

ILRI DISCUSSION PAPER

26

Evaluating value chain interventions: A review of recent evidence



RESEARCH
PROGRAM ON
Livestock and Fish

Evaluating value chain interventions: A review of recent evidence

Michael Kidoido¹ and Keith Child²

1. Economist, Impact Evaluation, International Livestock Research Institute

2. Principal Scientist, Impact Evaluation and Learning, ILRI

© 2014 International Livestock Research Institute (ILRI)



This publication is copyrighted by the International Livestock Research Institute (ILRI). It is licensed for use under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License. To view this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>. Unless otherwise noted, you are free to copy, duplicate, or reproduce, and distribute, display, or transmit any part of this publication or portions thereof without permission, and to make translations, adaptations, or other derivative works under the following conditions:

- ① **ATTRIBUTION.** The work must be attributed, but not in any way that suggests endorsement by ILRI or the author(s)
- ② **NON-COMMERCIAL.** This work may not be used for commercial purposes.
- ③ **SHARE ALIKE.** If this work is altered, transformed, or built upon, the resulting work must be distributed only under the same or similar license to this one.

NOTICE:

For any reuse or distribution, the license terms of this work must be made clear to others.

Any of the above conditions can be waived if permission is obtained from the copyright holder.

Nothing in this license impairs or restricts the author's moral rights.

Fair dealing and other rights are in no way affected by the above.

The parts used must not misrepresent the meaning of the publication. ILRI would appreciate being sent a copy of any materials in which text, photos etc. have been used.

Editing, design and layout—ILRI Editorial and Publishing Services, Addis Ababa, Ethiopia.

ISBN: 92-9146-353-1

Cover photos: ILRI/Jo Cadilhon

Citation: Kidoido, M. and Child, K. 2014. *Evaluating value chain interventions: A review of recent evidence*. ILRI Discussion Paper 26. Nairobi, Kenya: International Livestock Research Institute.

ilri.org

better lives through livestock

ILRI is a member of the CGIAR Consortium

Box 30709, Nairobi 00100, Kenya
Phone: + 254 20 422 3000
Fax: +254 20 422 3001
Email: ILRI-Kenya@cgiar.org

Box 5689, Addis Ababa, Ethiopia
Phone: +251 11 617 2000
Fax: +251 11 617 2001
Email: ILRI-Ethiopia@cgiar.org

Contents

Acknowledgements	iv
Abstract	v
Introduction	1
Challenges for evaluating impact of value chain approaches	2
Evaluation problems	4
Quasi-experimental study designs and estimation approaches	5
Experimental or Randomized Control Trials (RCTs) approach	8
Theory based approaches and mixed methods	8
Review of value chain impact evaluation studies	11
Review of the findings	11
Research designs, data types and empirical evaluation approaches	12
How were threats to validity addressed?	13
Lessons learned and ways forward	16
References	18
Appendix I List of reviewed value chain impact evaluation studies	21

Acknowledgments

We would like to acknowledge Nancy Johnson, Senior Research Fellow, International Food Policy Research Institute, for her numerous reviews that have greatly contributed to improving the quality of this discussion paper.

Abstract

Value chain interventions are rarely evaluated as rigorously as interventions in agricultural production or health. This is due to various reasons, including the intrinsic complexity of value chain interventions, intricate contextual support factors, presence of multilevel system actors, constant adaption to market and nonmarket forces and the cost associated with conducting an evaluation. This paper discusses a range of approaches and benchmarks that can guide future design of value chain impact evaluations. Twenty studies were reviewed to understand the status and direction of value chain impact evaluations. A majority of the studies focus on evaluating the impact of only a few interventions, at several levels within the value chains. Few impact evaluations are based on well-constructed, well-conceived comparison groups. Most of them rely on use of propensity score matching to construct counterfactual groups and estimate treatment effects. Instrumental variables and difference-in-difference approaches are the common empirical approaches used for mitigating selection bias due to unobservables. More meaningful value chain impact evaluations should be prioritized from the beginning of any project and a significant amount of rigor should be maintained; targeting a good balance of using model-based and theory-based approaches.

Key words: value chains, evaluation frameworks, impact evaluations

Introduction

Value chains are both operational and analytical models; they describe and analyse the vertical integration and disintegration of production and distribution systems (Roduner 2005; Roduner 2007). As operational models, value chains constitute a collection of entities and activities that characterize a production process. As most products are consumed away from the point of production, a value chain in the simplest sense constitutes all activities, institutions and entities involved in transforming, processing, transporting and adding value to the product before the product reaches the final consumer. Along the value chain, various actors exchange the ownership of raw materials, intermediate products and final products. These different actors are also linked by complex relationships, including demand for goods and services from each other.

As analytical models, value chains provide frameworks for analysing the movement of products from points of production to points of consumption. Specifically, they provide frameworks for examining how potential value chain actors might be barred from entering the value chain, the distribution of rents generated along the value chain, institutional and governance issues embedded in the value chain, upgrading along the value chain and how knowledge is acquired (Roduner 2007). Value chains are therefore appropriate for development interventions because they facilitate a clearer understanding of interactions among various actors and activities along product pathways. This ultimately allows development agents to: 1) increase efficiency and aggregate value generated along the value chain and 2) improve the relative share of benefits for various value chain actors.

Increasingly, development donors are applying substantive pressure on development agencies to prove that they are bringing about meaningful changes for intended beneficiaries (Deaton 2009; Sen 2012). Since the early 1990s, agricultural research programs, in particular, have come under intense pressure to prove that they are contributing to the goals of poverty reduction and environmental sustainability (Lilja et al. 2010). Consequently, a number of development agencies are searching for new and better approaches to delivering development aid. Value chain approaches (VCA) are widely considered good alternative approaches to delivering development aid, largely because VCAs involve several stakeholders. The appeal of VCAs is specifically linked to the 'leverage' they provide; value chains provide good platforms for delivering significant changes to thousands of beneficiaries, as opposed to just a few farmers or enterprises (Roduner 2007; Donovan and Dietmar 2010;). Moreover, VCAs are beneficial in that they can help identify the most critical entry points for development interventions. For instance, VCAs can encourage development actors who aim to improve the wellbeing of agricultural producers to consider both those on the agricultural production side and those on the agricultural inputs supply side, since the two are intimately linked in most agriculture value chains.

Value chain interventions, however, have not been immune to criticism. It has been argued that they deal predominantly with only a few selected well-to-do entrepreneurs and consequently cause insignificant changes to average poverty levels (Ton et al. 2011). Some authors, including Flores and Bastiaensen (2011) and Donovan and Dietmar (2010), have pointed out that value chain upgrading, consistent with most VCAs, comes at the cost of reducing female participation in the value chain. Moreover, there is a dearth of empirical evidence that value chains really can deliver development aid as effectively as their proponents claim. In light of these criticisms, it is increasingly clear that the creation of a convincing body of evidence to support VCAs is critical to their sustained application. Regrettably, not many value chain impact evaluations have been undertaken, documented or disseminated.

Structure of the paper

The purpose of this paper is to provide a strategic review of recent developments in approaches to value chain impact evaluation. The rest of the paper is structured as follows: The next section highlights the challenges of evaluating value chain approaches. Then, we review conventional impact evaluation approaches. We attempt to highlight the strengths and weaknesses of each method and the conditions under which a given method is used. We then present the most recent work on improving evaluations of VCAs, noting some benchmarks needed for well-designed value chain intervention evaluations. The ensuing section presents a review of 20 selected recent works on value chain impact evaluation. Here, we rely heavily on recent studies in order to judge the extent to which practitioners have attempted to address the weaknesses of conventional evaluation approaches. In the last section we discuss lessons learned from this review and how our work contributes to the literature on value chain impact evaluation.

Challenges for evaluating impact of value chain approaches

Humphrey and Navas-Alemán (2010) attributed the lack of impact evaluations to the complexity of value chains and the high cost of implementing evaluation studies. In a review of 40 value chain projects, they noted that most value chain projects did not follow up with well-designed impact evaluation studies, and also found that many evaluations focused only on reviewing and documenting project activities using qualitative methodologies to show their progress towards achieving intended outcomes.

There are other well-documented factors that impede robust value chain impact evaluation:

1. Market adaption: because value chains are continuously adapting to new market and non-market forces (Roduner 2007; Humphrey and Navas-Alemán 2010), evaluators face the challenge of designing impact methodologies that are suitable to rapidly changing dynamics. Traditional impact evaluation approaches generally aim at producing a static 'snapshot' that is inadequate when it comes to evaluating dynamic systems like value chains. Consequently, evaluation practitioners need to develop approaches that reflect fluctuations inherent in the value chain system; at a minimum, this requires multiple snapshots over the course of the project timeline (perhaps even in advance of the project start date).¹
2. Value chain interventions are typically time-, place- and commodity-specific (Ton et al. 2011). Interventions are thus unlikely to be repeated uniformly, and conclusions and generalizations made about the interventions will be less useful beyond specific value chains.
3. The specificity of value chain interventions makes it difficult to develop a monitoring and evaluation/impact evaluation framework that can leverage similarities across projects and minimize project costs. VCAs consist of several specific interventions, with each requiring a specific evaluation strategy; and yet they are usually implemented as a package.
4. Most intended outcomes of value chain interventions are complicated and influenced by several interconnected interventions, projects or policies besides those of the value chains themselves (Zandniapour et al. 2004; Ton 2012). Consequently, it is difficult to pin any single outcome to a specific value chain intervention.
5. Given the complexity of value chain interventions, the challenge that most value chain project managers face is balancing the allocation of resources to monitoring and evaluation and allocating resources to implementing the intervention itself. This balance needs to be achieved, since the evaluation of value chain effectiveness is important (Ashley and Mitchell 2008), yet it is one that most managers fail to achieve, preferring instead to allocate scarce resource overwhelmingly on project implementation.

¹. Also, constant changes in value chains lead to constant changes in the designs of value chain interventions, thus affecting the way project interventions are implemented and evaluated.

-
6. Finally (and related to the fourth factor) value chains are complex systems, heavily embedded in diverse social cultural settings which often influence how value chain interventions are implemented (Ashley and Mitchell 2008; Ton 2012). For the target group, however, the social, cultural and environmental impacts are as important as the tangible economic benefits derived from participating in the value chains. Yet the biggest challenge for evaluators has been measuring and quantifying these impacts both at the individual and community levels (especially using typical traditional impact evaluation approaches).

Evaluation problems

The aim of any impact evaluation exercise is to carefully measure the impact of a development intervention (value chain approach) on intended beneficiaries (value chain actors). This involves examining what would have happened to the entity (individual/enterprise) in the absence of the value chain intervention. It also involves establishing the behaviour of the entity both in the absence and presence of the intervention (causation claim). The challenge, however, is having the value chain actor in both states simultaneously (Asian Development Bank 2011; Cavatassi and González-Flores 2011). In fact, ‘true counterfactuals’ are never available (Alderman 2007; Deaton 2009). This is problematic because a rigorously designed evaluation project begins with an attempt to generate an acceptable counterfactual.²

The empirical problem for most impact evaluation approaches—and therefore for value chain evaluations—is to identify the ‘treatment effect’; i.e. the change in state of individuals who benefit from the intervention (treatment/participants) and those individuals who are similar to the beneficiaries in most aspects but are not exposed to the intervention (control/non-participants). It would seem that the non-participant group is a straightforward valid counterfactual group; however, the non-participation of the group might be a non-random outcome. Selection into participation might be based on unobservable factors which are simultaneously correlated with the outcome variable and bias the estimated size of impact (selection bias). Most impact evaluation approaches are, therefore, aimed at mitigating selection bias and rendering the non-participating groups a valid comparison group.

The choice of evaluation design depends on several factors, including treatment placement, study context, nature of interventions, ethical concerns and the timing of the evaluation study. Evaluation designs can be divided into three types: 1) quasi-experimental, 2) non-experimental and 3) experimental (also referred to as Random Control Trials, RCTs). Experimental and quasi-experimental designs are widely used in evaluating development projects, and have received considerable donor interest. Non-experimental designs that rely on qualitative methods are usually suitable for evaluating purely social programs (Getachew et al. 2011). Experimental designs largely rely on random assignment to treatment and control groups, whereas the quasi-experimental designs rely on constructed comparison groups. Quasi-experimental approaches are further categorized into: 1) before-and-after study designs (which focus largely on identifying the change in behaviour of economic agents participating in project activities before and after interventions), 2) with-and-without designs (which focus on tracking changes in behaviour of project non-participants and participants) and 3) several other approaches which rely on statistical modelling (in which treatment effects are estimated using parametric and non-parametric modelling).

In the following subsections, we present a detailed description of the strengths and weaknesses of quasi-experimental and experimental designs, closely following the categorization adopted by the Asian Development Bank (2011).

2. A counterfactual can be defined as a group of entities similar to the beneficiary group in all aspects apart from benefitting from the value chain intervention.

Quasi-experimental study designs and estimation approaches

Quasi-experimental evaluation designs rely on constructed comparison groups, or on the non-random assignment of units/individuals into treatment/control groups. In most cases, the comparison groups are constructed after the intervention has been implemented. For valid comparisons between control and treatment groups, however, the approach requires that the two comparison groups are as similar as possible with respect to key selected observable characteristics. The empirical estimation of the treatment effects is done using a variety of study designs and statistical modelling approaches (described below).

Before-and-after study designs

Before-and-after designs attempt to track changes in the behaviour of individuals participating in the study at two points in time (pre-intervention and post-intervention). The design requires that individual data—which vary with time and can also be tracked at two points in time—are available. The impact of the intervention (the treatment effect) is estimated using simple regression techniques, including Ordinary Least Squares (OLS). A regression model—in which the outcome variable of interest is specified as the dependent variable and time-varying variables and a time variable as the independent variables—is commonly used to estimate the treatment effects. The estimated parameter on the time variable represents the treatment effect, or the average effect of the intervention on participants (i.e. the average treatment effect (ATE)).

This study design assumes that the estimated treatment effect/impact is unbiased once all factors that influence the outcome variable are controlled for (Hulme 1997). The reality, however, is that in many cases observed and unobserved factors that have a significant impact on the outcome variables go unaccounted for; these factors end up biasing the estimated treatment effects. Factors such as covariant shocks, economic trends and influences of interventions from other development actors are likely to have a significant impact on the size and direction of the treatment effects. Unfortunately, the direction of bias is often not easy to predict (White 2007). Consequently, before-and-after designs are used or recommended only when the interventions are clearly defined or when the construction of a credible counterfactual group is not easy to achieve.

With-and-without study designs

The with-and-without designs compare changes in the behaviour of two groups (the participants and non-participants) at two points in time, most often between the pre-intervention and post-intervention periods. Assuming that the two groups are similar, the observed difference in the outcome variables between participants and non-participants and between the two periods is then considered as the impact due to the intervention. Two key conditions must exist for the treatment effects to be unbiased: 1) given a set of observable factors, the potential outcome for any individual should be independent of the probability of the individual participating in the intervention and 2) both participants and non-participants should have similar probabilities for any given value of a covariate (Deaton 2009).

Typically, the with-and-without designs are appropriate for cases where data—often cross-sectional and disaggregated by participation status—are available. Treatment effects are estimated using multivariable linear regression models such as OLS, Weighted Least Squares (WLS) or Instrumental Variables (IV) models. Estimates taken from linear regressions, however, normally suffer from specification biases (in which case the estimates are highly sensitive to variable exclusion and inclusion) (Asian Development Bank 2011). Moreover, linear regression methods impose the linear-in-parameter condition,³ which is unrealistic for certain treatment effects (Ravallion 2005). Semi-parametric methods, specifically the Propensity Score Matching (PSM) method, have therefore been widely adapted to estimate the treatment effects for with-and-without studies. Matching methods are attractive not only because they solve the specification bias problem but also because they do not impose any unnecessary conditions on model specification (Toledo and Carter 2010).

3. A condition which requires the independent variables to be linearly related with the outcome variables.

Propensity Score Matching (PSM)

Propensity Score Matching (PSM) is normally used for two purposes: 1) for constructing counterfactuals and 2) for estimating treatment effects. When used to estimate the treatment effect, the PSM estimator compares the change in the outcome of interest between participants and non-participants, conditional on a set of observable factors for both groups. PSM relies on two assumptions to address selection bias. The first assumption is the conditional independence (also referred to as the non-confoundedness assumption) which requires that potential outcomes/impacts are independent of how program participation is assigned (Rosenbaum and Rubin 1983). This assumption implies that selection is solely based on observables; thus, participation and outcome variables are observed simultaneously. The second assumption is the common support assumption, which requires that economic agents with similar characteristics should be equally likely to be participants or non-participants (i.e. they have similar probabilities of falling into either category).

The PSM approach can be problematic when matching is not done based on random observable characteristics. Nonetheless, a number of diagnostic checks are available to ensure that the two comparison groups are similar. For instance, one could consider checking for the sensitivity of the estimated treatment effects to small changes in model specification. Second, one can generate two counterfactual groups and make the first group a 'pseudo treatment' group and then compare the treatment effects of both groups. If a non-zero treatment effect for the 'pseudo treatment' group is observed, then one of the two groups is most likely the invalid counterfactual. The third diagnostic check involves the use of a variable that is most unlikely to be affected by the intervention. A non-zero effect on this variable implies that the counterfactual group is invalid. With or without a baseline, PSM can be used to estimate treatment effects. The PSM approach requires large sample sizes to ensure that bias due to non-random selection on observables is adequately addressed. Notably, matching on observables that are not truly random is likely to lead to estimates even more biased than those created by OLS, besides failing to eliminate the omitted variables bias. Matching methods are easier to implement where treatment variables are binary (Toledo and Carter 2010).

To address the problem of selection bias and thereby strengthen PSM, one can combine the approach with Weighted Least Square (WLS). In this approach, PSM is used to generate the propensity scores for the two comparison groups and then the calculated inverse propensity scores are used as weights in a regression model. The WLS approach is efficient even when the participants and non-participants are not randomly selected but are reasonably comparable (Khandker et al. 2010; Cavatassi and González-Flores 2011). In addition, unlike what happens when PSM is used alone, combining it with WLS generates consistent estimates of the Average Treatment Effects (ATE), since the approach utilizes information from all observations (Cavatassi and González-Flores 2011).

Difference-in-Difference (DD) design

The Difference-in-Difference (DD) approach estimates the change in the outcome of an intervention for two comparable groups (participants and non-participants), for two periods. Typically, data for the two comparison groups at two periods of time should be available for this method to be applied. The average treatment effect (ATE) is simply the difference in outcomes for the participant group between the pre-intervention period and the post-intervention period, and the outcomes of the non-participant group for similar time periods. For unbiased ATE estimates, the DD assumes that if the policy/intervention were not implemented, the outcomes for both groups would have been the same—this is the parallel trend assumption. The parallel trend assumption allows the DD approach to mitigate the effect of heterogeneity resulting from external shocks that affect both groups (Deaton 2009). Econometrically, the DD approach can also be achieved using a two-period panel model with treatment and time-fixed effects included. The ATE is estimated as the coefficient on the interaction term between time and treatment variables. The following diagnostic tests are suggested by Dulfo (2000) for validating the DD estimates:

1. Use data from the previous period (say, two years before the intervention), re-estimate the DD and compare with those estimated using data from one year before the intervention. If the generated DD is non-zero then the real DD is likely to be biased as well.

2. When several years of data are available, plot the average outcomes for the two groups and assess whether they are parallel or whether there is a jump just after the intervention.
3. Use alternative comparison groups. If the DDs of the alternative groups and the real one are different, it implies that the real DD is biased.
4. Replace the outcome variable with an alternative variable that is unlikely to be affected by the outcome of interest. If the DD of the alternative outcome variable is non-zero, then the real DD is also biased.

Regression Discontinuity (RD) design

The Regression Discontinuity design is not a common approach for evaluating development projects. It is mostly applied when assignment of treatment is based on a cut-off point for a continuous score of an index. It relies on the assumption that individuals just below the cut-off point and those just above the cut-off are very similar and only differ because of their respective treatment statuses. The group just below the cut-off point can be considered a good counterfactual for the one just above the cut-off point. As in PSM, comparison groups are first established and then the differences in outcome variables between the two groups are estimated as the average treatment effects (ATE). The RD estimator, however, faces similar problems as the PSM, and it is rarely used for evaluating development programs. In addition, the approach requires large sample sizes since it relies mostly on exploiting similarities around the cut-off point.

None of the empirical approaches reviewed above can adequately deal with selection bias due to unobservables. OLS, WLS, DD and PSM all rely on the assumption of exogeneity of the treatment variable. However, exogeneity of the treatment variable is unattainable because of several unobservable factors that affect outcomes and the treatment variables simultaneously (Duflo 2000). Hence, the Instrumental Variables (IV) approach introduced below, when used in combination with PSM or DD, can be a good approach for dealing with selection bias due to unobservables.

Instrumental Variables approach (IV)

An instrumental variable (IV) or instrument is a variable that affects an individual's participation in an intervention but does not directly affect outcomes (apart from through the individual's participation) (Deaton 2009). The IV approach exploits the existence of such a variable in mitigating selection bias due to unobservables. The advantage of the IV approach is that where good and valid instruments exist and are identified, the approach effectively addresses both biases due to unobservable and observables simultaneously (Duflo 2000). Also, the method can be used to test the exogeneity assumption, which forms the foundation for the matching and multivariate regression approaches previously reviewed (Cavatassi and González-Flores 2011). The IV approach, however, can lead to worse biases compared to an approach like OLS, especially when the identified instruments are not truly exogenous. Moreover, even true random instruments can be invalid (Duflo 2000; Deaton 2009; Cavatassi and González-Flores 2011). Hence, the toughest task with the IV approach is to find good and valid instruments.

Several approaches to finding valid instruments have been suggested. First, instruments can be identified through natural or quasi-experiments; i.e. where researchers exploit the fact that natural occurrences (such as floods) and policies could induce exogenous changes in some variables and yet are not themselves correlated with the participation status of individuals. Second, the IV-generated treatment effect can also be estimated as a Local Treatment Effect (LATE). LATE is mostly applied when the identified IV is found to maintain some degree of correlation with the unobserved anticipated outcome of participation, thereby affecting participation decisions (Khandker et al. 2010; Cavatassi and González-Flores 2011). In this case, the Intent-to-Treat (ITT) variable, representing the random assignment of treatment, is used as the instrumental variable. The program treatment effect caused by this instrument is also called the Local Average Treatment Effect (LATE) (Imbens and Angrist 1994).

Experimental or Randomized Control Trials (RCTs) approach

The core principles of Randomized Control Trials (RCTs) are randomization, control and comparison (Spillane et al. 2010). In designing RCTs, the evaluator's main aim is to randomly assign interventions, ensure that interventions are controlled so that there is minimal mixing of comparison groups and compare the groups to identify the effects of the interventions. Randomization implies that reasonable care is taken to ensure that every entity has an equal probability of being in either the treatment or control group. Thus, the suitability of RCTs for any evaluation study largely depends on how well the counterfactual groups are designed and managed. Usually when well designed, RCTs generate the least criticism because of the transparency with which the counterfactual groups are generated. RCTs are widely considered the 'gold standard' for impact evaluations (Deaton 2009). Also, they are the most appropriate approaches for evaluating single interventions constituting projects and for evaluating pilot projects intended for scale-up (Asian Development Bank 2011). Many of the empirical approaches used in the quasi-experimental evaluation approaches—for instance the DD and the IV—are also used in RCT designs to estimate treatment effects.

Nevertheless, like quasi-experimental approaches, RCTs also face randomization problems that stem from practical issues (Deaton 2009). Some of the implementation problems RCTs face include: high attrition rates threatening the internal validity of conclusions; threats to external validity resulting from RCTs' failure to capture general equilibrium effects (and consequently the failure of scaling-up results from pilot trials); failure to give a good representation of the distributional effects across entire populations (mostly relying on the average treatment effect); and threats to power due to overreliance on small samples given the high costs of implementing RCTs. Several remedies to these shortcomings have been suggested. For instance, targeting substantial sample sizes but balancing the cost of sampling and intervention implementation with randomizing at relatively aggregate levels could significantly increase the relevance of RCTs (Duflo 2000; Deaton 2009). Also, for RCTs to be relevant they must inform managers what works and what doesn't; this can be achieved by integrating qualitative methods with RCT designs. The approach greatly enhances evaluators' ability to identify instruments and clearly define the impact process.

In the following section, we discuss the extent to which these conventional empirical approaches have been combined with non-conventional impact evaluation approaches in assessing value chain interventions.

Theory based approaches and mixed methods

Value chains are complex systems in constant flux (Ton et al. 2011). Value chains are also heavily embedded in diverse social/cultural settings, and these settings greatly influence the implementation of value chain interventions. Moreover, as highlighted in the introduction section of this review, value chain interventions are time-, space- and commodity-specific. Although traditional impact evaluation approaches are efficient ways to solve common impact evaluation problems—such as bias due to non-observables, omission bias and specification bias—when used in isolation they fail to attribute the impacts that result from multiple value chain interventions. For this reason, traditional approaches need to be adapted to the complexities of value chain interventions. Hence, well-designed value chain impact evaluation studies should: 1) incorporate mechanisms for tracking constant changes in value chains, 2) incorporate appropriate mechanisms for tracking intervention-specific effects and 3) track the effects of individual interventions in VCAs by both time and place. Ton et al. (2011) and Ton (2012) have suggested several key benchmarks for guiding the design of a good value chain impact evaluation.

Ton et al. (2011) suggested that a good value chain evaluation design should be focused on answering three basic questions. First: Does the value chain approach work? Answering this requires finding evidence that shows that the approach is itself contributing to the positive outcomes gained by the intervention's beneficiaries. Second: How does the intervention work? Answering this requires that evaluators understand the specific processes and mechanisms through which value chain interventions deliver impact. And third: will the value chain intervention work in future?

Answering this requires using innovative and pragmatic approaches to examining whether the intervention can be replicated in other locations. Ton et al. (2011) pointed out that answering all these questions requires innovative ways of addressing common threats to the validity of conclusions generated from value chain impact evaluations. In value chain evaluations, the most common threats to valid conclusions are caused by poor measurement of outcome patterns, the failure to attribute impact to the various value chain interventions/actors and the challenge of generating generalizable value chain impact evaluation findings.

Ton et al. (2011) recommended that a well-designed value chain evaluation approach be based on a thorough analysis/examination of the value chain causal/logic model. White (2007), White (2009) and Sen (2012) have called this the 'theory based' approach. This approach relies on mapping out the pathways through which inputs are expected to achieve outcomes and then examining where the links at the various results levels are weak or missing (White 2007). An intervention is considered to be poor when links in the impact pathways are found to be missing or weak. Identifying missing or weak links, according to White (2007) and Sen (2012), should enable evaluators to have an enhanced understanding of cause and effect processes throughout the impact pathways. Also, having a clearer review of the causal model is critical to understanding key outcome indicators and their measurement. Approaches of this nature mainly exploit the Context-Mechanism-Outcome (CMO) framework (Pawson and Tilley 1997; Blamey and Mackenzie 2007). Other approaches in this category include: participatory evaluation, applied ethnographic evaluations and process-tracing designs. These approaches base their causal claims about impact on the processes linking program interventions with final outcomes, derived from either theory or from perceptions of stakeholders.

Regarding the question of whether value chain evaluation designs can identify the effects of VCAs, Ton et al. (2011) caution against reliance on methods that depend too heavily on regression modelling for RCTs and quasi-experimental designs. First of all, randomization (especially for RCTs) is impossible to achieve in value chain contexts, and even when it is attempted, many value chain stakeholders are locked out of the benefits of the interventions (White 2009; Ton et al. 2011). The second challenge of using RCTs in value chain impact evaluations is the difficulty of generating counterfactual groups, since value chain interventions often produce spill-over effects. According to Ton et al. (2011) quasi-experimental approaches could be more appropriate than RCTs; however, the former carry a great risk of non-random selection of treatments and hence the potential of compromising the validity of conclusions. Ton et al. (2011) also added that even the PSM model—which largely relies on exploiting all information related to the characteristics of the comparison groups—has been considered inadequate for use in value chain impact evaluations. Their argument derives from the fact that most of times the PSM model is constructed after fieldwork, and so the process often relies on limited information (especially when generating comparison groups).

White (2007) and Ton et al. (2011) proposed mixing methods and employing more 'ex ante theorization' to mitigate common threats to valid conclusions, particularly with value chain interventions. White (2007) called it 'using good contextualization to build confidence around results'. First, they propose relying on methods based on statistical analysis of patterns within case studies, instead of exclusively using regressions. They claim that such approaches respect the integrity of cases as systems that can also lead to unique outcomes given different combinations of variables/factors. They cite cluster analysis, contrasting case methods and qualitative comparative analysis as some examples of case-based approaches that can enrich regression-based approaches. These methods, when used in combination with regression methods, should help strengthen conclusions derived from value chain impact evaluations. Both studies also propose the use of mixed methods. White (2009), Ton (2012), however, cautioned that implementing evaluation designs based on a mixture of approaches requires creative thinking; Ton (2012) argued that it is not just about combining qualitative and quantitative methods in the traditional way. Ton (2012) argued for a strategic combination of qualitative and quantitative tools that possess different strengths and weaknesses in generating strong evidence to support conclusions. In an attempt to forecast the most likely threats to valid conclusions, analytical procedures, tools and methods have to be reviewed during the initial research designs (White 2009; Ton 2012).

In dealing with the common threats to the external validity of value chain evaluation findings, Ton et al. (2011) advised evaluators to rely on two key approaches. First, examine the differences and similarities between the intervention and the replication contexts. Next, focus on explaining the contextual differences that are relevant for replicating conclusions. Ton (2012) recommended that evaluators should heavily draw on case studies while making recommendations for replication of value chain interventions. Inclusion of well-structured and well-designed case studies, Ton (2012) believed, can provide good evidence of interventions that are likely to work in comparable contexts.

Review of value chain impact evaluation studies

Several criteria were used for selecting the evaluation studies included in this review. The overarching criterion, however, was that the studies had to be related to specific stages or to all stages of an agriculture value chain intervention. Based on this first criterion, over 50 evaluation studies were identified; however, many of these studies were unpublished, old or did not include good documentation of the evaluation methodology. Other criteria were adopted which reduced the number of studies to 20 (see Appendix 1). First, we restricted our selection to evaluation studies that were under three years old, which helped ensure that we were capturing the most recent developments in value chain impact evaluations. However, one study that was approximately four years old was included because of its strong value chain focus and its consistent documentation of the evaluation methodology. Other selection criteria included well-documented evaluation methodologies and the requirement that evaluations were done in Africa, Latin America, Asia, or in a transition economy.

In the 20 studies examined below, we identified the evaluation designs used (i.e. quasi-experimental, experimental or non-experimental), data and empirical approaches followed in estimating the treatment effects and how the studies handled threats to the validity of conclusions (statistical, internal, construct and external validity). We conclude the section with an evaluation of the extent to which the selected studies attempted to incorporate some of the benchmarks for well-designed value chain impact evaluations highlighted in the previous section.

Review of the findings

Of the 20 studies reviewed, 17 were designed as quasi-experimental studies, 3 were non-experimental studies and none were based on a purely experimental design. Among the 17 quasi-experimental studies, 5 studies used the difference-in-difference (DD) approach, 4 used the Instrumental Variables (IV) approaches in their varied forms, 2 were based on the with-and-without design, 1 was designed as a before-and-after study and 5 were based on a combination of approaches (see Appendix 1). The three non-experimental evaluations relied on qualitative approaches in establishing the changes in selected outcomes—specifically, qualitative changes in the five recognized livelihood capital assets—between participants and non-participants and between pre-intervention and post-intervention periods. Notably, study participants were not randomly selected.

Two of the 3 no-experimental studies adopted the Sustainable Livelihoods Framework (SLF) as their guiding framework for tracking changes in the behaviour of participants and non-participants over the two evaluation periods. The study by Zossou et al. (2012) in Benin used a participatory impact evaluation approach (the sustainable livelihoods approach) to track the impact of a video on rice parboiling technology (developed through participatory approaches on the livelihoods of women rice processors). Similarly, Katerberg et al. (2011) used the livelihoods-based impact assessment approach to evaluate the effect of a horticulture value chain project on women in Afghanistan. In Katerberg et al. (2011) qualitative changes in capital assets were adapted as proxies for measuring changes in poverty and well-being—and, therefore, impact. The three non-experimental studies were, however, also designed as before-and-after studies, since only information pertaining to the same respondents for two evaluation periods (pre-intervention and post-intervention) was used in the analysis.

Generally, we observe that the use of purely Randomized Control Trial (RCT) designs for assessing the impact of value chain approaches is still low. In several instances, significant changes to program implementation led to evaluation studies being changed from RCTs to quasi-experimental designs. Nonetheless, we also observe significant differences among the various quasi-experimental approaches that were used.

Research designs, data types and empirical evaluation approaches

The empirical approaches used by many of the evaluations were dictated by the research designs and type of data available to evaluators. For instance, Cavatassi and González-Flores (2011), Getachew et al. (2011) and Kamau et al. (2011) were designed as with-and-without studies. These evaluations relied on single-farm-visit data and focused mostly on households/individuals located in both intervention and non-intervention areas. Two of these three studies used Propensity Score Matching (PSM) to generate the required counterfactual groups and to estimate the treatment effects. Cavatassi and González-Flores (2011), however, after generating the counterfactual groups using PSM, used a combination of multivariate regression approaches to estimate the treatment effects.

Cosyns et al. (2011), Katerberg et al. (2011) and Zossou et al. (2012) used the before-and-after approach and focused on tracking changes in program/project participants alone. Both qualitative and quantitative data were used, although most of the quantitative data were collected using single-farm-visit surveys and were based on recall/retrospective data collection methods. A variety of simple non-parametric and parametric methods of mean comparisons were later applied to estimate the effect of program interventions on participants. In these studies, we observed that the greatest threat to the validity of their conclusions was a heavy reliance on recall/retrospective data collection methods (which are highly unreliable).

Jones and Gibbon (2011), Oduol et al. (2011), Carter et al. (2012) and Blair et al. (2012) used Instrumental Variables (IV) approaches in varied ways. Choices of selected IV approaches were in part determined by the type of data available, the nature of the intervention evaluated and potential for spill-over effects. For instance, Oduol et al. (2011) used cross-sectional data and applied the IV to estimate the Local Average Treatment Effects (LATE). On the contrary, Jones and Gibbon (2011), Blair et al. (2012) and Carter et al. (2012) had access to two-period data collected in various ways, and also used the IV approach in varied ways. Blair et al. (2012) and Carter et al. (2012) used the randomized phase-in design, where participation of randomly-selected groups of potential beneficiaries was delayed and eventually the late entrants were used as the counterfactual for the group that received interventions early in the project's implementation. In both Oduol et al. (2011) and Carter et al. (2012) the IV approach was instrumental in investigating and dealing with spill-over effects. For this, the LATE or Intent-To-Treat (ITT) estimator was used. In Carter et al. (2012), however, a quantile regression approach was also used to investigate the heterogeneity of the treatment effects across various categories of participants.

Baulch et al. (2009), Burki (2010), Bonilla and Cancino (2011), ISSER (2012) and Waarts et al. (2012) used the Difference-in-Difference approach. For these studies, multiple period data were available and attempts at creating counterfactual groups were made early in the projects. For instance, in ISSER (2012) a randomized phase-in design was used and three surveys were implemented: the first was exclusively done with early entrants and the two others were done with both early and late entrants. On the other hand, Baulch et al. (2009), Burki (2010), Bonilla and Cancino (2011), and Waarts et al. (2012) used two-period data. Across the five studies, various empirical strategies were used to estimate the Average Treatment Effects (ATEs). For instance, Burki (2010), ISSER (2012) and Waarts et al. (2012) used the difference-in-difference approach in its simplest form. This meant that ATEs were estimated as the differences in the outcome variables for the participation group between the pre-intervention period and the post-intervention period and for the non-participation group for similar evaluation periods. Bonilla and Cancino (2011), however, first used PSM to generate the counterfactual group and then estimated the treatment effects using the DD approach. On the other hand, although Baulch et al. (2009) used the conventional DD approach, they reweighted households based on their probability of participation. The DD approach is assumed to lead to robust results,

especially when sample selection biases are expected to be low or are adequately covered (which seems to have been the case with most evaluations that used this approach).

The last group of studies—Development Alternatives Inc. (2010), Forston et al. (2012), NORC (2012) and Oxfam GB (2012)—used a mixture of evaluation approaches. Despite the use of a mixture of designs, they all relied on longitudinal data for two comparison groups (treatment and control). For instance, Forston et al. (2012) used a randomized phase-in design and used regression modelling to estimate the treatment effects. In addition, Forston et al. (2012) used evidence from qualitative approaches to support the findings of regression modelling. NORC (2012) was designed as quasi-experimental in which several regression models were estimated in an attempt to strengthen the validity of the results. NORC (2012) combined PSM, IV and several weighted multivariate regression models (with the weights being the probability of participation). Development Alternatives Inc. (2010) was initially designed as an RCT; however, due to significant changes in program design that greatly affected the reliability of the RCT design, the study was adapted to the quasi-experimental design. Evaluators used simple analyses—including the analysis of mean differences—and used evidence from qualitative data to estimate the impact of project interventions (mostly at the outcome level).

Overall, PSM is identified as the most common empirical tool, serving several purposes across evaluation designs and empirical estimations of treatment effects. PSM has been used both for generating counterfactual groups and for estimating treatment effects (i.e. the PSM estimator). When PSM was used to generate the comparison groups, it was often combined with the DD or the IV approaches to improve the estimation of treatment effects. However, for evaluations where relatively good counterfactual groups were established early in the project planning phase and where no significant spill-over effects were encountered, the DD approach was found adequate and used exclusively.

How were threats to validity addressed?

Most of the evaluations we reviewed concentrated on addressing threats to statistical and internal validity. In other words, most studies concentrated on reducing bias resulting from poorly constructed comparison groups and non-random sample selections. Two studies also explored the heterogeneity of treatment effects on target groups. In addition, a few studies dealt with bias arising from spill-over effects, especially when individuals within selected participating communities chose not to participate in project interventions. When spill-over effects are not well addressed they can lead to under- or overestimation of project impacts (Imbens and Angrist 1994).

We assessed how the studies addressed threats to validity, focusing on the extent to which the studies constructed and used counterfactual groups. Notably, the weakest focus on generating reliable comparison groups was among the before-and-after approaches (i.e. Cosyns et al. 2011; Katerberg et al. 2011; Zossou et al. 2012). Similarly, in several evaluations counterfactual/control groups were not designed at the inception of the projects, but rather at the time of implementing the impact evaluations (i.e. *ex post*). This group of studies primarily consisted of evaluations designed as with-and-without studies, and included Cavatassi and González-Flores (2011), Getachew et al. (2011) and Kamau et al. (2011). For these studies, the Propensity Score Matching (PSM) approach was used to construct the counterfactual/comparison groups. PSM was used to create a common support or a matched group of participants and non-participants who were similar in all aspects (apart from the fact that the participants received the interventions).

The common support was then used to estimate differences in selected outcome variables across the two groups. In a few studies, the PSM-generated counterfactual groups were further verified at the community level using Focus Group Discussions (FDGs) and Key Informant (KI) interviews. For instance, Cavatassi and González-Flores (2011) and Waarts et al. (2012) went back to the community and used FDGs and KI interviews to verify the representativeness of selected counterfactual villages. In a few cases where PSM was used as an estimator—including Getachew et al. (2011) and Kamau et al. (2011)—a number of robustness checks were further done on the PSM estimator. These checks included: 1) bootstrapping the standard errors, 2) testing for the sensitivity of the estimator to unobservable variables that could have affected household decisions, 3) using various balancing methods for generating the comparison

groups and 4) using a dummy confounder to test the robustness of the estimated treatment effects (as applied in Getachew et al. 2011).

The strongest emphasis on generating and maintaining good counterfactual groups was found in studies that eventually adapted the DD approach. For some studies in this group—including Development Alternatives Inc. (2010) and NORC (2012)—the original construction of the comparison groups were abandoned because of significant changes in overall program implementation designs. Ultimately, these studies largely relied on qualitative approaches to track the effect of the interventions.

Potential biases resulting from selection on unobservables—especially where reliable counterfactuals could not be generated—were addressed using the IV approach. The generic approach was common, as was the case in Cavatassi and González-Flores (2011). Here they used an instrumental variable and then used a two-stage least squares (2SLS) approach to estimate the treatment effects. Other variants of the IV approach were also used. For instance, Jones and Gibbon (2011) adapted the IV procedure by substituting fixed effects with a set of location-continuous proxies (since the fixed effects were correlated with the participation variable). In Oduol et al. (2011), the Local Area Treatment Effect (LATE) estimator was estimated using IV. In both cases, however, the role of the IV was largely to induce a change in the behaviour of the participants in such a way that the IV would have an effect on the outcome. With the IV assumed to be random in the population, the LATE estimator is considered the mean impact of the IV on the outcome variable—and thus the impact of the intervention. It is generally argued that the IV approach deals more efficiently with all sources of bias due to both observables and unobservables (Ravallion 2005; Cavatassi and González-Flores 2011). The IV approach was also used for dealing with the potential bias due to individuals located in participating communities opting not to participate in project activities. In this case, a third group of individuals located within the participating community but who never themselves participated (Intent-to-Treat group) was modelled and used in several ways to investigate the impact of spill-overs. When the Intent-to-Treat group emerged to be similar to the treated, the two were merged into one treatment group and the ATE estimated using the basic IV approach. This is well illustrated in Cavatassi and González-Flores (2011) and Jones and Gibbon (2011).

Among the 20 evaluations only Blair et al. (2012) and Carter et al. (2012) attempted to investigate the heterogeneity of the treatment effects across the sampled population. Carter et al. (2012) used the generalized quantile regression analysis to compare the treatment effect between high performers and low performers within the study population. They focused on testing whether treatment effects were similar across the entire population, thus ensuring that results were generalizable. Similarly, Blair et al. (2012) used several exploratory approaches to test the stability of the estimated treatment effects across different segments of the population.

We further assessed the extent to which these studies incorporated some of the benchmarks suggested by Ton et al. (2011) and Ton (2012). First, of the 20 studies only 7 cases attempted to consistently use a mixture of qualitative and quantitative approaches. These included Baulch et al. (2009), Cavatassi et al. (2009), Development Alternatives Inc. (2010), Blair et al. (2012), Forston et al. (2012), Oxfam GB (2012) and Waarts et al. (2012). Most of these evaluations (5 of the 7), however, largely relied on the using the traditional Participatory Rural Appraisal (PRAs) tools—including Focus Group Discussions, Key Informant interviews and participant observations—to include qualitative analyses in the evaluations. On the other hand, Blair et al. (2012) effectively combined both qualitative and quantitative methods to validate the treatment effects. First, they categorized a subsample of participants (selected from the overall impact evaluation sample) into three groups: those who had experienced a high increase in income, those who experienced a modest increase and those who experienced no increase. They then conducted in-depth qualitative follow-up interviews with individuals in these subsamples to validate the estimated impacts. Second, they undertook additional analyses to provide a context for the impact estimates. For instance, they analysed the changes in consumer prices during the evaluation period and also did sensitivity tests to examine the robustness of the estimates. Blair et al. (2012) also undertook several exploratory analyses using various hypothetical levels of benefitting from project interventions. They also explored whether large estimated impacts were associated with other factors (such as belonging to farmer groups/cooperatives or to specific regions or departments).

Our review notes that only 7 of the 20 studies developed and applied logic/causal models. Cavatassi et al. (2009), Development Alternatives Inc. (2010), Toledo and Carter (2010), Blair et al. (2012), Forston et al. (2012), Oxfam GB (2012) and Waarts et al. (2012) relied on causal/logic models in defining the basic hypotheses and research questions. The casual/logic models were built early in the life span of these projects and they became the basis for developing the impact evaluation plans.

Lessons learned and ways forward

This paper has reviewed recent developments in value chain impact evaluation designs and methodologies. We revealed that very few value chain approaches have been evaluated for their impact on intended beneficiaries; moreover, the few that have been evaluated have not been well documented. We conclude that many value chain development projects prepare impact evaluation frameworks but do not implement them, and where they are implemented they focus on only a few stages of the value chain. In addition, we could identify only a few instances in which changes to conventional impact evaluation methods occurred. The majority of value chain studies follow quasi-experimental designs, as opposed to purely Randomized Control Trials (RCTs). Notably, the difference-in-difference (DD) and propensity score matching (PSM) approaches—either used exclusively or in combination—have been the most common empirical approaches applied to quasi-experimental designs. We noted limited proper development and use of counterfactual groups. In the majority of cases counterfactual groups were poorly conceived and poorly constructed.

In order to increase the numbers and quality of value chain impact evaluations, the following need to be emphasized:

1. To minimize the complexity and expense of value chain impact evaluations, value chain impact evaluation frameworks should be prioritized right from project inception and incorporated in project implementation as early as possible. Early consideration of value chain impact evaluations permits the adequate allocation of resources for impact evaluations.
2. It is nearly impossible to draw conclusions about the impact of value chain interventions without linking the changes in the value chain and the value chain actors to the component interventions. Proper construction and use of counterfactual groups—particularly at the initial stages of the value chain projects—can facilitate the identification of this linkage. Even imperfect counterfactual groups can provide valuable information for identifying the impact of value chain interventions, especially when qualitative approaches are deliberately incorporated into the study designs.
3. Quasi-experimental evaluation frameworks are more adaptable to value chain approaches. As such, evaluators should do the following: generate credible counterfactuals/comparison groups early in the project cycle; capture data at a minimum of two stages of project implementation; and ensure that relatively large sample sizes are used. At the design and analysis stages, evaluators should anticipate potential selection biases and determine how they can address them using statistical modelling. They should also understand and anticipate the potential effects of spill-overs. Above all, evaluators need to be aware of and plan to address the challenges of defining the research focus groups ('participants'), since most value chains are normally composed of several interventions aimed at various levels of the value chains.
4. Although we found scant evidence of the use of mixed methods in value chain impact evaluations, combining qualitative and quantitative approaches to data collection and analysis is preferable. Qualitative approaches provide an in-depth understanding of the impact processes and highlight the linkages between interventions and impacts. Evaluators should identify specific levels on the value chain where alternative methods can usefully be applied. Incorporating qualitative approaches should not be limited to merely using traditional Participatory

Rural Appraisal (PRA) tools in data collection; it should also include the innovative use of qualitative information to support estimated treatment effects. Furthermore, the use of causal/logic models is instrumental in clarifying research questions, identifying where the alternative qualitative and quantitative methods can be applied and monitor progress on key outcomes. Combining methods can also enhance the validity of findings and strengthen conclusions as equally emphasized by Stern et al. (2012).

References

- Alderman, H. 2007. Improving nutrition through community growth promotion: Longitudinal study of nutrition and early childhood development program in Uganda. *World Development* 35(8):1376–1389.
- Ashley, C. and Mitchell, J. 2008. *Doing the right thing approximately not the wrong thing precisely: Challenges of monitoring impacts of pro-poor interventions in tourism value chains*. London, UK: Overseas Development Institute.
- Asian Development Bank. 2011. *A review of recent developments in impact evaluation*. Mandaluyong, Philippines: Asian Development Bank.
- Baulch, B., Marsh, J., Linh, N.B., Linh, L., Trung, H. and Hoàng, V. 2009. *Key findings from the second Thanh Hoa bamboo survey: Assessment report*. Hanoi, Vietnam: Prosperity Initiative.
- Blair, R., Campuzaano, L., Moreno, L. and Morgan, S. 2012. *Impact evaluation findings after one year of the productive and business services activity of the productive development project, El Salvador*. Princeton, USA: Mathematica Policy Research.
- Blamey, A. and Mackenzie, M. 2007. Theories of change and realistic evaluation: Peas in a pod or apples and oranges. *Evaluation* 13(4):439–455.
- Bonilla, C.A. and Cancino, C.A. 2011. *The impact of the seed capital program of SERCOTEC in Chile*. Washington, DC, USA: Inter-American Development Bank.
- Burki, A.A. 2010. *Program evaluation of agribusiness support fund: Estimating the effects of treatment on farmer groups, agribusinesses and BDS market in Pakistan*. Lahore, Pakistan: Lahore University of Management Sciences.
- Carter, M.R., Toledo, P. and Tjernstrom, E. 2012. *The impact of business services on the economic well-being of small farmers in Nicaragua*. California, USA: University of California, Davis.
- Cavatassi, R. and González-Flores, M. 2011. Linking smallholders to the new agricultural economy: The case of the Plataformas De Concertación in Ecuador. *Journal of Development Studies* 47(10):1545–1573.
- Cavatassi, R., Gonzalez, M., Winters, P., Andrade-Piedra, J., Espinosa, P. and Thiele, G. 2009. *Linking smallholders to the new agricultural economy: An evaluation of the Plataformas program in Ecuador*. ESA Working Paper 09-06. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Cosyns, H., Degrande, A., de Wulf, R. and van Damme, P. 2011. Can commercialization of NTFPs alleviate poverty? A case study of *Ricinodendron heudelotii* (Baill.) Pierre Ex Pax. kernel marketing in Cameroon. *Journal of Agricultural and Rural Development in the Tropics and Subtropics* 112(1):45–56.
- Deaton, A. 2009. *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*. NBER Working Paper No. 14690. Cambridge, USA: National Bureau of Economic Research.
- Development Alternatives Inc. 2010. *Profit Zambia impact assessment*. Washington, DC, USA: United States Agency for International Development.
- Donovan, J. and Dietmar, S. 2010. *An asset-based approach for assessing the impact of value chain approaches on rural poverty: Methodological guidelines for development practitioners and private sector representatives*. Turrialba, Costa Rica: CATIE.
- Duflo, E. 2000. *Empirical methods*. MIT Lecture Notes. Cambridge, USA: Massachusetts Institute of Technology.
- Flores, S. and Bastiaensen, J. 2011. Gender bias in dairy value chains in Nicaragua. *Enterprise Development and Microfinance* 22(3):196–209.

- Forston, K., Rangarajan, A., Blair, R., Lee, J. and Valentine, G. 2012. *Evaluation of water-to-market training in Armenia*. Princeton, USA: Mathematica Policy Research.
- Getachew, Y., Jaleta, M. and Gebremedihin, B. 2011. Impact of input and output market development interventions on input use and net income of households in Ethiopia. *Journal of Economics and Sustainable Development* 2(9):22–33.
- Hulme, D. 1997. Impact assessment methodologies for microfinance: A review. Presentation given at the virtual meeting of the CGAP working group on Impact Assessment Methodologies April 17–19, 1997. Manchester, UK: Institute for Development Policy and Management.
- Humphrey, J. and Navas-Alemán, L. 2010. *Value chains, donor interventions and poverty reduction: A review of donor practice*. Vol. 2010. Brighton, UK: Institute of Development Studies, University of Sussex.
- Imbens, G. and Angrist, J. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–476.
- ISSER (Institute of Statistical, Social and Economic Research). 2012. *An impact evaluation of the MIDA FBO training*. Legon, Ghana: Institute of Statistical, Social and Economic Research, University of Ghana.
- Jones, S. and Gibbon, P. 2011. Developing agricultural markets in sub-Saharan Africa: Organic cocoa in rural Uganda. *Journal of Development Studies* 47(10):1595–1618.
- Kamau, M., Lawrence, M., Ricardo, F. and Ruerd, R. 2011. *The impact of certification on smallholder coffee farmers in Kenya: The case of UTZ program*. Nairobi, Kenya: Tegemeo Institute of Agricultural Policy and Development.
- Katerberg, L., Khan, A. and Ruddick, S. 2011. Evaluating value chain impact using a sustainable livelihoods approach: A case study on horticulture in Afghanistan. *Enterprise Development and Microfinance* 22(3):225–240.
- Khandker, S., Koolwal, G. and Samad, H. 2010. *Handbook on impact evaluation: Quantitative methods and practices*. Washington, DC, USA: World Bank.
- Lilja, N., Kristjanson, P. and Watts, J. 2010. Rethinking impact: Understanding the complexity of poverty and change—Overview. *Development in Practice* 20(8):917–932.
- NORC. 2012. *Impact evaluation of the farmer training and development activity in Honduras*. Vol. 01. Chicago, USA: NORC at the University of Chicago.
- Oduol, J., Binam J.N., Olarinde, L. and Diagne, A. 2011. Impact of adoption of soil and water conservation technologies on technical efficiency: Insight from smallholder farmers in sub-Saharan Africa. *Journal of Development and Agricultural Economics* 3(14):655–669.
- Oxfam GB. 2012. Guatemala highland value chains development alliance: Project effectiveness review. Oxford, UK: Oxfam.
- Pawson, R. and Tilley, N. 1997. *Realistic evaluation*. Los Angeles, USA: Sage.
- Ravallion, M. 2005. *Evaluating anti-poverty programs*. World Bank Policy Research Working paper 3625. Washington, DC, USA: World Bank.
- Roduner, D. 2005. Valuechains: What is behind this ‘new’ key word? And what is the role of development agencies in value chain development in developing countries? *Rural Development News* 2/2005.
- Roduner, D. 2007. Donor interventions in value chain development. Berne, Switzerland: Swiss Agency for Development and Cooperation (SDC).
- Rosenbaum, P. and Rubin, D. 1983. The central role of the propensity score in observational studies for casual effects. *Biometrika* 70(1):41–55.
- Sen, N. 2012. *Attribution: Measuring attributable change caused by a programme*. Cambridge, UK: The Donor Committee for Enterprise Development.
- Spillane, J.P., Pareja, A.S., Dorner, L., Barnes, C., May, H., Huff, J. and Camburn, E. 2010. Mixing methods in randomized controlled trials (RCTs): Validation, contextualization, triangulation, and control. *Educational Assessment, Evaluation and Accountability* 22(1):5–28.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befni, B. 2012. *Broadening the range of designs and methods for impact evaluations*. Department for International Development Working Paper No. 38. London, UK: DFID.
- Toledo, E.P. and Carter, R. 2010. *Impact of business services on the economic wellbeing of small farmers in Nicaragua*. Basis Brief: Assets and Markets Access (CRSP). Madison and Davis, USA: BASIS Assets and Market Access Collaborative Research Support Program.

-
- Ton, G. 2012. The mixing of methods: A three-step process for improving rigour in impact evaluations. *Evaluation* 18(1):5–25.
- Ton, G., Vellema, S. and de Ruyter de Wildt, M. 2011. Development impacts of value chain interventions: How to collect credible evidence and draw valid conclusions in impact evaluations? *Journal on Chain and Network Science* 11(1):69–84.
- Waarts, Y., Ge, L., Ton, G. and Jansen, D. 2012. *Tea production in Kenya: Impact assessment of two training models*. New York, USA: Rainforest Alliance.
- White, H. 2007. *Impact evaluation: The experience of the independent evaluation group of the World Bank*. Washington, DC, USA: The World Bank.
- White, H. 2009. *Theory-based impact evaluation: Principles and practice*. Working paper no. 3. New Delhi, India: International Initiative for Impact Evaluation.
- Zandniapour, L., Sebstad, J. and Snodgrass, D. 2004. *Review of impact assessments of selected enterprise development projects*. Washington, DC, USA: United States Agency for International Development.
- Zossou, E., van Mele, P., Wanvoeke, J. and Lebailly, P. 2012. Participatory impact assessment of rice parboiling videos with women in Benin. *Experimental Agriculture* 48(03):438–447.

Appendix I List of reviewed value chain impact evaluation studies

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Cavatassi et al. 2011. Linking smallholder potato farmers to the markets: Impact study of multi-stakeholder platforms in Ecuador	Ecuador	Formation of multi-stakeholder platforms for linking smallholder potato farmers to high-value agricultural markets. Helping farmers to meet the needs of high-value markets. They also provided training through Farmer Field Schools (FFS) to enhance productivity and to promote IPM	To increase yields and profits of smallholder potato growers in order to reduce poverty and improve food security	To understand whether and to what extent participating in platforms affected potato farmers' well-being through enhanced earnings from potato production	All farmers in treatment communities were eligible Used Propensity Score Matching to construct the counterfactual group Sites and communities were selected using the Ecuadorian census data	Evaluation targeted those farmers in the treatment communities who participated in the platforms Key indicators assessed Target crop yield Gross margins Use and practice of knowledge of agrochemical application	Study was with and without design, which also used cross-sectional data. Used Propensity Score Matching (PSM) for the counterfactual Used multivariate regression modelling (including Ordinary Least Squares (OLS), Propensity Score, Weighted Least Squares (WLS) and Instrumental Variables (IV)) to estimate treatment effects	The platforms had a positive impact on the well-being of participating farmers. Both yields and gross margins increased for benefitting farmers
Getachew et al. 2011. Impact of input and output market development interventions on input use and net income of households in Ethiopia	Ethiopia	Productivity and Market Success (IPMS) project implemented by ILRI, focused on selected crop and livestock value chains development through Input and output markets development Increasing cooperatives' access to credit to supply farmers with quality inputs in a timely manner Building capacity of extension workers and farmers	To improve farm productivity and production through market-oriented development as a driver of improved and sustainable livelihoods for rural populations	Assessing the impact of input and output market development interventions of the IPMS project on input use, productivity and total net income of participating households	Households located in Peasant Association (PAs), where IPMS activities were implemented	Households that participated in IPMS activities: Key indicators assessed Input use intensity Crop productivity Total net income	With and without evaluation design and used cross-sectional data. Used propensity Score Matching (PSM) estimator to estimate the treatment effects	Project interventions resulted in positive and significant effects on level of input use, productivity and total net income of treated households

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Kamau. 2011. The impact of certification on smallholder coffee farmers in Kenya: The case of UTZ certification program	Kenya	Implementation of the UTZ coffee certification program	To advocate and promote good practices to protect coffee consumers, the environment and the producers	To estimate the impact of certification on income, wealth and expenditures of households To assess the changes in farm households' perceptions of their economic situation, willingness to invest, risk attitude and royalty to their cooperatives arising from the certification programs	The treatment group was of farmers who belonged to UTZ-certified coffee cooperatives	Farmers participating in the UTZ certification program Indicators assessed: Coffee prices Coffee incomes	With and without evaluation design, data collected using a single farm visit survey and applied Propensity Score Matching (PSM) to estimate the treatment effects	The impact varied by region. It ranged from increased coffee prices and incomes to increased access to credit and to increased household income
Cosyns et al. 2011. Can commercialization of Non-Timber Forest Products (NTFPs) alleviate poverty?	Cameroon	The ICRAF marketing project activities included: Setting up institutional marketing arrangements Development of market information systems Provision of technical support for product processing Establishment of village nursery to support domestication and planting of trees	The project aimed to increase, diversify and stabilize incomes of poor small-scale farmers through increased participation in the benefits from the agroforestry tree products value chain	To determine the extent to which project interventions improved the financial situations of households in project villages	Farmers who belonged to the villages where the ICRAF tree product marketing project and partners were active	Households that participated in activities to commercialize <i>Ricinodendron heudelotii</i> (Bail. Pierre ex Pax., (njansang) Indicators assessed: Household income	Designed as a before and after evaluation, data gathered using retrospective data collection methods Differences in njansang incomes between control and project farmers and between pre-project and after project periods were compared. They also applied subjective parameters based on a 5-point Likert scale	Promotion of njansang had a positive impact on total cash income of participating farmers

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Zossou et al. 2012. Participatory impact evaluation of the rice parboiling videos on women in Benin	Benin	Zooming-in-zooming-out approach combined development of a rice parboiling technology with dissemination via videos. Aimed to improve the nutrition and market quality of marketed rice	Increasing the use of the rice parboiling technology developed with rice processors to enhance the nutritional value of rice and reduce grain breakage rate at milling	To assess impact of the farmer-to-farmer training video 'Cashing in with parboiled rice' on women's livelihoods in Benin	Treatment groups were women rice processors who resided in municipalities where rice parboiling was a tradition. Control group were women residing in communities located about 12 kms from where the video was screened	Target group was women in the selected video villages who watched the video Indicators assessed: Changes in five capital assets, i.e. human, physical, natural, financial and social capital	Evaluation was designed as a before and after study, using the Sustainable Livelihood Framework to assess changes in the five capitals. Used recall methods of data collection to capture data	Women who resided in the video villages and watched the videos experienced significant changes in their livelihoods (five capitals)
Katerberg et al. 2011. Evaluating value chain impact using a sustainable livelihoods approach: A case study on horticulture in Afghanistan	Afghanistan	Project MEDA worked through a group of women lead farmers to provide technical assistance, access to quality inputs and technology, postharvest handling information and linkages to higher value chain markets for horticulture products	To improve agricultural productivity and access to markets for women farmers, resulting in increased family incomes and enabling isolated rural women to become economic contributors through fresh vegetable value chain development and market integration	To evaluate the impact of the horticulture value chain project on women in Afghanistan	Random selection of women farmers where MEDA operated	Indicators assessed: Indicators related to the five capitals (natural, human, physical, social and financial)	Before and after evaluation design and used the Sustainable Livelihoods Framework (SLF) for tracking changes in the 5 capital assets. Recall methods of data collection were applied	There was a significant improvement in the five capitals for both households and lead farmers resulting from project activities
Jones et al. 2011. Developing agricultural markets in sub-Saharan Africa: Organic cocoa in rural Uganda	Uganda	Provided a contractual arrangement for cocoa buying combined with increased access to technical advice for farmers who participated in the scheme	To provide incentives to scheme members to process their cocoa to high-grade standards, following established organic standards	To measure the impact of the Esco (U) Ltd. Contract scheme on the income of participants and establish the process that led to the impact	Based on established characteristics of cocoa-growing areas, such as number of cocoa-growing households and their level of specialization in cocoa farming	Farmers in the scheme areas who sold organic cocoa to the scheme Indicators to Assess: Price of cocoa received Cocoa output Household income	Used data from two surveys. Used the Instrumental Variables (IV) approach to estimate the treatment effects	Scheme participation led to substantial improvement in household income Some overall positive changes in farm methods are also attributed to the scheme

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Oduol et al. 2011. Impact of adoption of soil and water conservation technologies on technical efficiency: Insights from smallholder farmers in sub-Saharan Africa	Uganda, Rwanda and DRC	FARA enhanced the levels of awareness and adoption of technologies among smallholder farmers in sub-Saharan Africa by establishing institutional linkages between farmers' organizations and other key stakeholders through an informal coalition of stakeholders known as Agricultural innovation platform (AIP)	To reduce the time lag between technology development, dissemination and its uptake by farmers internalizing the factors that constrain its adoption	To quantify the impact of adoption of SWC technologies developed and disseminated through past approaches to agricultural research and development on technical efficiency of smallholder farmers in East and Central Africa	Farmers in the Pilot Learning Sites (PLS)	Stratified Random Sample selection of farmers in the PLS Assessed indicators: Technical Efficiency (TE)	Used cross-section data. Used Instrumental Variables (IV) to estimate the treatment effects	Adoption of SWCTs generated and disseminated through previous approaches to agricultural research and development have not improved technical efficiency for farmers in the study areas
Carter et al. 2012. The impact of Rural Business Services (RBS) on the economic wellbeing of small farmers in Nicaragua	Nicaragua	Rural Business Development (RBD) project provided to either individual farmers or groups Technical and financial training Intensive training Expert technical assistance Marketing support Materials and equipment support	RBD was designed to support farmers in developing and implementing businesses plans around high-potential enterprises with the objective of increasing productivity and ultimately economic wellbeing	To estimate the impact of RBD interventions on the well-being of smallholder farmers	Selected areas had to have enough farmers to participate as leaders of the nuclei. Farmers also had to meet the project criteria	Targeted program compliers in early treatment RBS clusters Key assessed indicators Increased farm income Increased household expenditure	Quasi-experimental randomized program rollout design and data was collected using two surveys Estimation of program effects was done using Instrumental Variables (IV), specifically the Local Average Treatment Effects (LATE) estimators	Income from program targeted activities increased

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Blair et al. 2012. Impact evaluation findings after one year of the Productive and Business Services Activity of the Productive Development Project	El Salvador	The Production and Business Services (PBS) offered technical assistance and training, giving handouts such as agricultural inputs, setting up demonstration plots and group training sessions, provided technical and financial support for enterprises, investment in innovative agricultural and non-agricultural projects	To assist in the development of profitable and sustainable ventures for individual farmers	To evaluate the one year impact of PBS activities	Farmers in three value chains: handicrafts, dairy and horticulture	Farmers who participated in the program activities Key assessed indicators Production levels Business practice adoption Technology adoption Product diversification	Quasi experimental randomized rollout design and used Instrumental Variables (IV) approaches to estimate the treatment effects. Two surveys were used for data collection	Positive effect on employment in the horticulture value chains Increased net income in the dairy sector
Waarts et al. 2012. Sustainable tea production in Kenya: Impact evaluation of Rainforest Alliance and Farmer Field School (FFS) Training	Kenya	The Kenya Tea Development Agency Sustainable Agriculture Project implemented two major training activities: 1) Farmer Field Schools, which covered wide range of topics and 2) Rain Forest Alliance (RA) training and certification aimed at achieving RA certification	To enhance smallholder tea farmers' access to niche markets and enhanced tea value chains	To evaluate the impact of the project's training approaches on farmers' Good Agricultural Practices (GAPs)	Farmers who supplied coffee to four KTDA factories	Farmers trained using both or any of the training approaches Key indicators assessed Knowledge levels on Good Agricultural Practices (GAPs) Adoption of sustainable practices Income from tea	Quasi-experimental design and applied difference-in-difference to estimate treatment effects. Data were collected using a baseline and follow-up surveys	Farmers who participated in both training approaches had the highest increase in knowledge of GAPs. Training approaches led to increased use of chemical fertilizers. Both had a decrease in crop protection product use. Both approaches led to increased use of hired labour. Productivity increased for all groups

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Burki. 2010. Program evaluation of Agribusiness Support Fund (ASF) estimating the effects of treatment on farmer groups, Agri-businesses and BDS Market in Pakistan	Pakistan	The Agribusiness Support Fund (ASF) provided financial support to micro-agribusiness enterprises called Farming Enterprise Groups (FEGs), linked farmer groups with processors/exporters and built capacity of BDS providers under the training program	The aim was to support farmers in hiring of good Business Development Services (BDS)	To examine how matching grants affected FEGs, participating citrus and dairy farmers, and how the BDS market benefitted from the program	Sample farms drawn from a list of farms receiving program support in the priority sectors in the target areas. Active members of Farmers Enterprise Groups	Members of FEGs Leaders of FEGs Key indicators assessed Farm/enterprise level outcomes Enterprise profitability Technical efficiency and its components Employment generation Household level outcomes Household income and expenditure Household labour supply	Quasi-experimental design and applied difference-in-difference approach for estimating treatment effects. Data collection was done in two surveys for some parts of the study and once for others	ASF support has led to increase in profit of FEGs It led to increase in TE of FEG leaders It increased employment generation on treated farms
ISSER. 2012. An impact evaluation of the MIDA Farmer Based Organizations (FBOs) Training	Ghana	Millennium Development Authority (MIDA) provided training to farmers organized around Farmer Based Organizations (FBOs). FBO trainings were done in 3 thematic areas: business capacity building, technical training and the sales maximization. Besides the training, farmers also received starter packs consisting of fertilizers, seeds, protective clothing and cash to facilitate land preparation	To enhance the competitiveness of High Value Crops in local and regional markets	To measure the impact of the FBO training program on farmer's farm productivity and crop income	Members of FBOs that met the MIDA training criteria	Farmers who participated in the program trainings Key assessed indicators Crop yield Crop income Use of seeds and fertilizers Use of more formal sources of loans	Quasi experimental randomized phase-in design and estimated the treatment effects using difference-in-difference approach. Conducted 3 surveys during the life of the project	No impact on crop yields and crop incomes Increased used of inputs Increase in farmers' use of more formal source of loans

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Baulch et al. 2009. Key findings from the second Thanh Hoa bamboo survey: Second Prosperity Initiative (PI) impact evaluation report	Vietnam	Prosperity initiative (PI) work included undertaking research, networking and information sharing and supporting emergence of new businesses through technical advice and support to policy formulation	To support poverty reduction efforts in the target areas through promotion of bamboo marketing	To understand the contribution of bamboo (Luong) value chain of the industrial bamboo cluster to poverty reduction	Bamboo farmers located in the target areas	Farmers who had earned income from the bamboo value chain Key assessed indicators Per capita household income	Quasi-experimental design and employed a difference-in-difference approach, which reweighted households using their probabilities of participation. Two impact evaluation surveys were implemented	Prosperity Initiative interventions led to poverty reduction
Bonilla et al. 2011. The impact of the seed capital program of SERCOTEC in Chile	Chile	The Seed Capital Program provided a non-refundable financial subsidy to SMEs in their early stages	The Seed Capital Program (SCP) was aimed at mitigating the financing problems of SMEs in the chain of production	To measure the impact of programs on small businesses	SMEs that met SCP's selection criteria	SMEs that registered and participated in SCP's activities Key indicators assessed: Change in number of workers Total sales	Used Propensity Score Matching (PSM) to generate the counterfactual group. Relied on a single firm visit survey. Used difference-in-difference approach to estimate the treatment effects	The project had a positive impact on sales and number of workers for participating SMEs
NORC. 2012. Impact Evaluation of the Farmers Training and Development Activity in Honduras	Honduras	Farmer Training and Development Activity (FTDA) provided the following: Farmer training and development Facilitation of farmers' access to credit Upgrading of farm to market roads Provision of an agriculture public grants facility	To increase the productivity and business skills of small- and medium-sized farms and their employees To reduce transportation costs between targeted production centres, national, regional and global markets	To evaluate the impact of FTDA on household income (farm and off-farm) and on farm employment	Random allocation of villages into treatment and control groups Participation was based on whether a household belonged to the treatment village	Farmers who participated in the program activities Key assessed indicators Net income from horticulture crops and basic grains Net expenditure on horticultural crops and basic grains Net household income Total household consumption Labour expenses	Quasi-experimental approach. Treatment effects estimated using regression-based modelling, including: regression adjusted, basic Propensity Score Matching Estimator, regression-adjusted PS-based estimator, modified regression PS-based estimator and PS-based IV estimator	There was a positive effect of program activities on household income from horticultural crops

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Development Alternatives Inc. 2010. PROFIT Zambia impact evaluation	Zambia	Project PROFIT, a USAID-supported project, worked with lead firms and communities to develop agent networks. These networks reached rural areas lacking sufficient supplies of inputs and services in a number of sectors	The project was aimed at strengthening the connections within selected value chains to increase the provision of inputs and services to farmers, with a focus on improving output and quality (and therefore enterprise and household income)	To determine the links between project activities and the intended impacts on subsector growth and competitiveness, firm level growth and productivity, income increases and poverty alleviation	Participants were located in areas where the project worked	Targeted industry, farms and households Household indicators: Improved productivity Increased income Asset accumulation Sector-wide and firm indicators: Changes in knowledge, skills and behaviour	Used a longitudinal quasi-experimental design implemented through a mixed methods approach (qualitative and quantitative methods). However, estimation approach was not well-defined	Increased use of herbicides and chemicals among participants Increase use of veterinary services in the livestock sector Improved animal health
Forston et al. 2012. Impact evaluation findings after one year of the Productive and Business Services Activity of the Productive Development Project	Armenia	The Water-to-Market (WtM) project provided training, technical assistance and access to credit farms and agribusiness. In addition, the project provided long-term credit to qualified farmers. The project also had the Post-Harvest and the Processing and Marketing (PPM) trainings components	The aim of the project was to increase household income and reduce poverty in rural Armenia through improving the overall performance of the agricultural sector of the country	To assess the impact of WtM training on OFWM and HVA agricultural practices, agricultural production and household wellbeing	Whether the community had irrigation potential	Those households that participated in the program activities Key assessed indicators Cropping patterns Irrigation and other agricultural activities Crop yields Agricultural revenue and costs Other household expenditure Household employment Other sources of household income	Quasi-experimental randomized phase-in design, combining qualitative and quantitative approaches. Used regression modelling to estimate the treatment effects	Small positive impacts on adoption of HVA practices

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
Oxfam GB. 2012. Guatemala Highland Value Chains Development Alliance: Project effectiveness review	Guatemala	<p>The Guatemala Highlands Value Chains Development Alliance included three projects:</p> <p>1) GUAB38—provided support to food security and livelihoods through provision of technical support and fertilizers to support kitchen gardening; and provided high maize price selling outlets for members, established a revolving fund for members</p> <p>2) GUAB49—improved infrastructure for associations and linked broccoli-producing associations with export markets</p> <p>3) GUAB62—strengthened producer linkages with exporters</p>	The aim of the value chain development alliance was to support producers in diversification and increasing returns from agricultural production	The effectiveness review assessed the impact of GUAB38 and GUAB49 for crosscutting associations	Comparison groups were considered the producers who were at the time members of associations other than those supported by Oxfam GB (but similar)	<p>Key indicators investigated included:</p> <p>1) Women Empowerment Global index for assessing women's decision-making and roles</p> <p>2) Self-reported income changes</p> <p>3) Ability to meet basic needs</p> <p>4) Total household income</p> <p>5) Household consumption and expenditure data</p> <p>6) Household food security</p> <p>7) Household asset ownership</p> <p>8) Agricultural production and other productive activities</p>	<p>Applied the PSM and multivariate regression approaches</p> <p>Used data reconstruction techniques (specifically recall methods to acquire the baseline status)</p> <p>Also applied the logic along the intervention logic models; respondents were assessed where they received specific support</p>	<p>Mixed findings</p> <p>No difference between supported and non-supported in yield increase (agricultural productivity)</p> <p>No evidence that women have been encouraged to take up kitchen gardening</p> <p>Supported producers grew less maize but more of the commercial crops</p> <p>No evidence from any of the livelihoods measures that supported household had higher income</p> <p>Supported households reported improved diversity of diets</p> <p>Evidence of the project improving women's economic roles</p>

Study reference	Country	Description of intervention	Objective of intervention	Evaluation objective	How were sites and participants selected?	Target group for evaluation, and indicators assessed	Evaluation methods	Findings
UNIDO Evaluation Group. 2012. Crafting a green future: Bamboo in the curio and souvenir industry in Kenya	Kenya	<p>Project objective was to assist the government of Kenya in its efforts to preserve the forest and its environment while also aiming for sustainable social and economic development</p> <p>Project aim was to develop a bamboo-processing industry to add value to bamboo and use it as a cash crop</p>	<p>The project adapted an integrated agroprocessing value chain approach, prioritizing bamboo processing (with a rural development focus). Project activities included:</p> <p>Marketable products of bamboo crafts for income and general livelihoods for IDPs</p> <p>Development of pre-processing bamboo (e.g. provision of processing equipment</p>	<p>To assess the project for relevance, management and coordination, efficiency and effectiveness</p> <p>Largely evaluated on intended outcomes and sustainability</p>	Received training from project	Beneficiary income	<p>Extensive review of documents</p> <p>Qualitative and quantitative data gathering</p> <p>Focus group discussions</p> <p>Relied on the intervention logic to guide the data collection process</p>	<p>Short-term income increased for both men and women who were trained</p>

ISBN: 92-9146-353-1



The International Livestock Research Institute (ILRI) works to improve food security and reduce poverty in developing countries through research for better and more sustainable use of livestock. ILRI is a member of the CGIAR Consortium, a global research partnership of 15 centres working with many partners for a food-secure future. ILRI has two main campuses in East Africa and other hubs in East, West and Southern Africa and South, Southeast and East Asia. ilri.org



CGIAR is a global agricultural research partnership for a food-secure future. Its science is carried out by 15 research centres that are members of the CGIAR Consortium in collaboration with hundreds of partner organizations. cgiar.org