between beliefs and actions is often a major danger in using this technique in isolation.
5. **Trend analysis:** Participants draw a time line and discuss when certain events happened, things changed (e.g. land use change), or technologies were introduced.	+ Creates a dynamic view of the community that can also be used to discuss future changes. − Requires participants diverse in age.
6. **Resource flow diagram:** Participants draw the various components of the farming system (e.g. rice fields, vegetable fields, livestock, household, markets) and then draw the resource flows between these (e.g., nutrients, cash, labor).	+ Helps to understand interactions within the system. − Some participants can have difficulty comprehending the level of abstraction and consequently lose interest.
7. **Resource mapping:** Participants draw the boundaries of the community, the fields under various crops, water sources and differences in soil quality.	+ Helps to understand spatial variation in resources and land use. − The map might have little resemblance to reality.
8. **Problem tree analysis:** A tree with branches, trunk and roots is drawn. The trunk represents the problem, the roots the causes of the problem and the branches the consequences. Causes can be divided into immediate and underlying "root" causes, while consequences can be branched out into direct and indirect effects. The facilitator can guide the discussions, asking the participants to prioritize causes and effects or identify which causes and effects are getting better or worse.	+ Directly relates to the problem statement, objective of the project, and the proposed intervention. − A main problem first needs to be clearly identified and agreed upon, so it should not be the first tool to use. The high level of abstraction might limit participation.
9. **Objective tree analysis:** Similar to the problem tree except that the trunk is now the objective, the primary roots are the required interventions and the secondary roots are the enabling factors and assumptions.	+ Describes logical causal pathways of how project interventions are supposed to deliver the desired results (closely related to the Theory of Change). − High level of abstraction might limit participation.
10. **Daily activity clocks:** A small group of 2-3 women or men are presented with two circles representing the 24 hours of a regular working day (am and pm). They put all their activities, from getting up to going to bed, into the clock. Clocks of men and women are compared in a joint group discussion.	+ Suitable for gender analysis to assess how a technology affects the workload of men and women. The comparison will give a lively discussion. − More time allocation to an activity does not necessary mean that it is a burden. It does not reflect the priorities people give to different tasks.
11. **Spider diagram:** A small group of participants is asked to identify different priorities for their livelihood and score them as 5=very good, 4=good, 3=moderate, 2=bad; 1=very bad. These priorities can be identified in a focus group discussion.	+ Considers trade-offs between different priorities and needs. Can be useful for gender analysis by doing it separately for men and women. − Presenting different priorities as a spider web might be more confusing than helpful to some participants.

Tool	Main strengths (+) / challenges (−)
OBSERVATION TOOLS	
12. **Participant observation:** Direct observations in the community, the houses and the fields as a valuable source of information. The central idea is that a group of “outsiders" spends time in a village and has informal and open dialogues with the people on (all) aspects of their daily life.	+ The lack of structure minimizes observational biases and offers a more equal exchange of ideas and experiences, which builds mutual trust and interest. − Not all relevant community needs are observable.
13. **Transect walk:** A transect is a diagram that is produced during a walking discussion with community members and shows key features of different land use zones in a community. Transects are particularly useful when there is a range of different soil types or land use systems, for instance, as found in coastal, hilly or riverine communities.	+ Gives a basic understanding of agriculture in an area and its constraints and opportunities. − For the method to be effective, the group size should be limited. Recording information while walking is a challenge.

Source: Own compilation.

If little is known about a certain problem then it is useful to conduct a **scoping study**. Scoping studies help to quantify the extent of a problem, understand people’s perceptions about it and their current strategies used to address it. For instance, at AVRDC scoping studies are regularly conducted to better understand farmers’ perceptions about pest and disease problems in vegetable crops, quantify farm-level yield losses, and get information about farmers’ current pest management methods. A **value chain study** is a particular type of scoping study with the objective to identify constraints in how vegetables are moved from farmers’ fields to consumers’ plates. Such a study can reveal where product losses occur and where postharvest interventions consequently need to be introduced.

A clearly defined problem statement will logically lead to one or more project **objectives**, which in the case of development projects, are measurable changes in the situation (or behavior) of a target population which respond to its development needs. Well-defined project objectives should be specific, measurable, achievable, relevant and time-bound (“SMART”). It is important to clearly define a target population. An example of a well-defined objective is: “To increase the average vitamin A intake level among women and under five children by 25% in Coimbatore District of Tamil Nadu, India by 2020”.

2.3 Logical Framework Approach (LFA)

Having clearly defined what a project aims to achieve, the next step is to decide how to do it. The Logical Framework Approach (LFA) is the most commonly used project management tool for this purpose. Simply put, the LFA links project objectives to project activities. It was originally developed four decades ago for the United States Agency for International Development (USAID) but has since become a standard tool in project management and has been adopted and adapted by many private and public donor organizations.

A brief introduction of LFA using examples relevant to AVRDC follows, to show how the LFA can be used in projects that combine both research and development, and to define commonly used concepts to avoid misunderstandings in their use. LFA is more suitable for development than for research projects, but since nearly all projects at AVRDC are a combination of both, the tool is nevertheless useful. Using the LFA, a complex project can be summarized as a logical structure with a clear identification of basic inputs, outputs and outcomes so that others can understand it at a glance.

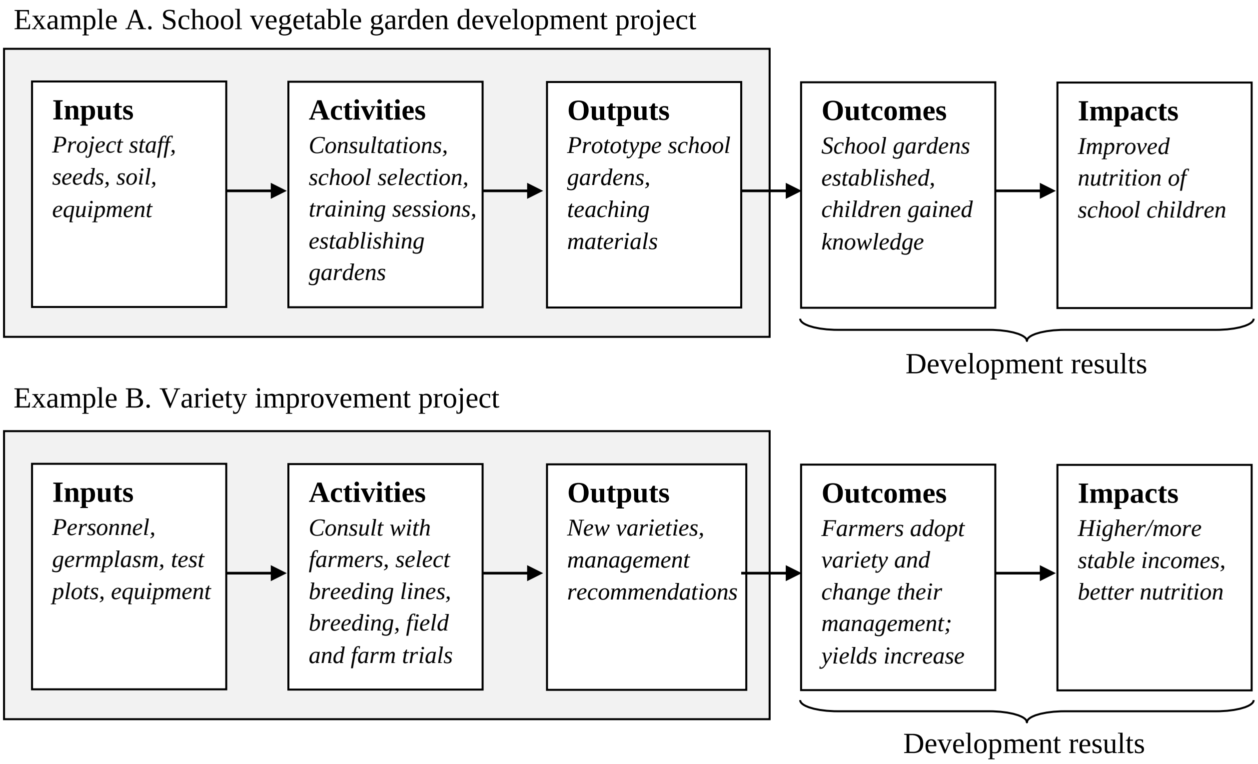

Figure 3 Example results chains for hypothetical research and development projects

The LFA depicts a causal or logical relationship between inputs, activities, outputs, outcomes and impacts. Figure 3 depicts these five elements as a results chain using examples of two research and development projects. Moving along the results chain, a project first needs inputs. **Inputs** are physical and non-physical resources that are put into the project, and for which outside funding is typically requested. Inputs are continuous and include personnel and overhead costs for maintaining the institution, as well as test fields and laboratories, vegetable seeds, chemicals, airfare, and the cost of organizing a workshop or conference.

Inputs are used to undertake project **activities** such as conducting field or laboratory experiments, collecting data for a baseline survey, interviewing stakeholders, analyzing data, writing reports, organizing workshops, and giving feedback to stakeholders. These activities generate project **outputs**, which are the direct results or deliverables of a project and can include innovations such as an improved vegetable variety or management method, and also workshops and conferences as well as journal publications and field manuals written within the context of the project. Outputs usually can be counted, have a direct link to a project activity, and occur within the project period. Inputs, activities and outputs are largely under the control of the project team.

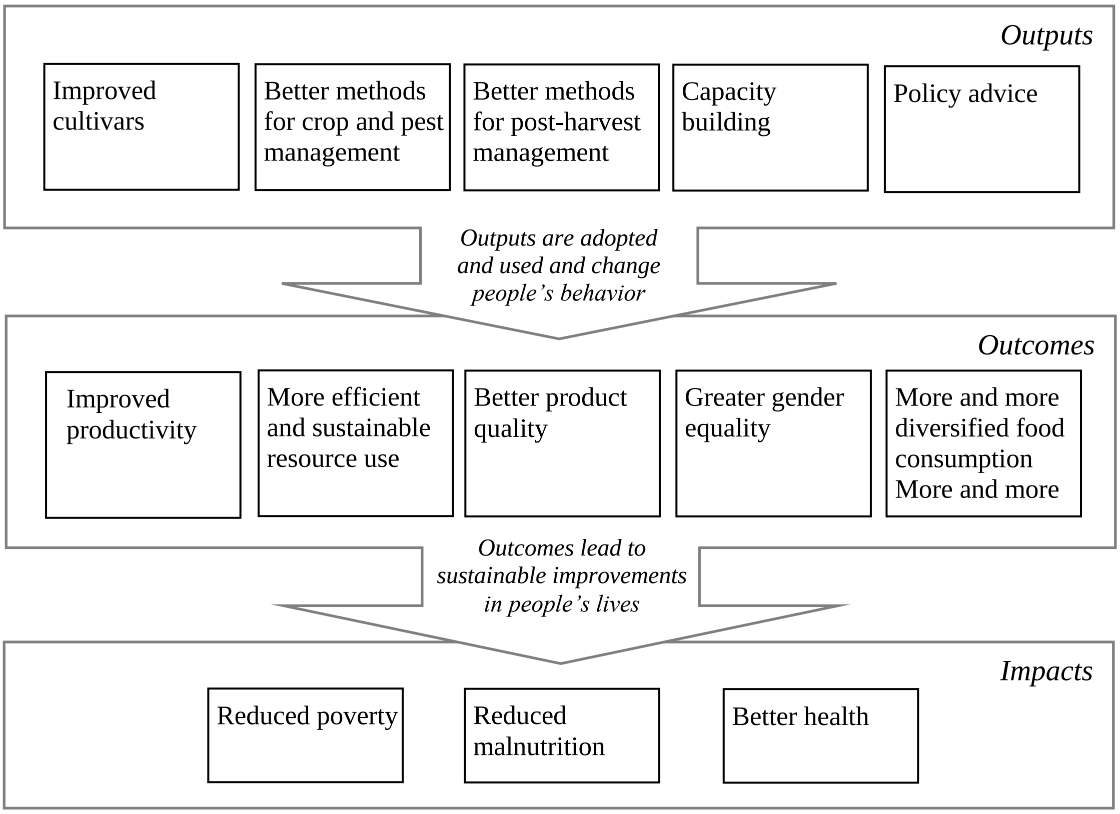

Figure 4 Outputs, outcomes and impacts in vegetable research and development projects

Outcomes refer to the adoption and use of the project outputs by the target population. Outcomes can include the adoption of improved varieties or management methods and the resulting gains in productivity, farmers' use of a new market channel for selling vegetables, or the use of a new storage facility for vegetables. Outcomes show a change in the conditions and behavior of the target population, but do not necessarily imply progress in development.

In the LFA, this progress in development of the target population is what is called **impact**. Simply put, impact is sustained progress toward meeting certain development goals; these are measurable and generate long-term improvements in the economic, health, social or environmental conditions of the target population. Figure 4 exemplifies the conceptual differences between outputs, outcomes and impacts for some of the main research activities at AVRDC. The distinction between outcomes and impacts is important in the LFA context, yet in practice these are not always easy to separate. This relates to the questions of how long is "long-term", or after how many years of higher incomes you can call income gains "sustainable". Some donors therefore prefer to distinguish between short-, medium-, and long-term outcomes rather than between outcomes and impacts.

2.3.1 The logframe matrix

Having defined inputs, activities, outputs, outcomes and impacts, a project needs to define outcome indicators that can be used to measure and track its progress. For this, the elements of the results chain are entered in a **logframe matrix** (or logical framework matrix) consisting of four rows and four columns. Table 3 gives a concise example of a logframe matrix for a school vegetable garden project.

Table 3 Example logframe matrix of a development project promoting school vegetable gardens

Results chain	Measurable indicators	Means of verification	Important assumptions
Impact/objective:			
- School vegetable garden based approaches improve nutrition among school children aged 6-12 years within 2 years	- Dietary diversity - Vegetable consumption - Anthropometry, e.g., stunting, wasting	- Food diaries, student weight and height at start and end of the school year for control and beneficiary schools	
Outcome:			
- School vegetable gardens adopted and awareness among children raised	- Number of school girls and boys reached - Food knowledge - Preference for healthy foods	- Monitoring forms. - Food knowledge and food preference test conducted at start and end of the school year for control and beneficiary schools	- School teachers are motivated and adjust the teaching curriculum - Students are interested in the school gardens
Outputs:			
- Prototype school vegetable garden designed - Project and control schools selected - Training materials available	- Production technologies made available (including crops, manuals, seed kits, etc.)	- Selected vegetables and technologies posted on website for each school - Seed distribution recorded using tracking forms	- Schools have no technical constraints that cannot be solved (such as lack of water)
Activities:			
- Training of trainers - Designing location-specific growing schedules - Randomly selecting control and beneficiary schools	- Names and number of trainers trained - Number of schedules developed - Names and number of project schools	- Workshop registration list and feedback forms. - Schedules posted on the website for each school. - Name and location of school posted on website.	- Local governments endorse the project - Randomly selected schools are willing to participate in the project

The first column of the logframe matrix describes the results chain described above. The second column lists measurable indicators to record and track if we are being successful. Well-defined outcome indicators, like the objectives, are “SMART”. The third column defines how we measure these indicators. The fourth column makes explicit the assumptions that underlie the structure of the project and the risk the project is exposed to. The logframe matrix also has a temporal logic running through the matrix from the bottom to the top: if the stated activities are implemented and the stated assumptions hold, then the outputs will be delivered; and if the outputs are delivered and the accompanying assumptions hold, then the outcomes will be achieved; and finally, if the outcomes are achieved and the stated assumptions hold, then the project objective will be achieved.

Table 4 lists some outcome and impact indicators relevant for AVRDC’s mission of reducing poverty and malnutrition and improving health. It is best to avoid indicators that are highly variable or that are likely to be only marginally affected by the project because they would require very large sample sizes to show a significant impact, as will be explained in Section 4.

Table 4 Selected measurable outcome and impact indicators of poverty, nutrition and health

Measurable indicator	Quantification
POVERTY/ FOOD SECURITY	
Per capita income	The sum of net output minus costs for each crop and livestock activity, minus annualized fixed costs and interest payments, plus non-farm and off-farm household income and remittances. Divided by the number of people permanently residing in the household.
Per capita food expenditures	Sum of household reported expenditures on food divided by the number of household members expressed in adult equivalents.
Food security index	Can use a standard Household Food Insecurity Access Scale to record how households experience food insecurity and calculate a food security index.
NUTRITION	
Dietary diversity	*Based on 24-hour recall:* The number of food groups consumed out of a set of standardized food groups during the previous day (FAO, 2010) *Based on 30-day recall:* For a complete list of possible food items, the respondent is asked if the household consumed it at least (a) every other day, (b) once or twice a week, (c) 1-3 times a month, (d) not at all. A dietary diversity index is calculated from the weighted sum of the number of food items and the minimum and maximum values as observed in the data.
Per capita daily nutrient intake	Calculated from household-level food consumption data using a 24-hour recall and collected from the person who usually prepares food in the household. Nutrient quantities calculated using food composition tables.
HEALTH	
Physical health	Incidence/prevalence rate of disease (e.g. diabetes, diarrhea) can be measured using questionnaire tools or clinical assessment.[2]
Malnutrition of children under 5 years	Measurement of weight and height (anthropometrics) and recoding of age and sex. Underweight (low weight for age), wasting (low weight for height), stunting (low height for age).
Micronutrient status	For instance, levels of hemoglobin, iron, and Vitamin A can be measured from blood samples (not generally done in AVRDC projects).
Cost-of-illness	Sum of medical expenses, transport costs to health center, and lost labor days converted to monetary units (using opportunity costs).

Notes: [1]For more details on food consumption, food security and dietary diversity see Smith and Subandoro (2007), on anthropometric indicators see Cogill (2003), on various measures of food security and nutrition see the Food and Nutrition Technical Assistance (FANTA) project: http://www.fantaproject.org/. For dietary diversity see FAO (2010) and Hoddinott (1999). For mental health see Beusenberg and Orley (1994). For an example of cost-of-illness see Templeton and Jamora (2010).

2.4 Theory of change and impact pathways

The Theory of Change (ToC) emerged in the 1990s as a means to model and evaluate comprehensive community initiatives within the Aspen Institute Roundtable on Community Change. The theory sets out the assumptions and causal links necessary to achieve the desired outcome/s rather than simply

[2] Incidence rate refers the proportion of people contracting the disease during a period of time. Prevalence refers to the proportion of people with a disease at a designated time, which is a function of incidence and average duration.

setting out the steps (inputs, outputs, outcomes) in the process of delivering project benefits. Assumptions change as learning develops or as the context responds to external influences including, but not limited to, the project itself. The components of a typical theory of change model include the following:

- A pathway of change that illustrates the relationship between a variety of outcomes that are each thought of as preconditions of the long-term goal.
- Indicators that are defined to be specific enough to measure success.
- Interventions that are used to bring about each of the preconditions on the pathway, and at each step of the pathway.
- Assumptions that explain why the whole theory makes sense.

Typically, changes for individuals are the first things that occur as a result of the projects, services, actions or planned strategies of a community initiative. As more individuals change, they may contribute to population-level changes. For example, if enough people increase their income, poverty rates may decrease. Individual impacts are the building blocks of community change; if they do not happen, it is unlikely that a community will improve. In the context of the Center's R&D interventions, these include: changes in attitudes (e.g. perceptions and beliefs toward use of a new technology), changes in knowledge and awareness, changes in farming skills, behavioral changes, changes in family income/savings and changes in vegetable consumption levels, institutions, seed policies, regulations, etc. However, by themselves these individual changes cannot ensure that positive changes will last.

Theories of change come in many shapes and sizes; for instance, there are different techniques to match beneficiaries' preferences for "planning" versus "doing". One recommended approach to visualize a ToC is to develop an outcome map that depicts relationships between strategies to be initiated and intended results. These results will include both short- and longer-term outcomes and may reflect changes at different levels, such as individuals, organizations, systems and communities. There is no right or wrong way to draw an outcome or impact pathway map; each map will look different, depending on the target beneficiaries' unique needs and preferences. The key is to listen to the views of beneficiaries in target communities and stakeholders so that your map reflects your beneficiaries' view of how change occurs.

2.4.1 Logical Framework Approach and Theory of Change

An LFA would tell you that setting up a farmer field school program on African eggplant seed production is an activity, farmers' attendance of the program is an output, and increased awareness and knowledge of seed production techniques is an outcome. A ToC, on the other hand, would tell you that farmers need to attend the field school over the entire season, at least two days per week for a minimum of 60 days, and that the curricula must focus on seed health, nursery production, pest and disease control, harvesting and processing in order for awareness and knowledge of African eggplant seed production techniques to increase. The major differences between logical frameworks and theory of change are summarized in Table 5. Combining the LFA and ToC can greatly improve project design and evaluation as these approaches are complementary. The recommended approach is to blend the strengths of logical frameworks with those of theories of change such that the program components of the logical framework can be complemented by pathway flow diagrams of mechanisms and linkages showing how outcomes can be achieved.

Table 5 Logical Framework Approach vs. Theory of Change

Logical Framework Approach	Theory of Change
Representation	Critical thinking
List of components	Pathway of change
Descriptive	Explanatory
Static	Dynamic
Linear	Complex

2.5 Monitoring and evaluating projects

2.5.1 Monitoring at various levels

Monitoring the performance of a project involves the systemized counting and tracking of resources used and outputs or results obtained, and a comparison of these to the original project plan. Monitoring is done continuously throughout the lifecycle of a project to ensure that a project stays within its budget, that resources are spent effectively, to strengthen accountability, and to identify possible problems so that these can be addressed by the project management to improve overall project performance. As monitoring is done continuously, it needs to be implemented as a routine operation by project staff. Monitoring can be done at the level of inputs and activities but also at the level of outputs.

First, the monitoring of inputs and activities involves the systematic recording of resources used. Through Maconomy, AVRDC's enterprise resource planning (ERP) system, project managers can keep track of all project expenses in real time. This allows project managers to see how resources are being used, identify possible delays in project activities, and assess whether project resources are being used as planned.

The second level of monitoring is to track project outputs. Many outputs can be quantified easily, such as the number of publications, the number of people (male/female) who participated in training workshops, and the number of training manuals distributed. Since 2013, the Center has used a centralized database to track such variables. This performance monitoring system, called VegOne, is maintained by the Center's Information Technology team.[3]

The third level of monitoring is to track outcome indicators. It is recommended to record costs and returns from on-station and on-farm technology trials. Farm-level impacts cannot be inferred from such data because research trials usually have few of the water, labor or cash constraints that real farmers experience, but such data nevertheless provide an indication of potential costs and returns, which is useful to assess whether a technology is ready for release.

2.5.2 Tools for tracking project resource use

To the extent that project activities involve the target population it is important to not only count the number of participants, but to also identify who they are. Using a one-page questionnaire, basic data can be quickly collected from the participants, including their age, sex, occupation, farm size, family size and relative poverty status. Project staff can use these data to find out if there was a selection bias among the participants and adjust the participant selection if necessary. In addition, training facilitators can be asked to write a brief workshop report, which summarizes the contents of the workshop, the feedback from participants and constraints and suggestions for future workshops.

For projects that distribute resources such as seed, seedlings, or promotional materials, it is important to identify the recipients at each stage of the distribution chain (e.g. ministries, local government, distributors, farm households, household members). This is, for instance, relevant for the distribution of emergency vegetable seeds kits. Carefully designed tracking forms can serve as an internal control system to ascertain that resources reach the target population, that they are used as intended and to

[3] VegOne can be accessed from within AVRDC at http://203.64.245.47:8080/VegOne/

identify possible constraints in the distribution system. Table 6 gives some examples of tracking tools that may be used for various purposes.

Table 6: Examples of the use of tracking tools for farmer field schools

Indicator	Importance	When to collect?	Monitoring tool
Farm household socioeconomic data (e.g. age, farm size)	Identify beneficiaries, baseline data for later evaluation	During the first training	Participant tracking form
Technical knowledge about vegetables	To evaluate the knowledge transfer	Before and after the training	Training evaluation form
Season-long farmer training	To assess farmer interest and effectiveness of training	Every training session	Facilitator session report; training evaluation form
Uptake of improved practices	To provide early evidence of adoption and to identify possible constraints	At some period after the training	Questionnaire-based follow up form
Post-season long training activities planned	To enable follow-up or linkage to other initiatives	At the end of the training cycle	Training evaluation forms

2.5.3 Performance monitoring plans (PMPs)

This is a tool for planning, managing, and documenting performance data. PMPs provide indicator definitions, sources of information, methods of data collection and analysis, and frequency/timing of data collection in a tabular format (Table 7). One purpose of PMPs is to collect comparable data from various project activities on a regular basis that can be then be aggregated to the program level. PMPs are used in USAID Feed-the-Future projects and the Humidtropics program, a CGIAR research program being implemented in the humid tropics.

Table 7 Example performance monitoring plan (PMP)

Indicator	Activity	Target	Achieved	% Achieved	Reasons for deviation
1. Number of farmers who participated in training activities (male persons)	Tomato grafting	30	30	100%	-
2. Number of farmers who participated in training activities (female persons)	Tomato grafting	30	30	100%	-
3. Area under improved technologies attributed to the project (ha)	Grafted tomato	10	8	80%	Insufficient capacity to produce seedlings

2.5.4 Project evaluation

The requirement to demonstrate results has highlighted the importance of project evaluation. Previously, project evaluations focused on **cost-benefit analysis**, which is the systematic accounting of project spending and project benefits expressed in monetary terms. Nowadays, the focus is on process and outcome or impact evaluation.

Process evaluation is concerned with how the project operates in terms of strategies and activities and can be used by project implementers and stakeholders as a self-learning exercise. Participatory project evaluation involves the project implementers, project beneficiaries and other stakeholders in the evaluation as an important learning tool for improving project operations. Focus group discussions are probably the most commonly used tool as interaction among stakeholders to help stimulate thoughts, recall experiences and generate new ideas. The evaluation can address questions such as:

- Did the project address the development needs of the target population?
- Who were the main beneficiaries and who in the target population was excluded?
- Is it likely that project outcomes can be sustained after the project?
- Under what conditions can project outcomes be replicated and scaled up/out?

Outcome or **impact evaluation** is concerned with the extent of accomplishment of a project's objectives. The terms are often used interchangeably, though outcomes generally refer to more immediate changes in the situation of the target population (e.g. technology adoption, productivity changes) and impact generally refers to sustained changes in livelihood indicators (e.g. long-term income, assets, health). Chapter 5 will discuss this in detail.

3 Dissemination and scaling strategies

3.1 Approaches to innovation dissemination and scaling

A wide range of dissemination approaches are available. Some are more effective for disseminating knowledge-intensive innovations, while others are more effective for reaching higher numbers of stakeholders with innovations that require less knowledge to apply. Therefore, choosing the optimal combination of dissemination approaches that fit the needs and characteristics of the target stakeholders is important for creating the maximum impact with available resources (Norton et al., 2005).

3.1.1 Training of Trainers (ToT)

For AVRDC, ToT is often an effective and efficient innovation dissemination approach because it allows the Center to bring its innovations more quickly to scale. In a ToT, the Center's scientists and partners train extension workers or lead farmers to disseminate innovations to farmers and/or other stakeholders through other approaches listed below. In a ToT, learning will be maximized among the participants if participatory methods are used more than lectures.

3.1.2 Farmer Field Schools (FFS)

FFS may be the best approach for disseminating knowledge-intensive innovations that require local adaptation because they are very participatory and typically cover an entire cropping season, which allows farmers/participants to learn in-depth about a range of topics and situations and adapt the innovation to their needs. Pontius et al. (2002) and Luther et al. (2005) provide details on the components and activities in FFS. For innovations that are less knowledge-intensive, it is not advised to use FFS, because some other approaches noted below may be more efficient for reaching larger numbers of stakeholders.

3.1.3 Field days

Field days are normally single-day events, which can reach large numbers of stakeholders but provide limited time to transfer innovations. The time therefore needs to be used efficiently to maximize the participants' exposure to innovations. If participatory exercises can be included, it is likely to increase innovation adoption versus participants simply viewing the innovations.

3.1.4 Publications: Fact sheets, booklets, leaflets, posters, etc.

When using this approach, the literacy level of target stakeholders needs to be assessed, and the publications designed to fit their level. Good-quality photographs and drawings should be incorporated as much as possible, to increase understanding of how to apply the innovations. Text should be concise. Attractive design and layout will make the publication more appealing, which may increase the stakeholders' interest in reading it and possibly adopting the described innovations.

3.1.5 Demonstration plots

Demonstration plots should provide a clear comparison between the newly-introduced and presently-used innovations, so that viewers can see the pros and cons of each. Strategically placed signs should explain clearly what is being demonstrated. If possible, the plots should be located by a road or other area where many people pass by, to maximize exposure.

3.1.6 Videos and TV or radio programs

The quality of videos is important for enabling adoption of an innovation. Proper footage of details will allow stakeholders to implement an innovation effectively, while poor attention to detail may eliminate key components necessary for adoption. Interviews for TV or radio can reach many stakeholders, so transmitting key details needed for adoption is instrumental.

3.1.7 Internet

The Internet becomes a more powerful dissemination tool daily as the number of users continually increases, but one needs to be aware of which stakeholders are likely to have access to it. Many resource-poor farmers in developing countries do not have Internet access and need to be reached through other means, but NARES and NGO extension officers rely increasingly on the Internet for information and they could be an important target for information on innovations.

3.1.8 Communication forums or workshops

While these are common, an often-neglected aspect is inviting people to attend who are most likely to disseminate the information or innovations to a broader audience; these people can greatly increase the sustainability and impact of a project after it is over. Effective information delivery during the workshop or forum is, of course, important for getting the main messages across. Proper planning for key information delivery can make a difference.

3.1.9 Dramas and songs

Entertainment-oriented innovation dissemination approaches such as dramas or songs can draw much attention to an innovation, particularly if the song is catchy or the drama has popular appeal. It is essential that the main message is clearly transmitted, however, and not overshadowed or lost in the overall performance.

3.1.10 Campaigns to spread simple messages

Short and simple messages conveying extension advice can be spread to large numbers of stakeholders through many means, including mass media. Mobile phones, billboards, radio broadcasts, and trucks with loudspeakers driving through villages are some means, and if the message is very simple it can be absorbed by many.

3.1.11 School vegetable gardens

School vegetable gardens can be used to disseminate a range of innovations and knowledge, including the nutritional and health benefits of vegetable consumption. Students may also learn how to grow vegetables successfully, and this may stimulate them to plant a vegetable garden at home. School vegetable garden programs may be integrated into the regular curriculum or be extracurricular activities.

3.1.12 Lessons in curriculum for school programs

Agricultural subjects in school curricula may be instrumental in keeping the general public in touch with important issues regarding where their food comes from, and the environmental implications of agricultural practices carried out in their region, which they may not be aware of. As the proportion of the population working in agriculture in many countries continues to shrink, it becomes increasingly important to keep the general public educated about key agricultural issues that affect everyone.

3.1.13 Home gardens

Successful promotion of home gardens may help households increase their vegetable consumption levels. Constraints to adoption often involve labor and time required from the household to successfully maintain the garden. Finding low-maintenance crops and other methods to minimize labor requirements may help increase adoption levels. AVRDC has developed a website tool for assembling a home garden.[4]

3.1.14 Mobile phone messages/broadcasts

As mobile phone use proliferates worldwide, so do the opportunities to use this increasingly important tool for innovation dissemination. Delivery methods need to fit the level of mobile phone technology that the target stakeholders are likely to have. In most cases, messages need to be concise, but with increasing internet access the possibilities are widening.

3.1.15 Plant health clinics

Under this approach, clinics staffed with plant protection scientists are held in rural areas so that farmers can easily bring their problems to scientists for consultation and help. Details can be found in Bentley et al. (2009). The approach is promising for creating impact, but needs detailed planning and consistent technical support to be implemented successfully.

3.1.16 Participatory breeding and farmer-managed variety trials

These approaches are strategic for obtaining input from farmers during variety development and final testing. Participatory breeding can also involve traders, processors, consumers and other actors along the value chain. If an open-pollinated line/variety is particularly appealing to farmers, they may adopt it immediately by saving seeds. The line/variety may be scaled-out quickly through farmer-to-farmer diffusion if large numbers of farmers find it an improvement over their currently used varieties.

3.1.17 Farmer-to-farmer diffusion

Farmer-to-farmer diffusion is likely to occur more readily with less knowledge-intensive innovations, because they can be transferred easily to other farmers. More knowledge-intensive innovations, such as many integrated pest management (IPM) methods, may diffuse more slowly because more instruction is required between farmers. These factors need to be taken into account if a project is counting on creating impact through farmer-to-farmer diffusion. This can be accomplished through a farmer field school or a video viewing club (VVC). The VVC approach was designed by the International Institute of Tropical Agriculture (IITA) specifically to encourage women's participation in training programs. However, VVCs are open to men just as farmer field schools are open to women. VVCs provide a more selective version of the topics covered in farmer field schools (FFS) and the training can be condensed into a shorter period not dependent on the growing season as is usually the case with season-long FFS. The clubs watch videos on technical topics such as pest and disease control, participate in discussions led by a trained facilitator, and practice the techniques in the field. The facilitator makes visits to participants' farms to provide additional support. VVCs have the advantage of relying less strongly on the quality of the facilitator since the videos provide the technical content, but are dependent on equipment (TV, videocassette player, computer, power supply) to function.

3.1.18 Demand creation activities

Advertising, exhibits and similar activities can greatly increase demand for an innovation or product. These need to be strategically planned to create maximum impact with the resources available. How to best target relevant stakeholders is just one aspect to consider. Such activity can also be aimed at increasing the demand for vegetables among consumers. For instance, in 2010, the Philippine government worked with celebrities in a national media campaign to fight malnutrition among children through raising awareness about the nutritional quality of common vegetables. The campaign was named 'Oh My Gulay!' (OMG!), meaning 'oh my vegetables'.[5] It played on the shorthand

---

[4] http://avrdcnutrition.gtdtestsite.comoj.com/homegarden

[5] The campaign was precipitated by the visit of Senator Edgardo Angara to AVRDC headquarters in 2008. See http://www.edangara.com/oh-my-gulay-2011-02-15 for more background information.

exclamation commonly used in mobile phone text messages 'OMG' for 'oh my god'. The awareness-raising campaign featured local celebrities posing with their favorite vegetables in advertisements in print media and television. Role models can have a big impact, particularly on children.

3.1.19 Mobile teaching laboratories

These have been used successfully in certain situations where resources are available to equip a vehicle with educational materials. Emphasizing hands-on and discovery-based learning methods makes these teaching labs more effective than just having reading materials.

3.1.20 Training and visit

This is an outdated approach which has fallen out of favor, but it may still be useful in certain situations, particularly for those with very limited extension budgets.

3.1.21 Public-private partnerships for quality seed bulking and distribution

Linking with private sector institutions can be advantageous for out-scaling an innovation, since they often have good facilities for mass-producing a product—for example, seeds of a new variety. They often have established advertising mechanisms in place that can greatly increase innovation adoption.

3.1.22 Innovation platforms

Agricultural innovation is not just about farmers adopting technologies. Innovation depends on a range of stakeholders in the value chain adopting different practices in a more or less concerted manner. Concerted action to innovate does not arise easily, but requires communication, learning, adaptation and negotiation (Boogaard et al., 2013). **Innovation platforms**, also called multi-stakeholder platforms, can offer a space for this to happen. Innovation platforms can be defined as: "[…] a group of individuals (who often represent organizations) with different backgrounds and interests: farmers, traders, food processors, researchers, government officials etc. The members come together to diagnose problems, identify opportunities and find ways to achieve their goals. They may design and implement activities as a platform, or coordinate activities by individual members" (Homann-Kee Tui et al., 2013). Innovation Platforms usually play an important role in complex large-scale systems-oriented programs such as in Humidtropics. The program makes a distinction between Innovation Platforms, which are more narrowly focused on a particular problem, innovation or commodity, and **Research for Development (R4D) platforms** that are more generally focused on agricultural and rural development. One R4D platform can lead to the initiation of several Innovation Platforms. Because R4D platforms are typically organized for a wider geographical level (e.g. province or region of a country) and may involve higher-level stakeholders such as representatives from ministries or national companies, they can facilitate scaling out successful innovations.

3.1.23 Learning alliances

A **learning alliance** is a multi-stakeholder network of individuals sharing knowledge and experiences on selected issues. Learning alliances can be used to share and develop knowledge about approaches, methods, policies that work or that did not work, and the reasons why (Lundy et al., 2005). Knowledge may come from within the alliance but also from the outside. They help to share knowledge across organizations or even across countries and promote a common understanding of problems and a collaborative search for solutions. Learning alliances have much in common with R4D platforms but have a stronger emphasis on knowledge and capacity building whereas R4D platforms tend to focus on problems diagnosis and innovation.

3.2 Studying adoption decisions

3.2.1 Quantifying adoption

Farm-level survey data are the most reliable for estimating adoption levels of innovations, but collecting them can be expensive and time-consuming and they only give an estimate for one particular year (Morris and Heisey, 2003). If adoption refers to the use of improved varieties, then seed sales data may be used as an alternative; but these only give a lower bound for the actual adoption if farmers also reproduce their own seed. A third method of estimating adoption is asking local experts such as extension agents, input suppliers or farm leaders to estimate the adoption rate. It is advisable to verify these estimates through triangulation by asking multiple experts to estimate the same or by letting various experts discuss together and reach consensus.

3.2.2 Adoption constraints

When an improved variety or other type of innovation has been adopted by a significant number of farmers then regression models can be used to understand the characteristics of farmers who use it and identify possible adoption constraints. The technique takes the adoption decision as the dependent variable and regresses it on a set of observable farm characteristics such as farm size, farm experience, access to credit and sources of information. Quantitative measures of farmer attitudes and risk preferences can also be included. Significant negative parameters can be interpreted as constraints on adoption, while significant positive parameters facilitate adoption. The marginal effect (or elasticity) of each variable informs how much the probability of adoption would increase if the farm characteristic would increase by one unit (or one percent in case of elasticity).

The estimation technique of the regression model depends on how adoption is quantified:

- If adoption is a binary variable (1 if adopted, 0 if not) then a logistic regression estimated with maximum-likelihood can be used.
- If adoption is a censored continuous variable (for example, the proportion of the crop area planted with improved varieties) then a Tobit regression or generalized linear modeling (GLM) can be used.
- If the dependent variable is a frequency (for example, the number of innovations adopted) then a Poisson regression can be used.

Some variables typically included in an adoption model are proxies for underlying factors (e.g. number of years of schooling might be a proxy for knowledge). Without a correct understanding of the underlying factors they can lead to the wrong conclusions. For instance, if the number of school years has a significant and positive effect on adoption of a tomato variety then we would need to find out what type of knowledge promoted adoption (as we should not conclude that children must be kept longer at school). As farmers gradually move from the non-adopter group to the adopter group over time, the characteristics of the two groups change and regression results can be expected to be different.

3.2.3 Effect of innovations on farm profits

To enhance the adoption of an innovation it is useful for farmers and other stakeholders to know how it would change their profits. This can be accomplished through a cost and returns analysis. To this end, three approaches are available depending on the kind and magnitude of the changes.

Partial budget analysis

This method can be used to evaluate changes in costs and returns from incremental changes in farm management; for instance, the adoption of a new tomato variety. The analysis is partial because it includes only those aspects of the farm affected by the change and assumes that all other factors remain constant. The method sums all positive and negative effects on revenues (e.g. better price, greater yield) and costs (e.g. higher cost of seed and fertilizer) to estimate the total net effect of the innovation. Partial budgeting can help farm managers decide the financial effect of an innovation on paper before they make the actual decision.

Whole farm budget analysis

Whole-farm budgeting is a detailed summary of the major physical and financial features of the entire farm operation. This is particular relevant if innovations change the farming system. Whole-farm budget analysis identifies all components of the farm and determines the constraints and resource flows among these. The method can be extended into whole farm mathematical programming, which is used in farm planning (Hazell and Norton, 1986). Simply stated, it is a method to determine the combination of farm activities (such as crop choices and resource allocations) that maximize the farmer's objectives but is feasible with respect to a set of fixed farm constraints (such as available amounts of land, labor, cash, irrigation water and access to technologies). Objectives are usually expressed in terms of profit maximization, but can also include other dimensions such as leisure, food consumption, or risk aversion. Firmly rooted in farm management theory, the method can be used in a normative way to maximize the profitability of a farm, but also in a positive way to explain a current situation and to explore the effect of changes in resource quantities, prices or the introduction of new technologies. The method is particularly useful to help farmers adjust to a changing economic and technological environment.

4 Attributing development outcomes and impact to interventions

When an improved disease-resistant tomato variety is introduced to a community of tomato growers, it will be observed that farmers have varying levels of interest in adopting it. Some farmers might have seen the new variety being grown in nearby test fields and eagerly start planting their entire tomato area with the new variety once they can obtain seed, others will want to try out some of the new seeds on a small plot near their house, while others might want to wait and see how the variety performs in their peers' fields before making a decision. As time passes and the new variety proves profitable, producing stable yields and generating consumer demand, more and more farmers are likely to adopt. Perhaps after five years the entire community will be growing the new variety. Although it is obvious that the new tomato variety has been a success and will have contributed to higher living standards of tomato growers, the question might arise "by how much?"

Outcome or impact evaluation is about assessing how much of the observed changes in outcomes, such as improvements in livelihoods, are due to a project intervention such as the introduction of a new tomato variety. Attribution forms a major challenge because changes in living standards, or any other type of socioeconomic change, are the result of multiple confounding factors. Due to the complexity of socioeconomic systems it is difficult to separate the effect of a project intervention from that of other factors. Yet, if an evaluation study is able to correctly establish a causal relationship we say that it has high **internal validity**.

However much a single farmer benefitted from adopting an improved technology, a meaningful evaluation must be able to generalize impacts to a larger target population. The ability to generalize from a smaller sample is referred to as **external validity**. If an evaluation study cannot be generalized to a larger population, then it suffers from a so-called **selection bias** as the sampled households are not representative of the target population at large. The challenges to evaluation studies depend on the type of outcome indicator that is being measured, as illustrated in Figure 5.

Attribution or selection bias is typically not an issue if the outcome indicator is the adoption rate of an improved technology. Changes in adoption rates can be directly deduced by comparing baseline and follow-up surveys; or just by asking farmers what year their adopted. However, if the outcome indicator is farm productivity, then we need to control for several confounding factors such as the use other inputs and weather. It becomes even more challenging at higher levels of outcomes such as the household income, nutritional status or health. Not only does the number of confounding factors increase, and with it the difficulty in correctly attributing any changes in outcomes to the technology, but the relative impact of the technology becomes smaller and might even become statistically insignificant.

Quantitative methods of ex-post impact evaluation fall into three broad categories of research designs: non-experiments, quasi-experiments, and true experiments. These methods differ in how they address attribution and selection bias. This chapter explains why the use of control groups in true experiments and quasi-experiments gives a higher internal validity than non-experimental designs.

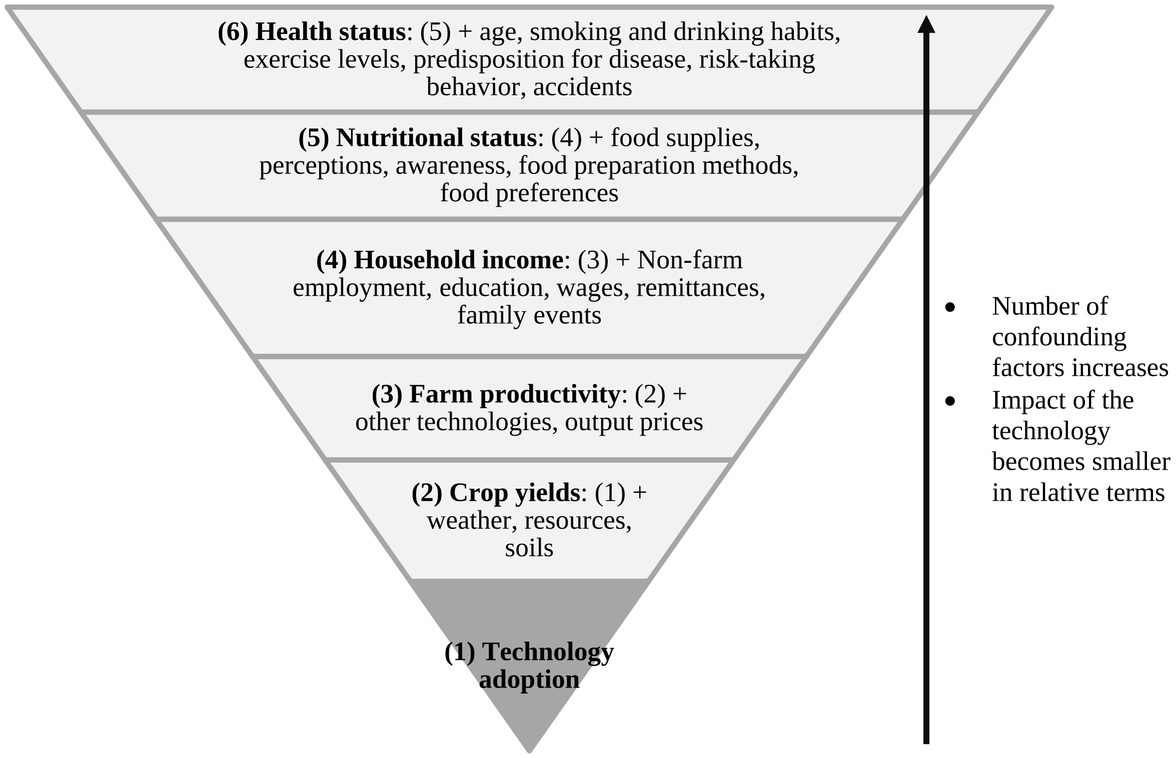

Figure 5 Attribution become more challenging at higher levels of outcome indicators

4.1 Non-experimental designs

Non-experimental designs are descriptive methods of impact evaluation (as opposed to analytical methods). A non-experimental design could be to collect baseline data from farm households before the technology becomes available and then to do a follow-up survey after households adopted the technology. The difference in outcomes between the two surveys could be interpreted as the impact of the technology. The problem with this type of comparison is attribution. The well-being of farm households is determined by many factors other than the adoption of the technology—price changes, rainfall, temperatures, the introduction of other technologies, resource changes—and we are unable to attribute how much of the total impact was because of the technology.

Another non-experimental design would be to compare data on adopters with data on non-adopters and assume that the difference was the result of the technology. This type of comparison suffers from low external validity as the benefit of adoption for current users of the technology is almost never representative of the benefits that other farmers would obtain if they would also adopt. The reason is that early adopters are almost always systematically different from non-adopters in many aspects. Adopters might have a larger farm size, greater supply of family labor, a higher education or more financial assets, which are called observable characteristics (as we can measure them). They might also be different in aspects such as skills, mentality, perceptions, creativity, entrepreneurial ability, personal interests, and risk preferences, which we cannot directly measure and therefore call unobservable household characteristics. These are likely to not only influence adoption but also to have a direct positive effect on farm performance. Adopters are therefore likely to have better farm performance even before they adopted the technology. This selection bias will lead us to overestimate a technology's impact.

4.2 Quasi-experimental designs

Ideally we would want to compare the same household at the same time with and without the presence of the new tomato variety so that any confounding factor is eliminated and attribute all observed differences to the new variety. Yet, as we cannot observe the same individual in two different states of the world at the same time, we need to find a valid counterfactual or control group against which to compare the performance of the adopters.

An ideal control group comparison must meet three conditions (Gertler et al. 2011): First, the group of farmers that adopted the technology and the control group must have identical characteristics in the absence of the technology. Second, if the technology would be introduced, both groups must react to it in the same way. Third, both the control and the treatment group must not be differently exposed to any interventions other than the introduction of the new variety. None of the three quasi-experimental methods discussed below perfectly meets these conditions. However, these methods try to approximate the ideal control group by applying statistical methods on household data for adopters and non-adopters.

4.2.1 Instrumental variable method

A first method to quantify impact is to use a linear regression of an outcome variable on a set of independent observable household characteristics with the technology adoption decision introduced as a dummy variable (1 if adopted, 0 otherwise).[6] If our outcome variable is crop yield or crop output then we can specify a production function, yet the outcome variable could also be per capita incomes, expenditures or nutrient intake. The sign and significance of the adoption dummy can be interpreted as the impact of adoption assuming constant levels of all other inputs. The inclusion of observable household characteristics in the regression acts as a control to reduce the effect of selection bias on the parameter for technology adoption. However, it only removes selection bias due to observable household characteristics, not due to unobservable characteristics such as skills or entrepreneurial ability, which are also likely to determine technology adoption. Since these characteristics affect technology adoption as well as the outcome variable, the error term in the regression equation, which includes these unobservable characteristics as well as measurement error, will be correlated with the technology dummy—thereby violating an important assumption of the linear regression model.

The challenge is therefore to replace or 'instrument' the adoption variable with a variable that is (a) strongly correlated with the adoption variable, but (b) conceptually unrelated to the outcome variable given adoption.[7] This new variable is constructed from the predicted values of a first stage regression of the adoption decision on a set of valid instruments.[8] Inclusion of predicted values in the second stage will give a more correct estimate of impact as it will have removed some of the selection bias due to observable household characteristics. Finding such a variable is practically difficult but if at project conception stage, a particular geographical area for promoting the technology was well targeted, then location dummies might be used as valid instruments.

[6] Note the similarity to the adoption model presented in Section 3.2.2, which used adoption as the dependent variable, while the impact model uses it as an independent variable to explain observed differences in outcomes.

[7] In other words, the instrument must not be correlated to the error term, which includes the unobservable household characteristics.

[8] This can be achieved through either a logit transformation of the dependent variable if the adoption decision is binary or, if the adoption is a percentage area under the new cultivar, by using generalized linear modeling.

Table 8 Strengths and limitations of various quasi-experimental designs

Method	Strengths	Limitations
Instrumental variable method	• Instrumenting the adoption variable partly controls for selection bias • Can be used without baseline data (but baseline data can yield better estimates)	• Finding good instrument is a challenge • Estimated impact only applies to a subset of adopters
Difference-in-difference method	• Combining before-and-after data with with-and-without data increases internal validity • Control and project populations are allowed to be different	• Assumes that adopters and non-adopters are exposed to identical dynamics • Assumes the absence of spillover effects
Statistical matching	• Can even be used without baseline data (but ideally, the matching is done on pre-intervention data) • Several alternative matching algorithms can be applied	• Assumes that selection bias is only due to observable characteristics • Dropping observations that have no common support creates a bias • Requires a large sample of non-adopters relative to the sample of adopters

4.2.2 Statistical matching

Matching methods construct an artificial comparison group by identifying for every farmer who has adopted the technology, the most similar farmer (or farmers) who has (have) not adopted. Matching is done on observable household characteristics and ideally uses baseline data to rule out any changes in observable characteristics that may have occurred because of the adoption. However, the method also can be used on pure cross-sectional data as some household characteristics such as age, land size, and education are not (or not immediately) affected by the technology adoption.

Similarity between adopters and non-adopters is determined using a matching criterion. Commonly used criteria include:

- **Distance function matching**: A distance function is a metric that can be used to define the similarity (or difference) between observations in a dataset based on their location in a multi-dimensional space. For instance, the Euclidian distance between two points on a two-dimensional graph is the length of the line segment connecting the points.
- **Principal component matching:** Principal component analysis is a data reduction method used to convert a set of possibly correlated variables into a smaller set of one or more linearly uncorrelated variables. The first principal component, which has strongest factor loadings, can be used as matching criterion.
- **Propensity score matching:** Propensity score matching first estimates a technology adoption model (similar to what was shown in Section 3.2.2 on page 18, but causality is not as important if used for statistical matching) and estimates the probability of adoption—called predicted propensity scores—for each farm household based on observable characteristics. Propensity score matching is not suitable if not all farm households had equal access to the technology (e.g. if the intervention targeted some locations only). If this is the case then the adoption model will give a poor fit and other matching criteria will be more suitable.

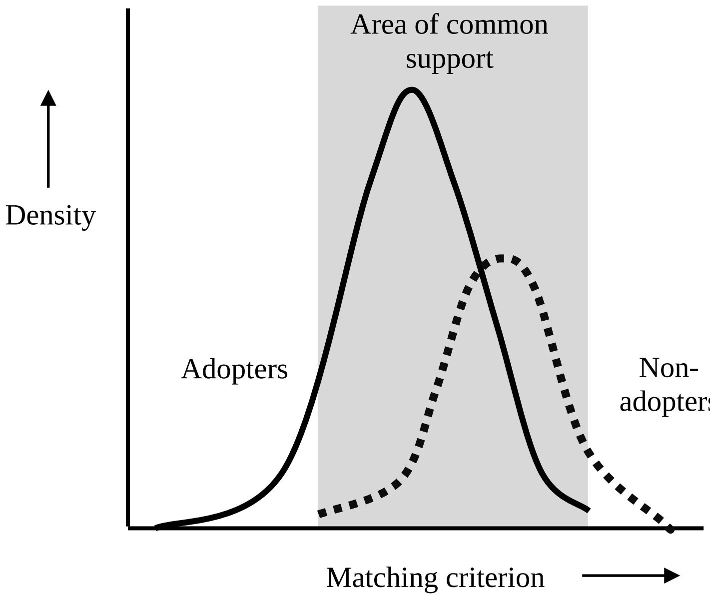

Figure 6 Statistical matching of adopters with non-adopters

These criteria can be used to rank all farm households in order of similarity and plot a density function for each of the populations as shown in Figure 6. It is not always possible to find a suitable match for each adopter in the non-adopter group. Farmers at the extreme ends of the score (those with a very high or very low matching scores) may have to be excluded from the analysis as there is not enough common support on the matching criterion. To avoid dropping adopters from the analysis, the sample of non-adopters should be relatively large. Matching studies therefore typically oversample non-adopters.

Methods to match adopters to non-adopters include:

- **Nearest-neighbor method**: This matches each adopting household to the non-adopting household with the closest matching criterion value. It is also possible to match each adopter to its five nearest neighbors. Matching with or without replacement is possible. For each matching pair the difference in outcome values is calculated and then averaged over all pairs.
- **Stratification or interval method**: This first divides adopting households in strata such as decimals and then calculates the average value for the outcome indicator for adopters and non-adopters within each strata and takes the difference between these two averages as the impact of the technology on the strata.

Other methods as well as variations on these methods are available.

As the matching methods only include observable characteristics, de Janvry et al. (2011) note that the obvious problem with this method is that unobservable characteristics are not considered. These are likely to be important determinants of adoption as well as the general farm performance. In addition they note that it needs to be assumed that there are no spillover effects between adopters and non-adopters. Both of these are relatively strong assumptions.

4.2.3 Difference-in-differences (DD)

The difference-in-differences (DD) method combines before-and-after comparisons and with-and-without technology comparisons to produce an estimate of impact that is better than each difference alone.

Figure 7 illustrates how the DD method works. The method compares the change in outcomes for the group that did not get the technology (C2–C1) with the changes in the group that did get the technology (P2–P1). Hence the impact of the technology is calculated as (C2-C1)-(P2-P1). The with-and-without groups must not necessarily have the same baseline conditions, as is illustrated in the figure with two different starting points. If unobservable characteristics do not change between the baseline and follow-up survey, then these cancel out and the problem of selection bias disappears.

However, for the DD estimate to be valid, the changes in the control group must accurately represent the changes that would have occurred in the treatment group had the technology not been introduced as is illustrated with the dashed line. Both groups should therefore be exposed to exactly the same dynamics and the control group should receive no effect from the technology adoption in the other group. This is important because any differences in trends between the two groups will be attributed to the new technology.

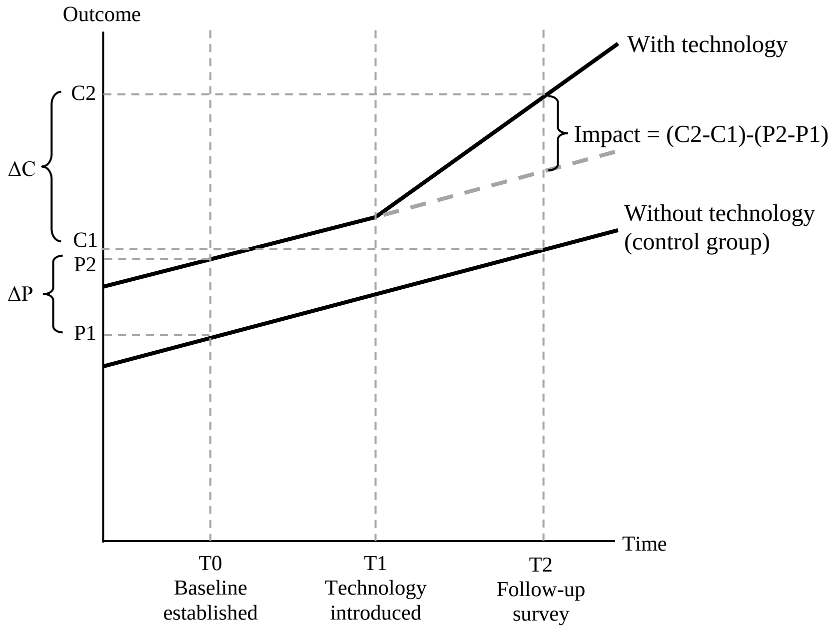

Figure 7 The difference-in-differences estimator of impact

The figure also shows that the impact estimate does not depend on the time that the baseline was conducted (if conducted at T1, the result would have been the same), but does depend on the time when the follow-up survey was conducted, as is a usual problem for any type of impact assessment.

The DD estimate of impact can be obtained from a regression analysis using farm-level survey data, specified as:

$$Y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 X_{it} + \beta_3 T_{it} X_{it} + \varepsilon_{it} \quad (1)$$

in which *Yit* is the outcome of household i at time t. The equation separates between a before-and-after effect (T=0,1), a with-and-without technology effect (X=0,1), and an interaction effect (TX=1,0); εit is the residual term. The parameters in this equation directly relate to parameters in

Figure 7 with β0 = P1, β1 = P2-P1, β2 = C1-P1 and β3 = (C2-C1)-(P2-P1). Additional variables can be included in the regression to control for any other factors that influence outcomes.

4.3 Experimental designs

Randomized Controlled Trials (RCTs) are an experimental design that is similar to a field experiment. RCTs avoid problems of selection bias and attribution by using a two stage randomization procedure as shown in Figure 8.

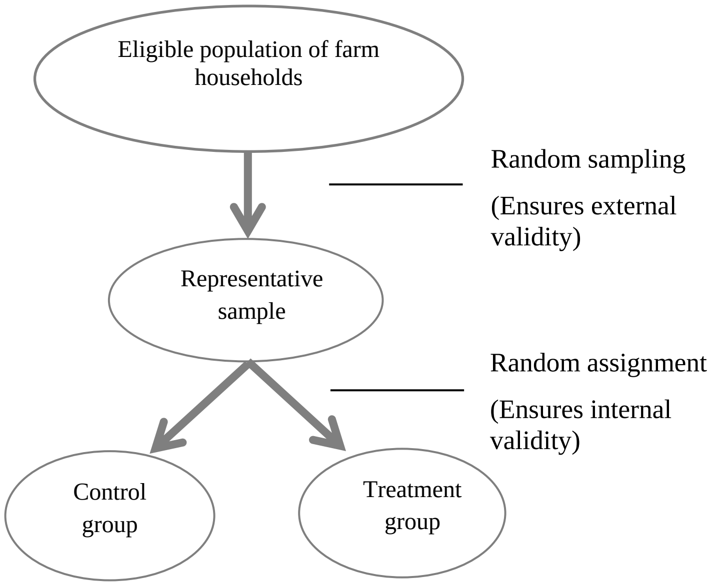

Figure 8 Internal and external validity in a randomized controlled trial

Randomization does not imply that all households or villages need to be included in the project. A more narrowly defined geographical area needs to be defined as the project cannot usually work in every part of a country. Most projects have narrowly-defined eligibility criteria such as poor smallholder households. A first step in the random selection is therefore the identification of eligible villages and eligible households. It is important that the eligible population only includes households for whom the technology is suitable and who therefore have a real adoption decision to make.

In the first stage a representative sample of farm households from the total population of households eligible to adopt the technology is randomly taken—ensuring external validity of the study. The use of randomization avoids selection biases that would exist if farmers themselves could approach the project to obtain seeds, or if project staff, extension agents or village headmen would decide on the distribution.

In the second stage, the sample households are randomly assigned to either the treatment group or the control group—ensuring the internal validity of the study since both groups, being random samples from the larger population, should have a high probability of being statistically identical in both observable and unobservable household characteristics. This will be the case if each sample contains a large enough number of households. Thus choosing the correct sample size is important for RCTs. Each group is a representative sample of the larger target population and baseline data can be used to verify this.

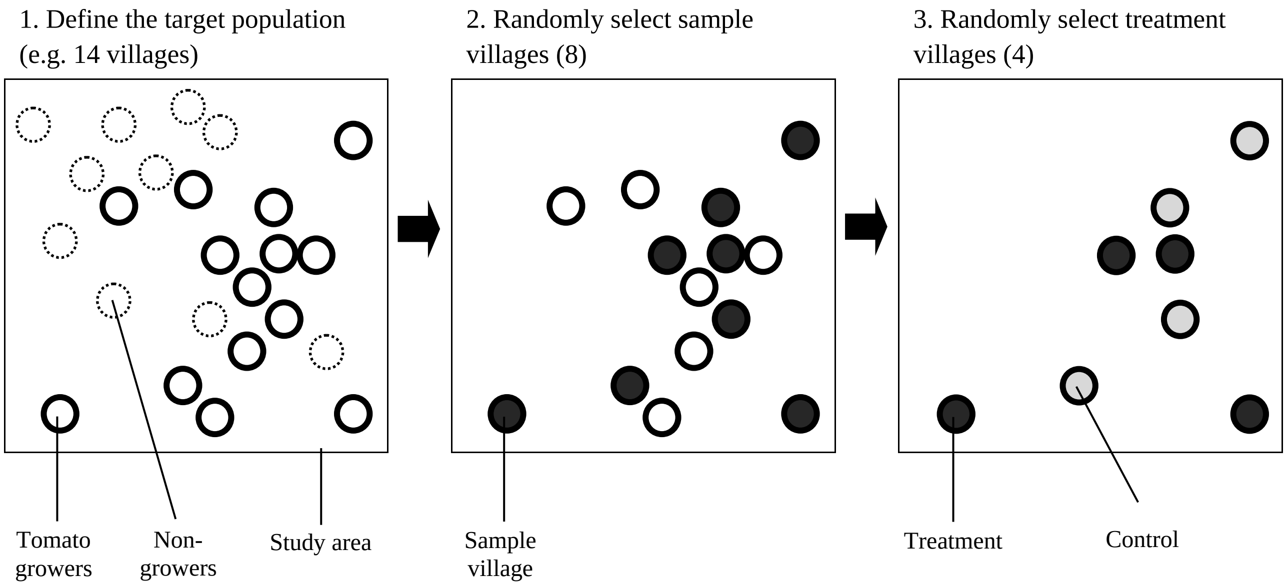

Figure 9 Randomized selection of control and treatment groups in RCTs

The use of RCTs with poor farmers might raise an ethical question because a potentially good technology is intentionally withheld from some farmers. Yet, when introducing a new technology such as an improved variety or better management method, a project can usually never make it available to all farmers at once. In the case of new varieties, not enough seed might be available to supply every farmer, or a technology might require training while training capacity is limited. Even without RCTs, a project has to make an ethical decision about who gets benefits now, and who gets it later.

In fact, RCTs are particularly useful in such situations as they avoid the use of subjective criteria in distributing project benefits. Randomization essentially uses a lottery to decide who among the eligible recipients will receive seed. If done in a transparent fashion, ensuring that each eligible farmer has equal chance of selection, then it avoids the use of arbitrary criteria and leads to a fair distribution of project benefits. To avoid the situation in which one farmer gets the technology, but his neighbor does not, randomization can be done at the level of villages rather than at the level of households. Such clustering also simplifies project logistics.

Although RCTs do not necessarily require baseline data, such data are useful to verify that both treatment and control groups are statistically similar. The random assignment to control and treatment groups needs to be repeated if this is not the case. Baseline data can also serve as a backup plan in case the RCT design fails—for instance, if an outside organization interferes in the treatment or control group; the DD or matching estimator can then be used as a second-best option.

4.3.1 Practical challenges

When using an RCT design for quantifying impact, a number of challenges can arise:

Control group contamination. Even if the project team can resist the temptation of giving seed to farmers who come and ask for it, some farmers might still be able to obtain it through other channels. This will be a particular problem in the case of open-pollinated varieties, when the treatment group starts reproducing its own seed. As a result, the counterfactual group is no longer a true control.

Spillover effects. The adoption of a new technology in the treatment group can have an effect on the control group through changes in the demand for labor, land prices or farm gate selling prices. Spillover effects can be positive or negative. The problem will occur if adoption is widespread.

Both problems can be alleviated by reducing the time span between the pre- and post-intervention data collection. Using a clustered randomization can also alleviate these problems as it reduces interactions between treatment and control groups. However, as clustering reduces sample variation, the disadvantage is that it can increase the required sample size (see Section 4.3.2).

Uncontrolled access to the technology. If the project team cannot control who gets access to the technology and who not then selection bias cannot be avoided and we can no longer establish the impact from comparing adopters with non-adopters. In such a case, a randomized promotion or encouragement design can be used. The researcher randomly selects non-adopters from the target population and promotes technology adoption among them. If the promotion leads to a sufficient increase in adoption, then the difference in adoption rate inside and outside the promotion area can be used to quantify the impact of adoption by using the promotion as an instrument for the actual adoption decisions in a two-stage-least squares (2SLS) regression. To correctly quantify the impact of the technology, rather than that of the promotion itself, it is important that promotion does not affect the outcome variables directly; monetary payments or the provision of free credit should therefore be avoided.

A disadvantage of randomized promotion designs is that they do not fully avoid the problem of selection bias. Farmers who were quick to adopt the technology when it became available probably had a higher expectation of its benefits and might have benefitted more than those farmers who only adopted after the technology was promoted to them (de Janvry et al., 2011). As a result, randomized promotion designs only provide a so-called local average treatment effect (LATE), which is the effect of the treatment on those farmers who only adopted after promotion. It is problematic to generalize this impact to the whole target population.

Imperfect compliance. The treatment group will include farmers who are not interested to adopt because they think the variety is unsuitable for their farm, as well as farmers who are unable to adopt because they may face land, cash or labor constraints. If there is no full compliance in the treatment group then the design will estimate the average impact for the whole population of adopters and non-adopters rather than the average impact for adopters, which is what we want to know.

If adoption is substantial but incomplete then it is possible to statistically correct for this using a 2SLS regression in which the randomized assignment is used as an instrument for actual adoption. This is possible because the randomized assignment is strongly correlated with actual adoption but, being random, is uncorrelated to project outcomes. De Janvry et al. (2011) pointed out that this approach will not avoid possible spillover effects unless the stratification is done at the village level and the village-level treatment is used to instrument adoption (1 if treatment village, 0 if control). A 2SLS regression with households as the unit of analysis will then give a correct estimate.

Yet, if a large share of farmers in the treatment group are not immediately convinced of a technology's benefits and therefore hesitant to adopt, then the 2SLS method would yield poor estimates on small samples. Either the sample size needs to be increased or farmers must be given some incentive to adopt, similar to the use of randomized promotion. The promotion can be used to instrument the actual adoption decision as described above. The promotion must be effective and support measures that directly influence outcomes must be avoided.

Figure 10 summarizes the above by showing three variations of RCTs suitable for setups in which the project can or cannot control access to the technology and whether or not the treatment group fully adopts the technology provided. It shows that the RCT is most straightforward to apply if the project team can control access to the technology and the technology has a high enough expected profit so that the entire treatment group adopts.

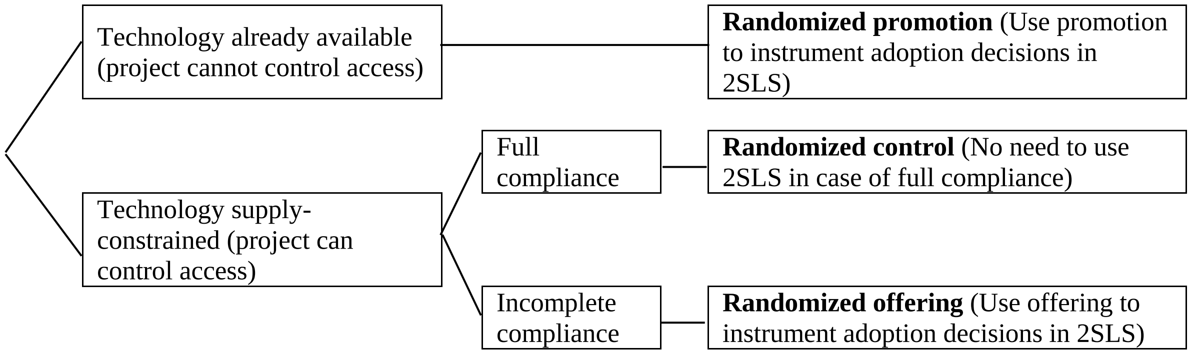

Figure 10 Three kinds of randomized controlled trials (RCTs)

4.3.2 Sample sizes

Sample clustering and sample size are important decisions in an RCT as these affect the reliability of the estimated impact. In general, the larger the sample, the greater the probability that significant impact will be detected. Yet, as data collection is expensive and time-consuming, the question of the minimum required sample size is relevant.

The answer depends on the outcome variable of interest, its mean and standard deviation. The higher the variation in the outcome variable, the larger the sample will have to be. It also depends on the desired level of significance of the impact estimate (usually set to either 0.90 or 0.95) and the desired statistical power of the evaluation, which is the probability that it will detect an actual difference between the treatment and control group (usually set to 0.80 or 0.90). The sample size furthermore depends on the minimum level of impact you would like to detect, as showing relatively small changes in the outcome indicator will require larger samples.

Assigning entire villages to the treatment and control groups as in a randomized clustering procedure also affects the required sample size. Clustering reduces the variation between households (two households randomly selected from the same village are more likely to be similar than two households randomly selected from the whole population) and therefore requires a larger sample. The sample size is therefore also a function of intra-cluster correlation and the number of households per village. With these seven pieces of information the minimum sample size can be calculated using various available sample size calculators.

4.4 What method to use?

Table 9 compares the four evaluation methods described in this framework by seven common challenges. It shows that RCTs have a clear strength in preventing selection bias, which is the most common challenge in evaluation studies. Control group contamination, spillovers, and divergent dynamics can be problems in RCTs and need to be avoided through the project design; for instance, by doing the randomization at the village level. Spillover effects are a problem for all estimation methods and can be addressed only through control group selection; however, this might be easier in an experimental setup. As DD and statistical matching methods do not rely on an experimental setup, they don't have problems of control group contamination or imperfect compliance and can be used if the control and treatment groups are not statistically identical. There are methods to address imperfect compliance in the design of the RCT, while DD or statistical matching estimators can also be applied to experimental data to correct for possible selection bias.

Table 9 Comparison of four methods of impact assessment

Challenge	Randomized controlled trial (RCT)	Difference-in-differences (DD)	Statistical matching	Two stage least squares (2SLS)
Selection bias	Random assignment minimizes selection bias. Else, correct for it using DD or matching methods	Is allowed, as long as adopters and non-adopters are exposed to the same dynamics	This method corrects only for bias in observable characteristics	This method corrects only for bias in observable characteristics (if instruments are available)
Spillovers from treatment to control group	Can avoided to some degree by using randomized clustering approach	This will bias the estimates	This will bias the estimates	This will bias the estimates
Control and treatment groups have divergent dynamics	This will bias the estimates	This will bias the estimates. Cannot use	Not a problem as the method only uses cross-section data	Not a problem as the method only uses cross-section data
Control group contamination	Will bias the estimates. Can be avoided through project design	This is not a problem in a non-experimental data	This is not a problem in a non-experimental data	This is not a problem in a non-experimental data
Imperfect compliance	Need to instrument adoption decisions in the analysis	This is not a problem in a non-experimental data	This is not a problem in a non-experimental data	This is not a problem in a non-experimental data
No baseline data were collected	Can use, but cannot verify if control and treatment are similar	Cannot use	Can use, but using baseline data is better	Not a problem as the method only uses cross-section data

Note: Another impact assessment method is regression discontinuity design (RDD). We have not discussed this here because RDD relies on the existence of a clear cutoff in eligibility (e.g. poor/non-poor) and agricultural technologies usually cannot be targeted in such way.

References

Asian Development Bank (ADB) 1998. *Using the Logical Framework for Sector Analysis and Project Design: A User's Guide*, Asian Development Bank, Manila, Philippines.

AVRDC - The World Vegetable Center 2010. *Prosperity for the poor and health for all: Strategic plan 2011-2025*, AVRDC - The World Vegetable Center. AVRDC Publication No. 10-738, Shanhua, Taiwan.

Bentley JW, Boa E, Danielsen S, Franco P, Antezana O, et al. 2009. Plant health clinics in Bolivia 2000—2009: operations and preliminary results. *Food Security*, DOI 10.1007/s12571-009-0033-z

Beusenberg M, Orley J. 1994. *A user's guide to the self-reporting questionnaire*, WHO Division of Mental Health, Geneva.

Boogaard B, Schut M, Klerkx L, Leeuwis C, Duncan A, Cullen B. 2013. *Critical issues for reflection when designing and implementing Research for Development in Innovation Platforms*. Report prepared for the CGIAR Research Program on Integrated Systems for the Humid Tropics (CRP Humidtropics), as part of Strategic Research Theme 3 'Scaling and institutional innovation', Knowledge, Technology & Innovation Group (KTI), Wageningen University & Research centre, the Netherlands.

Cogill B. 2003. *Anthropometric Indicators Measurement Guide*, Food and Nutrition Technical Assistance (FANTA) Project, Washington, DC.

de Janvry A, Dunstan A, Sadoulet E. 2011. *Recent Advances in Impact Analysis Methods for Ex-post Impact Assessments of Agricultural Technology: Options for the CGIAR*, Independent Science and Partnership Council Secretariat, Rome, Italy.

Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. 2011. *Impact Evaluation in Practice*, the World Bank, Washington, DC.

Hazell P, Norton R. 1986. *Mathematical programming for economic analysis in agriculture*, Macmillan, New York.

Hoddinott J. 1999. *Choosing outcome indicators of household food security*, International Food Policy Research Institute, Washington, D.C.

Homann-Kee Tui S, Adekunle A, Lundy M, Tucker J, Birachi E, Schut M, Klerkx L, Ballantyne P, Duncan A, Cadilhon J, Mundy P. 2013. *What are innovation platforms?* ILRI, Nairobi and Addis Ababa. Available at: https://cgspace.cgiar.org/handle/10568/34157

Lundy M, Gottret MV, Ashby J. 2005. *Learning Alliances: An approach for building multistakeholder innovation systems*. ILAC Brief No. 8 Rome, Institutional Learning and Change (ILAC) Initiative. Available at: http://www.cgiar-ilac.org/files/ILAC_Brief08_alliances_0.pdf.

Luther GC, Harris C, Sherwood S, Gallagher K, Mangan J, Gamby KT. 2005. Developments and innovations in farmer field schools and the training of trainers. Chapter 9 in Norton GW, Heinrichs EA, Luther GC, Irwin ME (eds.), *Globalizing Integrated Pest Management – A Participatory Research Process*. Blackwell Publishing, Ames, IA, USA, 338 pp.

Morris ML, Heisey PW. 2003. Estimating the benefits of plant breeding research: methodological issues and practical challenges. *Agricultural Economics* 29(3), 241-252.

Norad 1999. *Logical Framework Approach: handbook for objectives-oriented planning*, Norwegian Agency for Development Cooperation, Oslo, Norway.

Norton GW, Heinrichs EA, Luther GC, Irwin ME. 2005. *Globalizing Integrated Pest Management – A Participatory Research Process*. Blackwell Publishing, Ames, IA, USA, 338 pp.

Pontius J, Dilts R, Bartlett A. 2002. *Ten Years of IPM Training in Asia -- from Farmer Field School to Community IPM*. FAO Community IPM Programme, Bangkok. Available at: http://www.fao.org/docrep/005/ac834e/ac834e00.htm

SIDA. 2003. *The Logical Framework Approach - A summary of the theory behind the LFA method,* Swedish International Development Cooperation, Methods Development Unit, Stockholm.

Smith LC, Subandoro A. 2007. *Measuring Food Security Using Household Expenditure Surveys*. Food Security in Practic

Templeton DJ, Jamora N. 2010. Economic Assessment of a Change in Pesticide Regulatory Policy in the Philippines. *World Development* 38(10), 1519-1526.

The World Bank. 2012. *Designing a results framework for achieving results: a how-to guide*, Independent Evaluation Group & The World Bank, Washington, DC.

UNDP 2009. *Handbook on Planning, Monitoring and Evaluating for Development Results*, United Nations Development Programme, New York.

Walker T, Maredia M, Kelley T, La Rovere R, Templeton D, Thiele G, Douthwaite B. 2008. *Strategic Guidance for Ex Post Impact Assessment of Agricultural Research*. Report prepared for the Standing Panel on Impact Assessment, CGIAR Science Council. Science Council Secretariat, Rome, Italy.