GEF Evaluation Office

GEF IMPACT EVALUATION

Assessing the Potential for Experimental Evaluation of Intervention Effects: The Case of the Regional Integrated Silvopastoral Approaches to Ecosystem Management Project (RISEMP)

Impact Evaluation Information Document No. 15

AUGUST 2009

This paper is an input into the GEF Evaluation Office (GEF EO) program of **Impact Evaluation.** The findings, interpretations, and conclusions expressed herein are those of the authors and do not necessarily represent the views of GEF Evaluation Office, the GEF Council, or the Governments they represent.

The papers in the Impact Evaluation information document series are:

1. Approach Paper to GEF Impact Evaluation – *Brann and Todd*
2. Final Report on Proposed Approach to GEF Impact Evaluation - *Foundations of Success*
3. GEF Biodiversity Policy Review - *Foundations of Success*
4. Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility – Todd and Vaessen
5. Priorities and indicators for Global Environment Benefits from Biodiversity: The current international architecture – *Nair*
6. Case Study Methodology – *Conservation Development Centre*
7. Case Study: Bwindi Impenetrable National Park and Mgahinga Gorilla National Park Conservation Project - *Conservation Development Centre*
8. Case Study: Lewa Wildlife Conservancy – *Conservation Development Centre*
9. Case Study: Reducing Biodiversity Loss at Cross-Border Sites in East Africa *Conservation Development Centre*
10. Impacts of Creation and Implementation of National Parks and of Support to Batwa on their Livelihoods, Well-Being and Use of Forest Products – *Namara*
11. Protected Areas and Avoided Deforestation: A Statistical Evaluation – *Andam, Ferraro, Pfaff and Sanchez-Azofeifa*
12. GEF Protected Area Projects in East Africa – Conservation Development Centre
13. Applying Quasi-Experimental Methods to Evaluate the Social Impacts of Protected Areas – Andam, Ferraro and Holland
14. Evaluating the Local Socio-Economic Impacts of Protected Areas: A System-Level Comparison Group Approach – Sims
15. Assessing the Potential for Experimental Evaluation of Intervention Effects: The Case of the Regional Integrated Silvopastoral Approaches to Ecosystem Management project (RISEMP) – Vaessen and van Hecken

Global Environment Facility
Director of the GEF Evaluation Office: Robert D. van den Berg

Impact Evaluation Team (GEF Evaluation Office)
Task Manager: David Todd, Senior Evaluation Officer

Authors: Jos Vaessen and Gert van Hecken

Table of contents

List of acronyms

CATIE	*Centro Agronómico Tropical de Investigación y Enseñanza*
CG	control group
CIPAV	*Centro para la Investigación en Sistemas Sostenibles de Producción Agropecuaria*
ESI	environmental service index
FAO-LEAD	Food and Agriculture Organization – Livestock, Environment and Development
FDL	*Fondo de Desarrollo Local*
FONAFIFO	*Fondo Nacional de Financiamiento Forestal*
GEF	Global Environment Facility
GIS	Geographic Information System
NGO	non-governmental organization
PAD	Project Appraisal Document
PES	payments for environmental services
RISEMP	Regional Integrated Silvopastoral Approaches to Ecosystem Management Project
TA	technical assistance
WB	World Bank

Acknowledgments

We would like to thank staff from CATIE, Nitlapán and the World Bank for their collaboration. This report combines an overall perspective of the RISEMP's experimental design with a particular focus on the Nicaraguan case. We would especially like to thank the staff of Nitlapán (Nicaragua) for their openness and constructive comments, a clear sign of a professional learning organization.

1. Introduction

1.1. The Regional Integrated Silvopastoral Approaches to Ecosystem Management Project

This report presents the findings of an evaluation commissioned by the GEF Evaluation Office within the framework of its Annual Report on Impact (see for example GEF, 2007a). The Regional Integrated Silvopastoral Approaches to Ecosystem Management Project (RISEMP) was selected as a case study because it is one of the few recently completed conservation projects based on an experimental impact design, allowing (in theory) for an assessment of the net effects of an intervention.[1] This evaluation analyzes the strengths and weaknesses of the project's underlying experimental design.

The Regional Integrated Silvopastoral Approaches to Ecosystem Management Project (RISEMP) was initiated in 2002. It was a full-sized GEF/World Bank project, designed as an innovative pilot initiative, which would promote silvopastoral practices through technical assistance and payments for environmental services (generated by these practices). The project was implemented in three countries: Nicaragua, Costa Rica and Colombia. It was managed by the World Bank and coordinated by CATIE, an international research institute in Costa Rica. Country pilot sites were managed by national non-governmental organizations (Nitlapán, CATIE, and CIPAV). The intended total cost of the project was US$8.72 million; of which US$4.77 million was financed by a GEF grant and US$3.95 million through co-financing (from FAO-LEAD, Nitlapán, CATIE and CIPAV and other local donors). The project closed in January, 2008.

The main development objectives of RISEMP were to demonstrate and measure; a) the effects of the introduction of payment incentives for environmental services (PES) to farmers, based on their adoption of integrated silvopastoral farming systems in degraded pasture lands; and b) the resulting improvements in ecosystems functioning, global environmental benefits, and local socio-economic gains resulting from the provision of these services (see also the summary logical framework in Annex 1). There were *four project components*.[2] The *first component* aimed at strengthening local development organizations (especially the managing NGOs: CATIE, CIPAV and Nitlapán) to assist farmers in establishing and maintaining improved silvopastoral

[1] With respect to PES it might be the only completed PES project based on an experimental design (Wunder et al., 2008).
[2] A fifth component is project management activities, see also Annex 1.

systems, and in the technical and institutional aspects of silvopastoral systems. The *second component* concerned developing and implementing an improved monitoring system to provide accurate information and understanding on the potential of intensified silvopastoral systems in providing global environmental services and local socio-economic benefits. The *third component* was about creating and implementing a payment mechanism to provide incentives for establishing and maintaining improved silvopastoral systems on farms. The *fourth component* aimed to support policy formulation and dissemination, specifically developing a replication strategy, including exploration of potential sustainable financing mechanisms, to ensure the long-term sustainability of the project.

Important differences with other prominent PES programs, such as the Costa Rican national PES program,[3] should be noted. First of all, RISEMP focused on landscape restoration in agricultural landscapes whereas most other programs focus on land use conservation (e.g. forest conservation). This has implications in terms of costs of implementation (e.g. land use monitoring, market development) as well as the sustainability of the generated environmental services (threats to degradation). Second, relatively little attention was devoted to financial sustainability of payment mechanisms. In contrast, the project focused on testing the effectiveness of payment mechanisms on land use changes in agricultural landscapes and analyzing the relationships between different land uses and the generation of environmental services. Both issues are to a large extent unexplored territories of inquiry and thus illustrate the innovative nature of the project.

Box 1. Key concepts defined

Outcomes are defined as short-term, immediate effects attributable (in part) to intervention outputs.

Impacts refer to the "[p]ositive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (OECD-DAC, 2002: 24).

We use the term *effects* in the more generic sense, to refer to the direct and indirect

[3] Also supported by the GEF.

changes that are (in part) the result of an intervention. Effects comprise both outcomes and impacts.

The RISEMP project was in essence a research and innovation project. Apart from providing incentives to farmers to adopt silvopastoral practices in function of generating multiple environmental services, the project was designed to investigate:

1. on the one hand, the effects of different types of incentives on land use changes and the sustainability of these changes;
2. and on the other hand the effects of land use changes in terms of (global and local) environmental services and (local) socio-economic benefits

Thus, to some extent the project in itself was *about* outcome and impact assessment. As part of the project's objectives, the project teams (in the three countries) in collaboration with World Bank staff developed their own system of research and monitoring, the results of which have been published in project documents and books and journals (see for example Pagiola et al. 2004; Ibrahim et al., 2007; Pagiola et al., 2007).

The project was based on the experimental mechanism of targeting groups of farmers with different incentives. In principle, this would offer a solution to the attribution problem in impact assessment, as differences between otherwise similar groups could then be attributed to the differences in incentives received from the project.

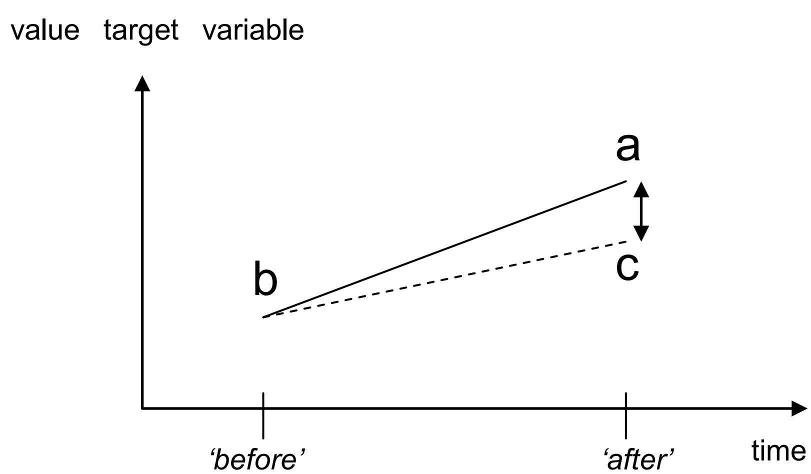
1.2. Logic and comparative strengths of (quasi-)experimental designs for evaluating intervention effects

A fundamental problem in outcome and impact evaluation is attribution. Can changes in certain variables be attributed to an intervention or are they the result of other factors? The project's underlying experimental design targeted exactly this question.

The most widely known and advocated types of methodological approaches that address the attribution problem are experimental and quasi-experimental approaches. The idea of (quasi-) experimental counterfactual analysis is that the situation of a participant group (receiving benefits from/affected by an intervention) is compared over time with the situation of an equivalent control group that is not affected by the intervention. Several designs of combinations of ex ante and ex post measurements of participant and control group have been used in this type of analysis (see for example Cook and Campbell, 1979; Shadish et al., 2002; for development interventions see for example Baker, 2000; Bamberger, 2006; White, 2006; Bamberger and White, 2007). Randomization of intervention participation is considered to be the best way to create equivalent groups. In case of random assignment to either the participant and control group, in sufficiently large samples the probability that both groups are equivalent on all observable and non-observable characteristics except for intervention participation is very high.[4]

The attribution issue can be briefly illustrated as follows. Consider a target variable *x*. In Figure 1 line *b-a* is the evolution of variable *x* in the participant group while line *b-c* represents the evolution within the control group. Randomization of membership of either of the two groups assures for the fact that the differential evolution between the two lines can be attributed to the intervention (since other factors can be considered equal). Consequently, the *net* effect of the intervention is the difference between *a* and *c*.

Figure 1. Graphical display of the net effect of an intervention



Source: own elaboration**.**

In sum, the approach offers the following advantages:

[4] As a second-best alternative, several matching techniques (e.g. propensity score matching) can be used to create control groups that are as similar to participant groups as possible (see below). Finally, regression-based approaches following the same logic of counterfactual analysis can be used in case design-based data are not available or designs are not feasible in practice, see for example GEF (2007b).

- It provides a robust estimate of the *net* effect of an intervention ‘controlling’ for other factors
- It provides an indication of the magnitude of an effect

Whether or not these advantages are realized in practice depends on the extent to which major threats to validity are effectively addressed by the project (see section 2.3.). We discuss this for the case of the Nicaraguan pilot site below.

1.3. The RISEMP project’s experimental design for outcome and impact assessment

The project was designed as a pilot project to test the effects of incentives (payments for environmental services and technical assistance) on land use changes and ultimately on environmental and socio-economic benefits. The experimental design was to be the basis for being able to attribute changes to different types of incentives. More specifically the following hypotheses were to be tested on the basis of the design (operational manual RISEMP):

1) Adoption of silvopastoral practices can be attributed to payments for environmental services (PES)
2) Adoption of silvopastoral practices can be attributed to technical assistance (TA)
3) Adoption of silvopastoral practices can be attributed to both payments and technical assistance

In addition, a fourth hypothesis was that different payment schemes (2 years and 4 years) would affect the speed and intensity of adoption behavior. Farmers receiving only two years of payments were expected to invest more heavily in their farms in order to benefit as much as possible from the PES payments (and get as close as possible to the maximum payment of 6,000 US$ per farm). Moreover, the comparison between the two groups would have to shed light on the question of sustainability of land use changes. It was hypothesized that the 2 years group would initially invest more and subsequently (after payments had ceased) less in land use changes (or even reverse some of the changes).

In order to be able to test these hypotheses, it was envisaged that the following ‘treatment’ groups should be established: a group of farmers receiving only PES (‘PES only’), a group receiving PES and technical assistance (‘PES + TA’) and a control group. In addition, the two

PES groups were again subdivided into a group receiving PES for 2 years ('PES 2 years') and another group receiving PES for 4 years ('PES 4 years')

The selection of farmers for the different 'treatment' groups ('PES only', 'PES + TA', 'PES 4 years', 'PES 2 years', control group) was to be done at random to assure equivalent groups. Consequently, the effects of different types of incentives can be directly deduced from a simple comparison of means (of changes over time)[5] between 'treatment' groups.

The experimental design can be used to compare groups in terms of the changes in land use over time. In addition, land use changes can be directly linked to environmental impact on the basis of an environmental services index (ESI). Consequently, the experimental design can be used in a fairly straightforward manner for estimating environmental impact (i.e. by multiplying the environmental value of a particular land use with the area of application and comparing this outcome between groups). The ESI was based on past research, and as part of the project's research activities was successfully validated and adjusted to better reflect the relationships between different types of land uses and environmental benefits. Annex 2 presents the ESI of the different land uses for biodiversity and carbon sequestration.

2. Objectives and methodology of the evaluation

2.1. Delimitation and objectives

On the basis of the project's strategy and logical framework one can discern three principal dimensions of project effects:

- Effects at field level: this refers to the processes of change induced by the project at pilot site level (among and beyond the participant farmers), from project outputs to outcomes and impacts;
- Institutional effects: in a narrow sense referring to the learning processes induced by the project at the level of the three implementing organizations (CATIE, Nitlapán, CIPAV) and the World Bank, reinforcing capacities and knowledge to further innovate as well as implement similar interventions; this is closely related to the third dimension;

[5] Also called double difference.

- Replicatory effects: which refers to the indirect processes of change in terms of diffusion and uptake of lessons learned; the nature and extent to which these lessons are taken up in research, policy design and implementation by research communities, governments and other institutional actors.[6]

As an innovative pilot project, all three dimensions are of importance and there is a link between the three. This report *exclusively focuses on the first dimension*, which in a sense provides the ingredients for the other two dimensions (the substantive content for institutional learning and replication). It focuses on the fundamental question of how to evaluate the effectiveness of payments for environmental services (and other incentives) in generating and sustaining such outcomes as land use changes and such impacts as environmental and socio-economic benefits.

The current evaluation focuses on the case of the Nicaraguan pilot site, one of the three pilot sites of the project. The Nicaraguan case was selected for its learning potential, as the implementation of the design experienced more problems than in the other two countries.

The objectives of the evaluation are the following:

- To assess the strengths and weaknesses of the experimental design underlying the RISEMP project; its design and implementation;
- To assess the potential of the experimental design as a basis for analyzing the effectiveness of project incentives on land use changes;
- To suggest alternative and/or complementary methods for outcome and impact assessment;
- To draw lessons on the viability and utility of (quasi-) experimental designs as a future evaluation component in similar projects as the RISEMP project, in which the GEF could play a role.

2.2. Methodology

2.2.1. Data collection

[6] Evidently, replicatory effects can also occur at the farmer level. In this case, we classify these under field level effects.

The evaluation relied on a variety of sources of data and methods of data collection including stakeholder interviews, interviews with farmers, document review and secondary data. Stakeholder interviews were conducted with staff from Nitlapán, CATIE and the World Bank (see Annex 3 for a list of interviewees). In addition, in Nicaragua a total of 29 farmers were interviewed (see next paragraph and Annex 3). Document review included documents produced by stakeholders (within the framework of the project) but also 'external' literature on impact evaluation. Project data on land use changes and adoption of silvopastoral practices were used in the inter-group comparisons that constitute the basis of experimental evaluation of intervention effects.

Some methodological observations on the interviews with farmers are in order. In the Nicaraguan pilot site of Matiguas-Rio Blanco 29 farmers were selected using a maximum variability sampling procedure. The interviews were conducted in name of the University of Antwerp in order to avoid being associated with the project and therefore eliciting socially desired responses. The latter phenomenon is quite common among farmers given the expectations they have vis-à-vis the large number of organizations offering support in the region, as well as more specifically, the long history of cooperation between several farmers and Nitlapán. Basically, the first part of every interview was devoted to getting to know the life history and livelihood strategies of every farmer (on average 1 hour), after which the interview was gradually directed to the topic of projects and institutions with which the respondent had been collaborating, eventually talking about the RISEMP project (on average 45 minutes).

Interviews with stakeholders and farmers were semi-structured and covered a list of topics described in the next section on the conceptual framework that was used for evaluating the project's experimental design. Triangulation between opinions and findings from different interviews, documents and data analysis was used to validate findings.

2.2.2. Methodological framework

The basic idea of the experimental design is that one compares the intervention situation with the counterfactual, the situation that would have occurred without the intervention, in order to determine whether and to what extent changes in variables of interest can be determined to the intervention. More specifically, one compares a participant group (affected by/receiving benefits from the intervention) with a control group, a group that exactly resembles the participant group in all aspects but for participation in the intervention.

The validity of the tests of the main hypotheses underlying the experimental design (see section 1.3.) depends on the extent to which the group comparisons actually represent unbiased estimates of the net effects of particular incentives. In other words, in order to be able to analyze in a credible and valid way (and subsequently accept or reject) these hypotheses, the following three inter-group comparisons would need to be bias-free:[7]

- comparing the average change over time of the ‘PES only’ group with the CG;
- comparing the average change over time of the ‘PES only’ group with the ‘PES + TA’ group;
- comparing the average change over time of the PES 2 years group with the PES 4 years group.

Several aspects of design and implementation of an experimental design in development interventions can potentially threaten its validity. In analyzing the strengths and weaknesses of the experimental design the following aspects were taken into account, which are deemed most pertinent as threats to the validity and utility of an experimental design in projects such as the RISEMP project:[8]

1. Selection bias: refers to the problem of under- or overestimating project results due to uncontrolled differences between different (treatment) groups of farmers that would lead to differences in result variables if none of the groups would have received project benefits (Rossi et al., 2004; Shadish et al., 2002). One can differentiate between selection bias on the basis of observable variables (e.g. farm size, education level) and unobservable variables (e.g. motivation, risk aversion).
2. Contagion (or treatment diffusion): refers to the problem of groups of farmers that are not supposed to be exposed to (or receiving) certain project benefits are in fact benefiting from a project in one or more ways: by directly receiving the benefits from the project, by indirectly receiving benefits through other participating farmers (e.g. knowledge transfer), or by receiving similar benefits from other organizations (see Shadish et al., 2002).

[7] Not affected by any of the problems described
[8] For detailed discussions see for example Shadish et al. (2002) or Morgan and Winship (2007).

3. Behavioral responses: several unintended behavioral responses not caused by project incentives or 'normal' conditions might disrupt the validity of comparisons between groups and hence the ability to attribute changes to project incentives. The most important are the following (see Shadish et al., 2002):
    a. Expected behavior or compliance behavior: participants react in accordance with project staff expectations for reasons of compliance with the established contract, due to the (longstanding) relationship with staff, or due to certain expectations about future benefits from the organization (not necessarily the project).
    b. Compensatory equalization: discontent among staff or recipients with the inequality between groups might result in compensation of groups that receive less than others.
    c. Compensatory rivalry: differentiation of incentives between groups of farmers might result in social competition between those receiving (many) project benefits and those that receive less or no benefits.
4) Other aspects that might weaken attribution analysis:
    d. Characteristics of the intervention.
    e. Quality of the data collected.
    f. Timing of the data collection activities.
    g. Characteristics of the design.

3. Assessing the design and implementation of the experimental framework in the Nicaraguan pilot site

3.1. Introduction: project design and preparation[9]

In Nicaragua, in 2002, in order to start project activities farmers were selected in a systematic manner and assigned to groups receiving different types of incentives. After the selection of communities (seven communities in two watersheds in the Matiguas-Rio Blanco region) in which the project would intervene, a census was held among all farmers. On the basis of the census the project staff invited farmers to meetings to explain the rules of the game and

[9] Findings in this section are primarily based on interviews with staff from Nitlapán, and to a lesser extent interviews with staff from CATIE and farmers.

promote the project. The project objectives were explained to farmers and farmers were selected for participation on the basis of the following criteria (operational manual RISEMP):

- small and medium farmers;
- secure land tenure;
- livestock as principal income activity;
- willingness to sign a contract with the project;[10]
- willingness to collaborate with project monitoring activities regarding the following information: socio-economic, carbon, water, biodiversity data;
- willingness to participate in training and receive technical assistance;
- willingness to develop a farm development plan in order to generate environmental services and improve productivity;
- willingness to continue to manage silvopastoral systems after project closure.

In addition, in practice the following criteria for selection were applied:

- proximity to the road;
- farmer should live in the farm;
- farmer should have between 8 and 100 hectares of land.

At the time of project initiation meetings with farmers, the message was that all farmers participating in the project would receive payments for environmental services generated by their changes in land use. Some farmers left as they did not believe that benefits would come forth or they lost interest in the project.[11] Consequently, apart from the formal selection criteria a kind of natural selection process took place in which the most motivated farmers, i.e. those that continued to attend the meetings, would be the first to qualify for project benefits.

The interested and selected participants were then assigned to two groups, those that would receive payments and technical assistance (PES + TA) and those that would receive payments only (PES). For the two groups preliminary quota were established per community. Subsequently, independent of the previous subdivision, the total group of people receiving PES (with or without TA) was again divided into two groups, one receiving only payments during the

[10] Among other things the contract stipulated some land use restrictions such as the prohibition to burn fields or forested areas, or the prohibition to deforest.

[11] Other reasons for leaving/not participating were the following: general feeling of distrust towards institutions, reluctance to sign a contract with the project, reluctance to take risks when investing in the farm, resistance to experimentation (in terms of trying out new practices).

first two years of the project, and one receiving payments for four years (until the end of the project). The control group was established later on (see below).

The number of farmers per category was more or less fixed beforehand (PES: 30; PES + TA: 70; CG: 30; see operational manual RISEMP). Actual numbers of farmers per group in 2003 are depicted in Table 1 (see also Annex 4 for the other countries).

Table 1. Subdivision of farmers according to type of incentive, Nicaraguan pilot site

Group	*2003*	*2007*
PES	**30**	**28**
2 years	7	6
4 years	23	22
PES + TA	**77**	**70**
2 years	24	20
4 years	53	50
CG	29	25
Total	136	123

Source: RISEMP data

3.*2. Validity of comparison of means: 'PES only group' versus 'control group'*

3.2.1. Selection bias

The groups receiving PES only, PES + TA, and the crosscutting groups receiving PES for a period of 2 or 4 years were established more or less at random from the population of farmers attending project meetings and falling within the pre-established criteria described in the previous section.

The control group was selected after the other groups and its subdivisions (as described above) had already been established. The urgency to find a sufficiently large group of willing farmers and the timing of the selection made it impossible for project staff to select farmers randomly (from the same population as farmers selected for the other groups) or even on the basis of certain selection criteria (see section 3.2.3.). In the end, the control group comprised farmers who had continued to attend the project meetings (but did not comply with selection criteria for PES), farmers who had ceased to attend the meetings, and others. As a result, comparisons between the 'PES only' group and the CG would be biased due to severe problems of selection bias on the basis of observables as well as unobservables.

Selection bias on the basis of observables. Landowners not complying with selection criteria for receiving PES or TA and therefore rejected for receiving PES, in some cases were asked to become part of the control group. This was part of a pragmatic solution to rapidly define groups of sufficient size. The downside of this type of measure was that it introduced a clear selection bias on the basis of observable characteristics (see Table 2). CG farmers had on average more land, livestock and a relatively smaller proportion of the CG (in comparison with the 'PES only' group) had a history of receiving TA prior to the project. In addition, in the case of land and livestock the standard deviation is much higher in the control group because to a large extent the CG contained farmers with properties that were either too small (less than 8 hectares) or too large (more than 100 hectares) to be considered eligible for receiving PES. Especially the latter type of farmer was quite different from the average participant. Apart from having more land and assets and thus more capacity to invest, several of these farmers did not spend much time on their farms, instead having a manager to run their farm for them. In other words, decision-making in these cases was divided between owners and managers. Apart from variables such as farm size, assets, living and working on the farm, there were other differences. In the PES groups there were subgroups of people that were very well organized. This was mostly due to the fact that the initial group of potential project beneficiaries included networks of members of some farmer associations with a previous experience of working with and benefiting from projects implemented by Nitlapán. There was no such social structure linking control group farmers to each other.

Selection bias on the basis of unobservables. Some of the farmers that had lost interest in the project at the time of preliminary meetings (before the experimental design was established), were later asked to become part of the CG. While some of the CG farmers were likely to be more reluctant to adopt innovations than the average participant, in practice a subgroup of the CG was triggered by the project to invest in silvopastoral practices (see discussion below).

<u>Table 2. Evidence of observable selection bias between the different 'treatment' groups of the RISEMP project, Nicaragua pilot site</u>

	Farm size	Units of	Received	Received credit

	(ha) *mean (std.dev.)*	livestock *mean (std.dev.)*	technical assistance in the 3 years prior to project *(% yes)*	in the 5 years prior to project *(% yes)*
Control group (n = 25)	46,7 (37,1)	51,0 (41,1)	12,0 %	60,0 %
PES + TA (n = 69)	31,9 (25,8)	30,2 (29,0)	42,0 %	47,8 %
PES only (n = 28)	29,5 (25,1)	32,6 (35,6)	21,4 %	64,3 %
PES 2 years (n = 26)	27,2 (22,7)	22,9 (23,9)	26,9 %	46,2 %
PES 4 years (n =71)	32,7 26,4)	33,8 (32,7)	39,4 %	54,9 %

Source: RISEMP baseline data 2002

3.2.2 Treatment diffusion

Several types of contagion affected the experiment, reducing the differences between 'treatment' groups and rendering part of the comparison between groups invalid.

There was no contagion effect with respect to PES. Payments were restricted to the groups selected for payments. However, payments were expected to alleviate the capital constraint and consequently boost adoption of silvopastoral practices. Any evidence of this effect could be distorted by selection bias problems; i.e. a substantial number of CG farmers (see above) did not face a capital constraint and would be able to invest without receiving PES.

The objective of technical assistance was to relieve the knowledge constraint for being able to invest in silvopastoral practices. The potential of using the experiment as a framework for isolating technical assistance effects from the influence of other effects was in practice completely compromised by the following types of contagion.

The first type of contagion refers to the issue of farmers that were not supposed to receive TA from the project, in fact benefiting directly from project TA activities. Project staff admirably tried to separate the PES + TA group from the other groups ('PES only' and CG). Nevertheless, given the proximity between the farmers, the social relationships among farmers and also the social relationships between farmers and staff, this was very difficult. Information about project activities was widely available. Although to a large degree, workshops, exchanges and personal visits were restricted to the technical assistance group, other farmers from other groups, although not formally invited, would also sometimes attend the sessions. It was difficult for project staff to prevent these farmers from attending.[12] In other occasions, information and

[12] Nor did they feel very motivated to exclude farmers from courses.

advice on land use techniques was given to farmers from whatever group when they asked for it.

The second type of contagion refers to the diffusion of knowledge from farmer to farmer. Farmers from the different groups were often neighbors, friends, members of the same networks or even family members. Farmers from the technical assistance group would often share their newly acquired knowledge on silvopastoral techniques with others. Evidently, when some of the silvopastoral practices began to manifest their benefits, interest from other farmers increased. What in development projects is usually considered as an important benefit, i.e. the diffusion of project knowledge beyond the participant group, turned out to be a substantial impediment on the validity of the experimental design.

The third type of contagion refers to knowledge acquired by other institutions. Before the project initiated its activities in the region, several institutions, including Nitlapán, had already been working with farmers on livestock production and land use systems. In fact, most of the practices promoted by the RISEMP project were not new. During the project implementation period, Nitlapán successfully negotiated a division of labor with the principal institution delivering TA in the region, Technoserve. The implication was that Technoserve would focus on other topics and other farmers than Nitlapán. Despite this agreement, basic knowledge of silvopastoral practices continued to be available to farmers through different channels.

While both the ‘PES only’ group and the CG were meant not to have access to TA, both groups in fact had access to knowledge about silvopastoral land use practices through the three mechanisms described above. It is unclear whether the ‘PES only’ group on average benefited more from project TA than the CG. Probably, the problem of treatment diffusion of TA has had no substantial effect on the validity of the comparison between the groups as an indicator of the net effect of payments on land use changes. In contrast, treatment diffusion rendered group comparisons between the ‘PES + TA’ group and ‘PES only’ group invalid, as this comparison was precisely designed to test the effectiveness of TA on land use changes (see section 3.3.2.).

3.2.3. Unintended behavioral responses

Previously, we identified several unintended behavioral responses that might negatively affect the validity of group comparisons.

The first type of unintended response concerns expected behavior or compliance behavior: participants react in accordance with project staff expectations for reasons of compliance with the established contract or the relationship with staff or due to certain expectations about future benefits.

Although unintended[13] from an experimental design point of view, project staff in fact intentionally tried to influence participant behavior through other mechanisms than payments and technical assistance. For example, the practices of burning plots and deforestation were by formal agreement prohibited and farmers not complying with this directive would be expelled from the project. The implication of the latter is that any change or lack of change in forest cover cannot be uniquely attributed to project incentives given this contractual obligation.

Another way in which project staff tried to ensure contract compliance as well as stimulate adoption of silvopastoral practices was the promise of a second phase. The impression was created that if farmers did well they would be eligible for a second phase[14] of payments and project benefits. The promise of a second phase was also used to motivate farmers to join the CG. This had a clear behavioral effect in terms of motivating several CG farmers to start investing in silvopastoral practices with the expectation of becoming a future participant eligible for payments.

The longstanding relationship between several farmers receiving PES (and TA) and Nitlapán was another element that triggered unintended behavioral responses. For example, participant farmers in the community of *San Ignacio* had been working with Nitlapán for years and by the end of the project were likely to continue the collaboration in the future. Compliance with contractual obligations as well as land use changes in this community were substantially influenced by a variety of mechanisms such as trust, reputation and friendship that were part of the ongoing collaboration between the farmers and Nitlapán.[15]

[13] We use the term unintended to refer to other responses than those generated by project incentives.
[14] In fact, project staff and WB staff themselves were convinced that they would be able to successfully define a follow-up project.
[15] In fact, adoption of silvopastoral practices in this community has been more successful than in other communities.

The second type of behavioral response is called compensatory equalization: the idea that discontent among staff or recipients with the inequality between groups might result in compensation of groups that receive less than others. The selection of the control group in many aspects had been problematic. The late introduction of the idea of a control group made it extra hard to find farmers willing to be part of this group.[16] Apart from the promise of eligibility for a second phase of the project, in order to arrive at a sufficiently large number of farmers willing to be in the control group, Nitlapán relied on the following mechanisms:

- (in some communities) the longstanding relationship of collaboration between certain farmers and Nitlapán was invoked to persuade farmers to be part of this group;
- farmers were paid a small fee (10 US$ per year) for collaborating with the data collection activities in the farm;
- farmers were offered a map of their farm, based on a satellite photograph and GIS information; this was considered to be useful for further on-farm planning and production.

A final unintended behavioral effect is compensatory rivalry: differentiation of incentives to groups of farmers might result in social competition between those receiving (many) project benefits and those that receive less or no benefits. In the Nicaraguan case, this has probably been the strongest unintended behavioral effect. Among a number of CG farmers there was substantial resentment for being rejected for participating in the project.[17] As a result, these farmers wanted to show the project staff and the other participating farmers that they could innovate without project support. The fact that they lived near farmers who were receiving TA (and PES), which facilitated learning about the new land use practices, made it easier for CG farmers to compete.

3.2.4. Conclusion on validity of comparison of means

Brief summary on threats to validity:

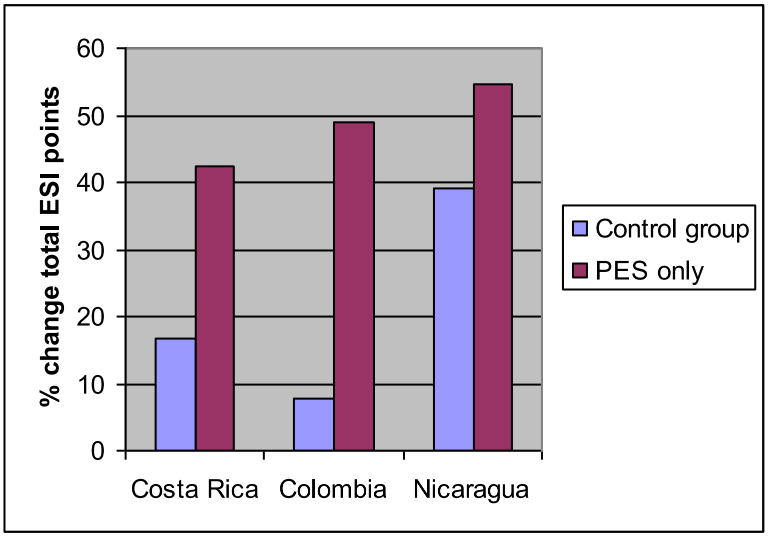
- selection bias: problematic
- treatment diffusion: not problematic
- unintended behavioral responses: problematic

Validity of comparison of means: low

[16] would be deprived from previously promised benefits.
[17] Interviews with farmers and Nitlapán field staff.

Figure 2 shows a comparison of increases in ESI points (see Annex 2) per hectare for the three countries.

Figure 2. Incremental ESI points per hectare (2003-2007), three countries – PES only versus CG



Source: own calculations based on RISEMP project data, January 2008

In theory, the net effect of the project would be the difference between the average incremental ESI values of the 'PES only' group and the CG. On the basis of the previous discussion we can conclude that given the problems in the design and implementation of the experiment, both the CG and 'PES only' group values are distorted in such a way that an unbiased comparison is impossible. With respect to the CG, we do not know to what extent problems of selection bias, contagion and unintended responses have distorted the value of the CG from the value it would have had in the absence of these problems. We can only conclude that several elements have contributed to a relatively high increase in ESI points in the CG. For example in the case of Nicaragua:

- Selection bias: several CG farmers on average were wealthier than 'PES only' farmers, with access to capital (and labor) to implement changes (positive effect on adoption of silvopastoral practices).
- Contagion: several CG farmers had access to technical assistance from other organizations as well as the project itself (positive effect on adoption). The fact that the

‘PES only’ group also had access to TA makes contagion less problematic for this particular comparison.

- Unintended responses: several CG farmers harbored the expectation of becoming a future beneficiary of the project (positive effect on adoption).

With respect to the ‘PES only’ group, unintended responses such as compliance behavior or the expectation of maintaining access to future benefits of Nitlapán positively affected adoption behavior.

To conclude, the true net effect of the project for this variable (difference in incremental ESI points) is different from what we can deduce from the inter-group comparisons of ‘PES only’ and CG farmers. Given the systematically higher incremental change in ESI points (or environmental value of the land use) among ‘PES only’ farmers in comparison to CG farmers in the three countries we can safely conclude that PES has had an effect on land use changes. Nevertheless, for the Nicaraguan case, given the problems affecting the design, and on the basis of these comparisons only, we cannot say whether payments have been important or even decisive in bringing about land use changes and corresponding indirect effects.

It is worthwhile to mention that the lower values in the CG in Costa Rica and Colombia suggest that the experimental designs in these sites were probably less problematic than in Nicaragua. However, also in these cases the differences between ‘PES only’ and CG values do not represent unbiased estimates of the net effect of the project. Given the scale and diversity in problems affecting the experimental design in Nicaragua it is unlikely that the other two sites were unaffected by any of these problems (see Annex 5).

3.3. Validity of comparison of means: ‘PES only group’ versus ‘PES plus technical assistance group’

3.3.1. Selection bias

The allocation of farmers to the ‘PES + TA’ group, the group that was considered to be the most attractive to farmers, was not entirely random. Interviews with staff and farmers confirmed that there was some favoritism towards the most motivated farmers (those that had attended all the meetings) and farmers with a history of collaborating with Nitlapán on previous projects. The higher percentage of previous access to TA among ‘PES +TA’ farmers appears to confirm this.

Nevertheless, at group level, the two groups ('PES + TA' and 'PES only') are largely similar with little observable selection bias (see Table 2). In addition there is likely to be some unobservable selection bias as assignment of farmers to either of two groups was not entirely done in a random manner.

In sum, the principle randomization of groups was respected to some extent and there were no substantial differences between the 'PES only' and PES + TA' group that might seriously invalidate the comparison of means as a measure for the net effect of the project.

3.3.2. Treatment diffusion

On the basis of the discussion in section 3.2.2 we can conclude that group comparisons between the 'PES + TA' group and the 'PES only' group are rendered useless due to the different treatment diffusion problems. Knowledge on silvopastoral practices was widely available through the three mechanisms described earlier. 'Treatment' differences between the two groups are considered to be too small to allow for any meaningful interpretation of changes attributable to the project TA on the basis of the experimental design.

3.3.3. Unintended behavioral responses

Although the experiment was severely affected by unintended behavioral responses, the effect was primarily on the 'PES only' group versus CG comparison. There are no reasons to assume that there were any systematic differences in unintended behavioral responses between the 'PES only' and 'PES + TA' groups.

3.3.4. Conclusion on validity of comparison of means

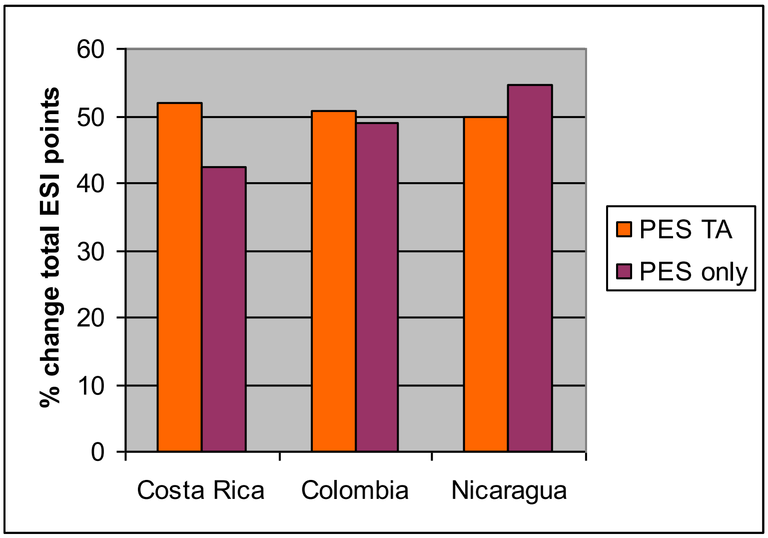
Brief summary on threats to validity:

- selection bias: not problematic
- treatment diffusion: problematic
- unintended behavioral responses: not problematic (no systematic differences between the groups)

Validity of comparison of means: low

On the basis of the previous discussion we can conclude that group comparisons between the 'PES + TA' group and the 'PES only' group are rendered useless due to the different contagion problems. In other words, 'treatment' differences between the two groups are considered to be too small to allow for any meaningful interpretation of changes attributable to the project.

Figure 3. Incremental ESI points per hectare (2003-2007), three countries – PES only versus PES + TA



Source: own calculations based on RISEMP project data, January 2008

Figure 3 confirms this. Differences between the two groups in terms of incremental ESI points are very small in Nicaragua. This is also the case for the other two countries. However, in Costa Rica and Colombia, incremental change is higher in the 'PES + TA' groups. This is a more plausible result as it suggests that the combination of PES and technical assistance is more useful than PES only (see also section 3.5.).

3.*4. Validity of comparison of means: 'PES 2 years' group versus 'PES 4 years' group*

3.4.1. Selection bias

The decision to assign a particular farmer to a group receiving either 2 years or 4 years of payments was done more or less at random. The groups were rather similar in terms of observable characteristics (see Table 2) and (given the random nature of assigning farmers to either of two groups) most likely also for unobservable characteristics influencing adoption behavior. There is a slight bias in terms of farmers in the 4 years payment group having more livestock and a higher percentage of farmers having received TA.

3.4.2. Treatment diffusion

There were no systematic differences between the two groups in terms of access to TA or knowledge on silvopastoral practices in general. Both groups included farmers with access to TA from the project and in addition, knowledge on silvopastoral practices was also widely available to farmers from both groups through the three mechanisms described earlier. There was no treatment diffusion with respect to PES modality (i.e. the differentiation between 2 and 4 years of payments was correctly implemented).

3.4.3. Unintended behavioral responses

Although the experiment was severely affected by unintended behavioral responses, the effect was primarily on the 'PES only' group versus CG comparison. There are no reasons to assume that there were any systematic differences in unintended behavioral responses between the 'PES 2 years' 'PES 4 years' groups.

3.4.4. Conclusion on validity of comparison of means

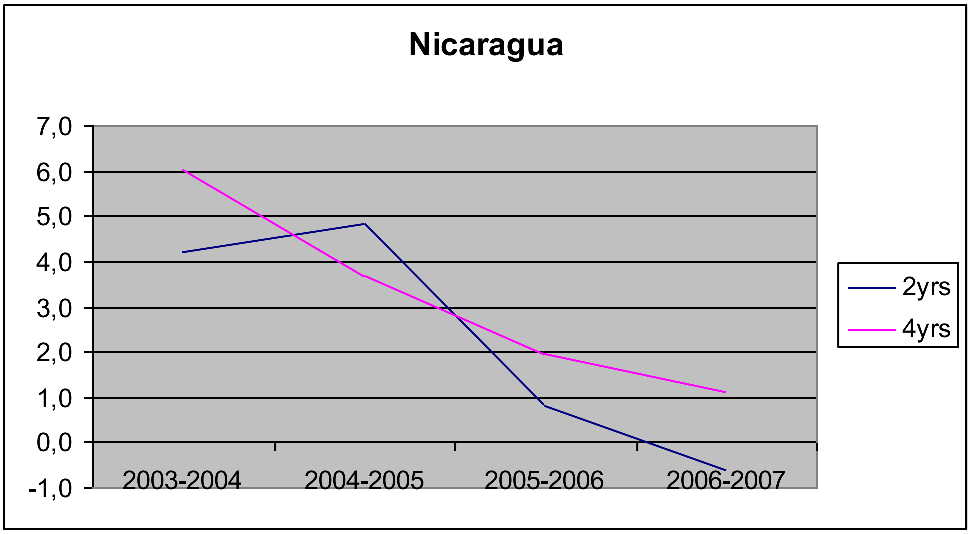
Brief summary on threats to validity:

- selection bias: not problematic
- treatment diffusion: not problematic
- unintended behavioral responses: not problematic (no systematic differences between the groups)

Validity of comparison of means: high

The comparison between farmers receiving only 2 years of payments and farmers receiving 4 years of payments is considered to be quite valid. There were no severe problems of selection bias. In addition, treatment diffusion of PES modality did not occur. Finally, there is no indication that there were systematic differences between the two groups of farmers in terms of unintended behavioral responses.

Figure 4. Adoption behavior of the PES2yrs vs. PES4yrs group (average incremental points per farmer per year, in relation to previous year), Nicaraguan pilot site



Source: own calculations based on RISEMP project data, January 2008

Interesting findings that can be deducted from Figure 4 are the following. First, the trend appears to confirm hypothesis 4 (see section 1.3.) that adoption behavior was initially higher in the 2 years group, whereas it substantially declined in the last two years of the project (trends in other countries confirm this, see Annex 6). In the 2 years group, the incremental ESI even became negative in the last year of the project, indicating a decline in the ecological value of the farm. This had not happened (yet) in the other two countries (see Annex 6), although similar patterns can be discerned there.

The declining trend can be explained by several factors. First, once payments for land use changes from the project stopped, an important incentive for adoption was taken away. Many farmers wanted to get as close as possible to the maximum amount of PES that could be earned (6,000 US$ per farm) given the constraints they were facing. After payments ceased, adoption slowed down as the incentive to earn money had disappeared. Second, resource constraints in combination with the fact that farmers were not (yet) convinced of the (economic) advantages of some of the practices led to a decline in investments in new practices and possibly a decline in maintenance of (some) existing practices (resulting in a declining incremental ESI value (per farm) or even a negative one). The comparison depicted in Figure 4 can be useful for predicting the sustainability of land use changes in the near future. It is likely that some land use practices once they have demonstrated their economic value will be sustained and even expanded into the future (e.g. improved pastures). The initial steep decline and subsequent flattening of the curve appears to support this (see also Annex 6). However, at the same time resource constraints and the lack of economic pay-off for some of the practices (e.g. conservation of shrubs) might result in a decline in the ecological value of the farm (as indicated by the downward trend in Figure 4). The overall sustained gain/loss in terms of ESI points per farm is a question that warrants further inquiry. In practice, whether there will be a net decline or gain in ecological value at the farm level depends on individual-specific factors such as farm household characteristics, on-farm and off-farm income activities, market integration (and correspondingly market conditions), and the existing mix of land use practices as well as recent innovations and trends. Without further insight into the question of how incentives affect types of farmers in different ways, predictions as to the overall sustainable effect of the project on land use changes and indirect benefits are impossible to make (see section 5 for further discussion).

3.5. Other issues that affected the utility of the experimental design

Other issues affecting the usefulness of the design were the following:

- Treatment change. The amount of PES per ESI point was increased during the project (after already having received a first payment). Payments per ESI point were increased to 75 US$ and 110 US$ per point respectively for farmers receiving 4 years and 2 years of payments. The behavioral effect of this increase on adoption behavior has probably been small.
- Quality of the data and measurement problems. In the first years of the project, staff was still learning how to measure and characterize land use activities (for example in distinguishing improved pastures from natural pastures)[18]. This particularly affected the quality of the baseline data. Later on, when more experience was gained in recognizing and measuring particular land uses, the quality of the monitoring increased substantially. As a result, baseline data are probably overestimating silvopastoral land use systems in use. The implication is that actual adoption has been higher for some practices than shown in the data.

[18] Interviews with Nitlapán field staff.

- Timing of the baseline and ex post survey. The timing of the baseline was adequate; the survey was implemented before initiating payments and technical assistance activities. Land use data were adequately collected annually allowing for dynamic comparisons between groups.
- Sample size. Given the small group sizes (especially of the CG) and possible high variance among farmers, statistical power is quite low. The probability of type II errors is high (failing to find a difference between groups when in fact there is a real difference). In addition, a sample size of 30 is more or less the lower boundary for applying the central limit theorem which allows for the application of parametric statistical tests. The implication is that formal confirmatory statistical analysis in some cases is not warranted (especially the comparison between CG and 'PES only'). In those cases, group comparisons should be limited to descriptive comparisons.[19]
- Other characteristics of the design. The design did not include a group that was targeted with technical assistance without payments. The current design if well implemented could have tested for the added value of technical assistance to farmers already receiving payments. However, it could not test for the effectiveness of technical assistance as such. Testing the effectiveness of technical assistance was not an original objective of the project. However, our interviews with field staff and farmers suggest that for many farmers knowledge rather than money is the most direct constraint for successful innovation. Monitoring land use changes of separate groups of farmers with only PES or only technical assistance could have elucidated what on average would have been the most pressing constraint for land use change in the region.[20] Additional analysis (e.g. using household survey data) would have helped to shed light on the question as to what types of farmers would require PES, technical assistance or both (see also section 5).

3.6. Final remarks on the design and implementation of the experimental design for the Nicaraguan pilot site

[19] Non-parametric tests can in principle be used to test for statistically significant differences between groups. However, not only are group sizes relatively small, given the non-random selection of some of the groups, statistical inference is not particularly useful.
[20] This suggestion was raised by staff from Nitlapán who would have preferred this type of design over the current one.

The basic conditions for managing a controlled experiment were not fulfilled. The different stakeholders, WB staff, project staff in Nitlapán and farmers did not share the same vision about the importance of the experiment. In case of trade-offs with the development objective of improving the environmental conditions of farms and the livelihood conditions of farmers in the project, project staff acted in favor of the latter, for example by not exercising a strict control on the technical assistance component of the project. In the management of the experiment there was little or no quality control. As a result the majority of problems in terms of selection bias, contagion and behavioral responses were not addressed or even identified.

The setup of the experimental design was described in the operational manual of the project. The manual provided clear instructions on the allocation of farmers into different groups and on group sizes. In addition, the manual specified the information to be collected in the baseline survey, covering all groups of farmers. In general, the basic logic of the experiment was well understood among staff. However, most of the project staff had no experience with managing such an experiment. The majority of project staff in Nitlapán did not have a background in research, but were experienced livestock specialists and extensionists, having worked all their professional life in rural development projects. For several staff members in Nitlapán, especially the field staff who interacted most frequently with farmers, the primary goal was to implement a project that would bring benefits to the people, not to carefully manage an experiment for research purposes. There were no clear guidelines (at least not on paper) on how to manage the experiment in the field, and staff, having had no training in the logic and management of this type of experiments, when aware of problems that could affect the validity and utility of the experiment had to improvise along the way.

Communication about the experiment to the farmers suffered one serious setback. The project started out with the message that all farmers would receive payments. In a subsequent supervision mission by the WB the scheme was changed and the idea of a control group to be monitored by the project while not receiving any benefits was introduced.[21] This caused quite some resentment among farmers. In the beginning CIPAV (Colombia) and Nitlapán (Nicaragua) and to a lesser extent CATIE (Costa Rica) did not quite agree with the idea of a control group as it was considered not to be ethical. Despite several objections, the project went through with the inclusion of a control group.

[21] Interviews with staff from CATIE and Nitlapán.

The idea of differentiating treatments to farmers was not welcomed by all farmers. The differentiation between participants and control group, the division between those that would receive technical assistance and those that would not and the division between 2 years and 4 years of payments caused substantial resentment as well as confusion among farmers.

In all, the experimental framework failed on two of the three group comparisons that were to support rigorous claims on the effects of PES and technical assistance on land use change and corresponding environmental effects. The ‘PES only’ versus CG comparison is rendered invalid due to severe problems of selection bias and unintended behavioral responses (especially in the CG). The ‘PES only’ versus ‘PES TA’ comparison is rendered invalid due to problems of treatment diffusion. The ‘PES 2 years’ versus ‘PES 4 years’ comparison is quite valid. The data and their subsequent interpretation illustrate the utility of the experimental design in terms of providing reliable evidence on land use behavior under different types of incentives.

4. Utility of the experimental evaluation design in the assessment of project outcomes and impacts

The previous discussions allow us to present an overall picture of the strengths and weaknesses of the project’s experimental design as a basis for outcome and impact assessment. In the presentation of this picture, we will focus on the concrete situation of the project (taking into account design and implementation failures discussed above). In addition, we will highlight some aspects from the point of view of an ideal experimental setting.[22]

4.1. Strengths of the experimental design for assessing outcomes and impacts

A critical comparative advantage of (quasi-)experimental designs is that they allow for a quantitative estimation of the net effect of an intervention, ‘controlling’ for other external factors. In other words, they have a comparative advantage in establishing to what extent changes in target variables are brought about by an intervention vis-à-vis other factors. Both positive and negative effects can be identified, and both direct and indirect effects, depending on the set of variables that is taken into account.

[22] We will restrict ourselves to arguments that directly relate to the type of evaluation context of projects such as RISEMP. We will not go into the general debate on the applicability of (quasi-) experimental methods in development interventions (see for example Bamberger and White, 2007).

In addition, although design and implementation of the experimental framework may be costly and require the necessary technical expertise, subsequent analysis and interpretation of data is fairly straightforward. Experimental design data can provide reliable and easy to interpret evidence to policy makers.

In the case of the RISEMP project, if selection bias, contagion and unintended behavioral responses would have been kept under control, the design would have allowed for unbiased estimates of the net effects of the project incentives on land use changes and environmental target variables. The utility of experimental approach has been illustrated by the valid comparative analysis between groups of farmers that received 4 years of payments versus those that only received 2 years of payments.

4.2. Weaknesses of the experimental design for assessing outcomes and impacts

<u>Disentangling the effects of project incentives from other factors</u>

Although the comparative advantage of the experimental approach lies precisely in its potential to isolate project effects from other effects, the analytical benefits of the design can only be realized if extensive and careful attention is paid to the different threats to validity. Problems of selection bias, contagion, and unintended behavioral responses have compromised the utility of the experimental design. Given the substantial threats to validity, especially in the Nicaraguan case, but probably also in the other two countries, the effects of the project incentives cannot be satisfactorily disentangled from the influence of other factors.

<u>Scope of the evaluation</u>

Experimental designs are not equipped to address the full scope of impact of projects such as the RISEMP project. We briefly highlight two aspects of scope. First, as presented in the introduction of this report, we only focus here on effects at field level. The other two dimensions, institutional effects and replicatory effects, are not amenable to being assessed by means of the experimental approach.[23] The two dimensions cannot be captured in terms of discrete

[23] One might think of ways to assess for example institutional effects through randomized experiments. However, at least three constraints come to mind: the unit of analysis for randomization, the range of institutional effects to be

‘treatments’ being applied to a discrete (and sufficiently large) group of subjects. Tracing and measuring the institutional and replicatory effects of the project would require a completely different approach, starting out from a theory of impact (see below).

Second, even within the confines of impact at field level, the experimental design cannot address the full range of effects of the project. It is restricted to a limited set of indicators. Ideally, as is the case in the RISEMP project, these indicators are included in the baseline study of an intervention. Unexpected and unintended effects are usually not (adequately) captured by baseline studies and therefore not taken into account in the analysis. Examples of effects which cannot be adequately assessed (mainly) due to lack of data are the following:

- Indirect environmental effects such as the displacement of ecologically destructive land uses;
- Other indirect effects: e.g. diffusion of adoption by other farmers, (price) effects on local inputs and commodity markets (through production effects), (price) effects on land markets.

Moreover, in the specific case of the RISEMP project socioeconomic impacts cannot be assessed on the basis of the experimental design. This is not a shortcoming of the design itself but of data collection efforts; socioeconomic variables were only monitored for a subgroup of farmers, not including all groups. Socioeconomic impacts can be inferred from analyses of the private profitability (based on intensive monitoring of a small group farmers) of land use practices in combination with adoption grades. The baseline survey of the project includes relevant socioeconomic data to be used in impact analysis. As to date, no follow-up socioeconomic survey covering all project farmers has been organized.

Timing

Typically, adoption processes of new technologies do not take place over night. Farmers continuously experiment and assess the attractiveness of innovations on their farms. Consequently, given the relatively short time span of the project (five years) and recent closure of the project (in January 2008), one might expect farmers to continue to expand investments in land use practices that they perceive to be attractive while at the same time ceasing to invest in

---

included in the experiment, and the cost-effectiveness of doing such an analysis, especially when compared to other methods of inquiry.

other land use practices (or even undo some of the changes made). Therefore, the full long-term impact in terms of the environmental and socio-economic benefits brought about by land use changes cannot be assessed yet.

Usually, an outcome and impact evaluation based on an experimental approach would rely on a few snapshots of reality, the before and after situation. In addition, experimental evaluations usually look at (short- and medium term) outcomes rather than (long-term) impacts. In fewer cases, periodic measurements of target variables (as in the case of the RISEMP project) are available. In any case, the long-term effects are usually not yet apparent at the time of outcome and impact evaluations. While this is the case for any type of outcome and impact evaluation, there are other methodological approaches, such as sustainability analysis in combination with theory-based evaluation (see below) that are helpful in developing a line of argumentation on the likely sustainability of certain effects (see for example, GEF 2009a). In the case of the RISEMP project, trends in adoption data and group comparisons between the two groups receiving payments for respectively 2 and 4 years are useful for complementing such alternative analyses in assessing the likelihood of sustainability of land use changes and corresponding environmental and socioeconomic effects.

Measurement issues

Technically, this is not just about the weaknesses of the experimental design as such, but applies to a broader range of techniques for statistical analysis and beyond. However, measurement issues are especially important in this context as (quasi-) experiments focus on a narrow set of indicators only, which are used to refer to broader phenomena and processes of change. We briefly highlight two issues: the issue of construct validity, whether a variable correctly captures the phenomenon it refers to, and the type of data.

Any type of statistical analysis (whether descriptive or inferential) is based on a succinct abstraction of reality, not only in terms of the relationship or the model that is the focus of the analysis, but also in terms of the number and choice of variables that are to represent reality. An example from the RISEMP project is the following. Using the variable quantity of land use changes both as a proxy and a basis for assessing environmental effects has its limitations. A distinguishing element not entirely adequately captured by the project's monitoring system concerns the quality of land use changes.[24] When looking at the effectiveness of TA, the distinguishing effect of TA delivered by the project on land use changes will probably not be visible in the quantity of land use changes but rather in the quality of land use changes as well as the sustainability of these changes. The quality of land use change has not been expressed in measurable indicators and consequently is not included in the experimental design-based comparisons between groups.

Regarding the second aspect, type of data, it should be highlighted that there is a substantial difference in working with data that are based on direct measurement (e.g. areas with a certain land use directly measured by project staff using GIS tools) and data generated from survey questionnaires. In general, survey data are more liable to suffer from measurement errors. In the case of the RISEMP project, despite the fact that information-sharing was part of the contract, farmers often felt reluctant to share confidential information within the framework of (socioeconomic) surveys. This was especially problematic in Colombia. The land use changes used in the group comparisons (based on direct measurement) are less likely to suffer from measurement error than survey data used in many other settings.

<u>The rationale underlying changes</u>

Experimental designs are equipped for determining the outcomes and (to a lesser extent) impacts of interventions on target variables while 'controlling' for known (and unknown) external factors. In reality, the changes in target variables are the result of an interplay of factors, of which an intervention is but one of many. Experimental approaches directly relate target variables to 'treatments' and do not address the underlying issue of how changes have come about. What are the causal pathways underlying processes of change influenced by the intervention? Under what circumstances do project incentives positively (or negatively) affect target variables? How do project incentives affect particular types of farmers in different ways? What are potential unintended, indirect or long-term results of interventions? Are there other instruments that might have achieved similar results more effectively? These and other questions regarding the nature of processes of change influenced by interventions and the

[24] According to project staff this aspect is especially important in practices such as fodder banks or live fences. Two fodder banks of the same size (in land area) might differ substantially in terms of the diversity of species of grasses and other plants, with implications for biodiversity in terms of the fodder bank's role in supporting insects and bird species.

resulting effects over time are best addressed using a theory-based evaluation approach (see below).

5. Reinforcing the experimental design-based analysis with other methods

Promising methodological options to reinforce the outcome and impact evaluation in the case of the current project as well as in other similar contexts are the following:

- Additional statistical analysis;
- Theory-based evaluation;
- Sustainability analysis.

5.1. Additional statistical analysis

Despite the problems found in the experimental design, further statistical analysis will be possible if another effort is made to collect ex post data at farm (household) level. This will open up new possibilities for using matching techniques (creating better control groups) or regression-based approaches using statistical controls to reduce observable selection bias problems (and to some extent, if measured) contagion problems. Nevertheless, some of the validity threats to attribution analysis (unobservable selection bias, contagion problems, unintended behavioral responses) cannot be corrected by further quantitative analysis.

Another option for further quantitative analysis would be to focus less on attribution of changes to project incentives and more on the general question of associations between levels and patterns of adoption on the one hand and different incentives, farm (household) characteristics and contextual variables on the other. To some extent, these data are available in the project's baseline survey. The problem is that important variables like land sales or purchases, access to credit and/or technical assistance in the past few years, are crucial explanatory variables without which further explanatory analysis would be markedly incomplete. This provides another reason for implementing an ex post survey covering all PES and CG farmers.

5.2. Theory-based evaluation

Theory-based evaluation focuses on the underlying assumptions of how an intervention is supposed to work (see for example Weiss, 1997; Rogers et al., 2000; Leeuw, 2003). A distinction can be made between process theory and impact theory, the latter focusing on the causal assumptions connecting project outputs (and some process variables) with outcomes and impacts. Several pieces of evidence can be used for reconstructing the intervention theory, for example:

- an intervention's existing logical framework (see Annex 1) provides a useful starting point for mapping causal assumptions linked to objectives; other written documents produced within the framework of an intervention are also useful in this respect;
- insights provided by as well as expectations harbored by policy makers and staff (and other stakeholders) on how they think the intervention will affect/is affecting/has affected target groups;
- (written) evidence on past experiences of similar interventions (including those implemented by other organizations);
- research literature on mechanisms and processes of change in certain institutional contexts, for particular social problems, in specific sectors, etc.

Figure 5. Basic impact theory of the RISEMP project at field level

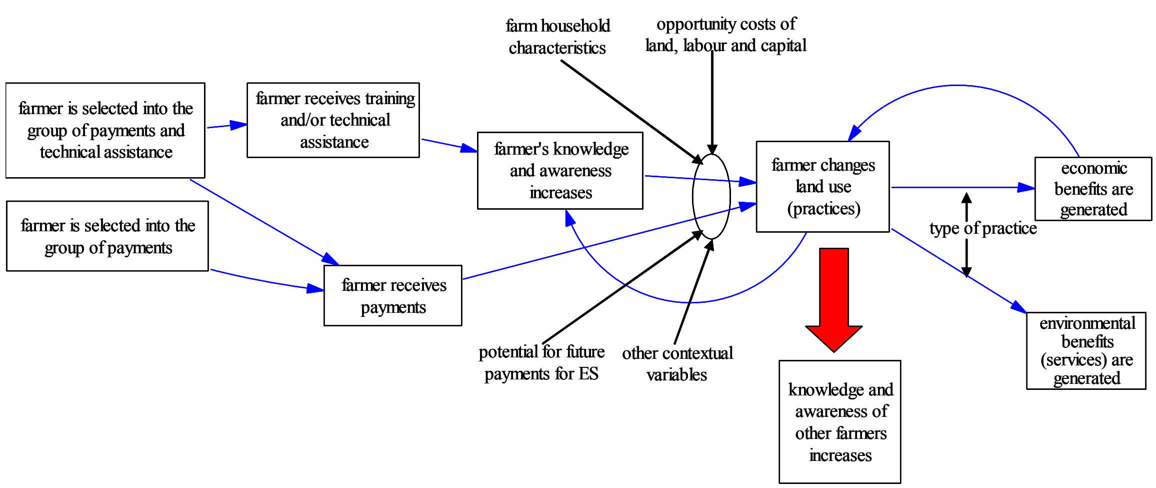


Figure 5 presents a graphical representation of the impact theory (restricted to effects at field level) of the RISEMP project. It shows the main causal assumptions linking project incentives (TA and PES) to final impact variables.

In general, the impact theory can serve multiple purposes within the framework of an outcome and impact evaluation. We briefly mention three principal purposes:

- *Explanatory-analytical function*: what are the main causal pathways through which the project influences processes of change leading to outcomes and impacts? Once the evaluators have reconstructed a workable theory that adequately reflects the existing knowledge on processes of change for the evaluation context at hand, the theory can be used as a basis for argumentation, supporting the analysis on impact when needed. In addition, the theory can be helpful in explaining differences in effects among farmers, i.e. conditioned by contextual variables (e.g. farm gate prices of milk, agro-physical conditions) and characteristics of farmers (e.g. age, wealth).
- *Methodological design function*: the initial impact theory provides a useful framework for determining what type of evidence is needed for testing and refining the theory and responding to the principal questions of the outcome and impact evaluation (see Box 2).
- *Predictive function*: the impact theory can be used to support predictions on what is likely to happen. The theory can serve as guidance for assessing the likelihood that certain changes will come about in the future or will be sustained into the future. While these effects are unknown at the moment the outcome and impact evaluation is undertaken, certain conditions for sustainable impact can be identified which increase the likelihood of sustainable change (see below). This analysis will be strengthened when combined with an analysis of extrapolation of current trends into the near future.[25] Current work by the GEF EO on reviews of outcome to impact using a theory-based approach focus on these conditions (see GEF, 2009a).

It is beyond the scope of this report to provide a comprehensive theory-based analysis of project effects. To illustrate the value of an impact theory we briefly present a few corroborated aspects of the impact theory regarding the determinants of adoption behavior of silvopastoral practices:

- An important determinant of adoption behavior was the level of organization. In the Nicaraguan pilot site in two of the seven communities (San Ignacio and Paiwita) farmers were well organized in an association called '*Asociación de San Jose*'. Group solidarity and shared norms (in terms of mutual assistance during workshops or in the field, extensive knowledge exchange, group discussions, etc.) positively influenced adoption behavior.

[25] See for example the previous discussion on the comparison between those farmers receiving 2 years and those receiving 4 years of payments.

- Many farmers while receiving PES payments also had access to credit. For example, in order to effectively utilize the fodder banks, one would need a 'picadora' (equipment for processing grass into fodder). Nitlapán (in collaboration with a rural bank called the FDL) facilitated access to this equipment, respectively by offering a low interest loan (a so-called 'green loan' provided for these purposes) or directly delivering the equipment to the farmer to be repaid without interest over a period of two years. This to a large extent explains the high adoption rate of fodder banks in Nicaragua in comparison with the Costa Rican and Colombian site where access to this type of assistance was more limited.
- High milk prices had a positive effect on farmers' willingness and capacity to invest in silvopastoral land use practices. The additional incentive of these prices (next to the ex post payments of the environmental services generated by the silvopastoral land uses) reinforced farmers' preferences to invest in silvopastoral practices that generated a combination of low or intermediate environmental benefits (and corresponding relatively low amounts of PES) with direct economic benefits (e.g. improved pastures with trees), over land use practices that offered no productive benefits but high environmental benefits and corresponding high amounts of PES (e.g. reforestation).
- Small farmers are more likely to maximize value per unit of land, which explains why they had higher incremental ESI points per hectare than big farmers, who tend to maximize value per unit of labor.

Box 2. An example of a mixed-method design for impact evaluation

- E.g. a randomized experiment could be used to assess the effectiveness of different incentives (PES and TA) on land use change and subsequent environmental and socio-economic effects of these changes (potentially strengthens the internal validity of findings);
- E.g. survey data and case studies could tell us how incentives have different effects on particular types of farm households (potentially strengthens internal validity and increases external validity of findings);
- E.g. semi-structured interviews and focus group conversations could tell us more about the nature of effects in terms of production, environment, poverty, etc. (potentially enhances construct validity of findings).

5.3. Sustainability analysis

An impact theory provides a useful basis for developing an argumentation on the sustainability of changes brought about by an intervention. As the long-term results and the sustainability of the results that are visible at the moment of the outcome and impact evaluation are unknown and cannot be observed, evaluators will need to assess sustainability in an indirect manner. Outcome and impact evaluations can focus on other results that are observable in the short term, such as the institutionalization of practices and the development of organizational capacity, which are likely to contribute to the sustainability of outcomes and impacts for participants and communities in the longer term.

Mog (2004) describes different dimensions of sustainability relevant to projects such as the RISEMP project. The GEF Evaluation Office has developed a framework of five dimensions of sustainability that should be considered in outcome and impact evaluation (GEF, 2009b; see also GEF, 2009a). For each dimension we provide a brief example from the perspective of the RISEMP project. These examples should not be regarded as comprehensive and conclusive results on sustainability. They merely serve the purpose of illustration.

- *Socioeconomic sustainability*: The extent to which project activities lead to long term improvements in the social and economic situation where the project is found and where such changes are essential to ensure improved environmental management. In case of the RISEMP project, researchers at CATIE found that several of the promoted silvopastoral practices such as improved pastures with trees and fodder banks are privately profitable in the medium term. It is very likely that farmers will maintain and even increase those practices that are privately profitable.
- *Programmatic Sustainability*: The extent to which the actions that are taken during the life of the project continue after the formal project ends. At pilot site level the provision of PES and TA ceased by the end of the project. In different guises, the three local implementing organizations continue to support (a part) of the farmers that participated in the RISEMP project with such activities as environmental education, credits for

sustainable land use practices and technical assistance on topics related to silvopastoral land use.

- *Institutional Sustainability*: The extent to which necessary institutional structures are in place and secure for the long term as a result of the project. Institutional structures such as PES administration mechanisms and research and monitoring units ceased to exist after project closure. No new institutions were created in the pilot regions. However, an important result of the project has been the strengthening of capacities in the three implementing organizations: CATIE, CIPAV and Nitlapán.
- *Financial Sustainability*: The extent to which post-project activities can sustain themselves financially or mechanisms are in place to provide a constant flow of external financial resources. The project did not establish links with markets for environmental services such as they exist in some countries at local, national, or international level. Examples are local markets for hydrological services or international mechanisms such as the clean development mechanism (carbon sequestration).
- *Replication*: The extent to which successful implementation of actions in one project can be repeated in other project sites. Replication at pilot site level is fairly limited. Replication (replicatory effects) at other levels has been very successful. The project has generated many useful lessons for replication at other sites. Currently, an upscaled new version of the RISEMP project is under preparation in Colombia. The ESI, improved and validated during the course of the project, is currently being used in several new projects in different countries. Research results on the relationships between silvopastoral land uses and environmental benefits have been widely published.

6. Conclusions and recommendations

6.1. Conclusions

The Nicaraguan case shows how an experimental design that is implemented without the necessary knowledge and institutional support in the field can lose its utility. It should be emphasized that the problems with the experimental design are essentially strategic and planning failures and not implementation failures as such. Project staff were not trained or in any way prepared to manage an experimental design and could not be expected to deal with the various problems that threaten the validity of the design. The analysis shows that the utility of the experimental design in terms of resolving the attribution problem is heavily compromised by several threats to validity.

In all, the experimental framework failed on two of the three group comparisons that were to support rigorous claims on the effects of PES and technical assistance on land use change and corresponding environmental effects. The ‘PES only’ versus CG comparison is rendered invalid due to severe problems of selection bias and unintended behavioral responses (especially in the CG). The ‘PES only’ versus ‘PES TA’ comparison is rendered invalid due to problems of treatment diffusion. The ‘PES 2 years’ versus ‘PES 4 years’ comparison is quite valid. The data and their subsequent interpretation illustrate the utility of the experimental design in terms of providing reliable evidence on land use behavior under different types of incentives.

The fundamental question of the cost-benefit ratio of using an experimental design should be raised. Implementing such a design involves substantial costs:

- implementation costs: designing the experiment, selecting the farmers, managing and controlling the quality of the experiment, etc.
- costs in terms of facing ethical dilemmas or possible resistance from farmers or other stakeholders;
- foregone benefits to farmers (withholding benefits to certain groups of farmers, less outreach than without an experimental approach).

These costs can only be justified if the experiment is done carefully, thereby delivering its analytical potential. In the Nicaraguan case (and possibly the other two sites), the costs of implementing the experiment, without the necessary quality control and supervision clearly outweighed the analytical benefits of doing an experiment.

6.2. Recommendations

Despite the limited utility of the experimental design in Nicaragua and potential unidentified problems of the design in the other two countries, the logic of experimentation potentially provides a powerful tool to test the effectiveness of particular incentives on outcomes and impacts, controlling for other factors. Experiments can be especially useful in the following cases:

- when knowledge on attribution (and effectiveness) is important; for example in the case of innovative instruments when little is known about their effectiveness; in case there is a lot of existing evidence about the effectiveness of a particular approach or instrument then the benefits of an experimental design might not outweigh the costs;
- when there is an interest in the magnitude of effects (caused by the project).

However, they should only be applied:

- if sufficient attention and resources are dedicated to training and quality control of the experimental design in practice;
- if attention is paid to possible combinations of experimental approaches with other methods, which would reinforce each other and together would allow for a more comprehensive coverage of the outcome and impact dimensions of an intervention (as well as address more adequately questions of both average effects attributable to the intervention as well as heterogeneity in effects).

More specifically, we propose the following recommendations for future implementation of experimental designs in similar projects:

- Implementing an experimental design outside a laboratory in complex social environments requires a clear protocol and a shared vision among those actors involved in implementing the design on what the design is about and how it should be managed in practice. The different threats to the validity of the design should be considered before implementation.
- An experimental design (based on randomization) fundamentally affects the way an intervention is implemented. Therefore the Project Appraisal Document (PAD) should provide sufficient information on:
    - what the central hypotheses are to be tested with the experiment as well as other knowledge that is expected to be gained from the experiment;
    - the basic characteristics of the experiment;
    - how the experiment will be implemented;
    - the likely threats to validity and how they will be addressed;
    - a budget for staff training as well as for the incremental costs of managing the experiment;

    - an assessment weighing the costs against the benefits of experimentation.
- Provide training to project staff on management and quality control of experimental designs.
- Select a control group in a different region than participants/beneficiaries in order to avoid treatment diffusion; this avoids treatment diffusion effects, possible resentment and other unintended behavioral responses.
- In the communication to participants/beneficiaries try to avoid using references to the experimental design, as the idea of being part of an experiment might trigger a range of unintended responses. This might become easier if a control group is selected and monitored in a different (but similar) region.
- At national level, try to coordinate efforts by different institutional actors planning outcome and impact evaluations within (and possibly beyond) the agricultural and environmental sector. This might have several advantages:
    - Outcome and impact evaluation activities might be pooled and budgets and capacities shared, improving the prospects of high-quality evaluations;
    - Periodic surveys on target groups might incorporate variables relevant to particular interventions. As a result, individual outcome and impact evaluations can make better use of existing surveys, which is especially relevant in the case of quasi-experiments and regression-based quantitative outcome and impact evaluations.

References

Baker, J.L. (2000) *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*, The World Bank, Washington D.C.

Bamberger, M. (2006), *Conducting quality impact evaluations under budget, time and data constraints*, The World Bank, Washington D.C.

Bamberger, M. and H. White (2007) "Using strong evaluation designs in developing countries: experience and challenges", *Journal of Multidisciplinary Evaluation*, 4(8), 58-73.

Cook, T.D. and D.T. Campbell (1979) *Quasi-experimentation: design and analysis issues for field settings*, Houghton Mifflin, Boston.

GEF (2007a) *GEF Annual Impact Report 2007*, GEF Evaluation Office, Washington D.C.

GEF (2007b) *Protected Areas and Avoided Deforestation: A Statistical Evaluation*, prepared by K. Andam, P. Ferraro, A. Pfaff, and G. Sanchez-Azofeifa, GEF Evaluation Office, Washington D.C.

GEF (2009a) *Review of Outcomes to Impacts – Guidelines and Procedures*, prepared by the GEF EO in collaboration with the Conservation Development Center, GEF Evaluation Office, Washington D.C.

GEF (2009b) *GEF Annual Impact Report website*, http://www.thegef.org/gefevaluation.aspx?id=22444 [last consulted, June 17, 2009].

Ibrahim, M., M. Chacón, C. Cuartas, J. Naranjo, G. Ponce, P. Vega, F. Casasola and J. Rojas (2007) "Almacenamiento de carbono en el suelo y la biomasa arbórea en sistemas de usos de la tierra en paisajes ganaderos de Colombia, Costa Rica y Nicaragua", *Agroforestería en las Américas*, 45, CATIE, Turrialba.

Leeuw, F.L. (2003) "Reconstructing program theories: Methods available and problems to be solved", *American Journal of Evaluation*, 24(1), 5-20.

Mog, J. M. (2004) “Struggling with sustainability: A comparative framework for evaluating sustainable development programs”, *World Development*, 32(12), 2139–2160.

Morgan, S.L. and C. Winship (2007) *Counterfactuals and causal inference – methods and principles for social research*, Cambridge University Press, Cambridge.

OECD-DAC (2002) *Glossary of Key Terms in Evaluation and Results Based Management*, OECD-DAC, Paris.

Pagiola, S., P. Agostini, J. Gobbi, C. De Haan, M. Ibrahim, E. Murgueitio, E. Ram’rez, M. Rosales and J.P. Ruíz (2004) “Paying for Biodiversity Conservation Services in Agricultural Landscapes”, *Environment Department Paper*, 96, World Bank, Washington D.C.

Pagiola, S., E. Ramírez, J. Gobbi, C. De Haan, M. Ibrahim, E. Murgueitio, and. J.P. Ruíz (2007) “Paying for the environmental services of silvopastoral practices in Nicaragua”, *Ecological Economics*, 64(2), 374-385.

Rogers P.J., T.A. Hacsi, A. Petrosino and T.A. Huebner (Eds.) (2000) *Program theory in evaluation: Challenges and opportunities*, New Directions for Evaluation, 87, San Francisco, Jossey-Bass.

Rossi, P.H., M.W. Lipsey and H.E. Freeman, (2004) *Evaluation: A systematic approach*, Thousand Oaks, Sage Publications.

Shadish, W.R., T.D. Cook, and D.T. Campbell (2002) *Experimental and quasi- experimental designs for generalized causal inference*, Boston, Houghton Mifflin Company.

Weiss, C.H. (1997) “Theory-based evaluation: Past, present and future”, in D.J. Rog and D. Fournier (Eds.), *Progress and future directions in evaluation: Perspectives on theory, practice and methods*, New Directions for Evaluation, 76, San Francisco, Jossey-Bass.

White, H. (2006) *Impact evaluation – The experience of the independent evaluation group of the World Bank*, Independent Evaluation Group, World Bank, Washington D.C.

Wunder, S., S. Engel and S. Pagiola (2008) “Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries”, *Ecological Economics*, 65(4), 834-852.

Annexes

Annex 1. Summary logical framework RISEMP

Project Development Objectives	**Outcome / Impact Indicators**	**Outputs**	**Output Indicators**
To obtain local environmental benefits through reduction in erosion and improvement in soil and water quality while at the same time increasing production, income and employment in rural areas. To provide global environmental benefits, through improved biodiversity and carbon sequestration services. To gain initial experience in the management of incentives required to produce global environmental benefits. To develop recommendations for sector and environmental policies in terms of land use, environmental services and socio-economic development provided by the introduction of	Sustainable silvopastoral systems established in three Latin American countries and improved water quality in six watersheds in Latin America. Improved habitat for diverse types of bio-diversity provided and stable carbon sequestered in the soil and in commercial wood under silvopastoral systems in six watersheds in three countries. Improved resource monitoring methodologies developed for measuring carbon sequestration , biodiversity conservation developed and sustainable funding mechanism established which provide appropriate incentives to induce farmers to provide global environmental benefits. Increased awareness of the potential in	Increase in area with improved eco-systems functioning of 12,000 ha. currently degraded pasture land, as demonstrated by specific indicators for soil and water quality and bio-diversity The increase in numbers of livestock producers, community leaders, and policy decision makers at the local, regional and national level that are familiar with the ecological and economic benefits of more intensive silvopastoral systems in livestock production. The extent of dissemination of improved monitoring methodologies developed for measuring carbon sequestration, biodiversity conservation, water quality in watersheds and socio-economic aspects.	1.1.1 About 4000 ha silvopastoral systems, established, improving the eco-system in at least 12,000 ha to demonstrate the benefits of silvopastures for carbon sequestration and biodiversity in three countries. 1.1.2 Increased biodiversity conservation (at least 50 bird species/production system): 1.1.3 Increased carbon sequestration (about 25,000 ton carbon sequestered per year). 1.1.4 Increased water quality in watersheds (reduction on Biochemical Oxygen Demand (BOD) and suspended total solids (mg/l). 1.1.5 Increased socio-economic impact: Farm income to increase by 10 percent per year. 1.2.1 Local stakeholders trained in 3 countries (about 30,000 farmer days of training over 5 year period). 1.2.2 Local organization's capacity strengthened (20 organizations in 3 countries). 2.1 Methodologies to assess biodiversity, carbon sequestration, water quality on farm, watershed and community level and socio economic impact developed and tested. 2.2 Monitoring systems for biodiversity conservation, carbon sequestration, water quality using biological indicators and socio-economic impact established (monitoring systems in 3 countries).

silvopastoral systems to rehabilitate degraded pastures.	environmental services provided by integrated ecosystem management and experience gained for future development of the integrated ecosystem management approaches to restore degraded pasture. Guidance for future funding, lessons for replication/best practice, and policy requirements for environmental services in livestock production defined.	Understanding of farmer reactions to incentive systems for global environmental benefits obtained. The availability of policy guidelines to promote silvopastoral systems and establish sustainable benefits sharing mechanisms related to global and local environmental services provided by integrated ecosystem management.	3.1 Eco-Services payment systems implemented in each of the target countries. 3.2 Certification of ecological services conferred (results of monitoring analyzed at farm and landscape level, and environmental services paid to the farmers). 3.3 Farmers and community reaction to environmental services incentives and change of attitude and perception to local and global environment measured (measured by changes on land use, in particular in area set aside for forest regeneration). 4.1 Socioeconomic data available on key factors affecting farmer adoption of silvopastoral systems. 4.2 Alternative sources of funding for payment for eco-services, and alternative measures to promote silvopastoral systems identified and secured. 4.3. Specific recommendations for best ranching practices and land use that improve habitat heterogeneity to sustain higher biodiversity, and increase ranch yield disseminated among minimum 1200 farmers 12 NGO's and/or community-based groups, policy-makers and regional networks.

Source: PAD

Annex 2. Environmental Services Index (ESI) used by the RISEMP project

Land use	Biodiversity index	Carbon index	Total index
1 Crops (annual, grains and tubers)	0	0	0
2 Degraded Pasture	0	0	0
3 Natural Pasture without Trees	0,1	0,1	0,2
4 Improved Pasture without Trees	0,1	0,4	0,5
5 Semi-Permanent Crops	0,3	0,2	0,5
6 Natural Pasture + Low Tree Density (<30/ha)	0,3	0,3	0,6
7 Natural Pasture enriched with low tree density	0,3	0,3	0,6
8 Living Fences with new trees	0,3	0,3	0,6
9 Improved Pasture Low tree density	0,3	0,4	0,7
10 Fruit Crops (Monocrop)	0,3	0,4	0,7
11 Graminous Fodder Banks	0,3	0,5	0,8
12 Improved Pasture Low Tree Density	0,3	0,6	0,9
13 Ligneous Fodder Banks	0,4	0,5	0,9
14 Natural Pasture High Tree Density	0,5	0,5	1
15 Fruit Crops (Diverse)	0,6	0,5	1,1
16 Multistrata living fences	0,6	0,5	1,1
17 Diversified Fodder Banks	0,6	0,6	1,2
18 Commercial Tree Plantations (Monocultivation)	0,4	0,8	1,2
19 Shaded Coffee	0,6	0,7	1,3
20 Improved Pasture with High Tree Density	0,6	0,7	1,3
21 Guadua (bamboo) forest	0,5	0,8	1,3
22 Diversified Commercial Tree Plantations	0,7	0,7	1,4
23 Shrub habitats (tacotal)	0,6	0,8	1,4
24 Riparian Forest	0,8	0,7	1,5
25 Intensive Silvopastoral Areas	0,6	1	1,6
26 Secondary Forest (intervened)	0,8	0,9	1,7
27 Secondary Forest	0,9	1	1,9
28 Primary Forest	1	1	2

Source: Based on RISEMP data

Annex 3. Persons interviewed

Nicaragua (March, April and July, 2008)

Farmers		
Name	**Community**	**Group**
Absalon Guerrero	Paiwita	PSA+AT (4)
Agusto Robles	Limas Abajo	PSA+AT (4)
Albertina Jarquín	Patastule/El Gavilan	PSA+AT (4)
Alberto Saravia	San Ignacio	PSA+AT (4)
Angela Alvarado	San Ignacio	PSA+AT (4)
Carlos Urbina Luna	Limas Arriba	PSA+AT (4)
Eusebio Mendoza Dias	San Ignacio	PSA+AT (4)
Guillermo Garcia Polanco	Patastule	PSA+AT (4)
José Andrés Amador Martinez	San Ignacio	PSA (4)
José Rolando Castillo Ramirez	Patastule	PSA+AT (4)
Juan José Jarquín Jarquín	Limas Abajo	PSA (4)
Julia Gadea Amador (Jaime Robles)	Limas Abajo	PSA+AT (2)
Julio Gutierrez Obando	San Ignacio	CONTROL
Kairo Torres	Patastule	/
Orlando Urbina	El Gavilan	/
Pastor Flores Rodriguez	Paiwita	CONTROL
Pedro Reyes Urbina	Patastule	CONTROL
Pedro Talavera Valle	Paiwita	PSA+AT (4)
Hector René Zeledon Alvarado	Patastule	CONTROL
Richard José Robles Ortega	Limas Abajo	PSA (4)
Roberto Urbina	El Gavilan	PSA+AT (4)
Rosario Ramirez García	El Gavilan	PSA (4)
Santos Genaro Sevilla Suarez	El Gavilan	PSA+AT (4)
Severino Vega Martinez (Fermin del Socorro Vega Vega)	Limas Abajo	PSA+AT (4)
Tomas Castro Torres	Limas Abajo	CONTROL
Tomas Soza Morales	Limas Abajo	PSA+AT (2)
Trinidad Lanzas (Ramona del Socorro Garcia)	Limas Arriba	PSA+AT (4)
Victorina Ortega Mondoy	Limas Abajo	/
Zoyla Martinez Rubio	Patastule/El Gavilan	CONTROL

Project staff	
Alfredo Arguëllo	Project staff, Nitlapan
Omar Davila	Project staff, Nitlapan
Yuri Marin	Socio-economic analyst, Nitlapan
Guillermo Ponce	Carbon sequestration analyst, Nitlapan
Elias Ramirez	Coordinator Nicaraguan pilot site, Nitlapan
Bismark Reyes	Water analyst, Nitlapan

Costa Rica (July, 2008)

Project staff and other stakeholders

Oliver Bach	Rainforest Alliance
Francisco Casasola	Coordinator Costa Rican pilot site, CATIE
Omar Davila	Project staff, Nitlapan
Leonardo Guerra	Consultant carbon sequestration
Muhammad Ibrahim	Project leader, CATIE
Jose Ney Rios	Water analyst, CATIE
Joel Saenz	Biodiversity specialist
Claudia Sepulveda	Project staff, CATIE
Diego Tobar	Project staff, CATIE
Cristobal Villanueva	Socio-economic analyst, CATIE

World Bank

Gunars Platais	Environmental economist
Cees De Haan (2007)	Silvopastoral specialist – team RISEMP
Stefano Pagiola (2007)	Environmental economist – team RISEMP

Annex 4. RISEMP project experimental design (situation 2006)

Group	Colombia	Nicaragua	Costa Rica	Total
Group A	29	27	28	84
Group B	50	75	69	194
Group C	25	29	27	81
Total	**104**	**131**	**124**	**359**
According to type of payment scheme*				
Payment Scheme 1 (4 years)	36	75	50	151
Payment Scheme 2 (2 years)	39	39	46	124
Total	**75**	**104**	**96**	**275**

Group A = control (without payments nor technical assistance); Group B = (payments and technical assistance); Group C = only payments.
* Only groups B and C.

Source: progress report RISEMP, 2006

Annex 5. Comparing the three sites: a broad impression

The other two cases were not analyzed in detail in this study. We restrict ourselves to some broad impressions about the quality and implementation of the design.

On the basis of interviews with staff from CATIE and looking at the data, our impression is that the experimental design in the Costa Rican site has been the least affected by the different threats to validity. The Nicaraguan case has been the most problematic one. The Colombian design can be positioned in between the others.

Notwithstanding the design-related problems in Nicaragua, the longstanding relationship between Nitlapán and many of the farmers had a positive effect on data quality. In contrast, data quality has been particularly problematic in Colombia, especially in the control group. One of the reasons for this has been the history of conflict and drug trafficking in the country, fueling an atmosphere of distrust between farmers and institutions. In terms of data quality, Costa Rica would be the intermediate case, with the Nicaraguan and the Colombian sites respectively having had the least and the most problems with data quality.

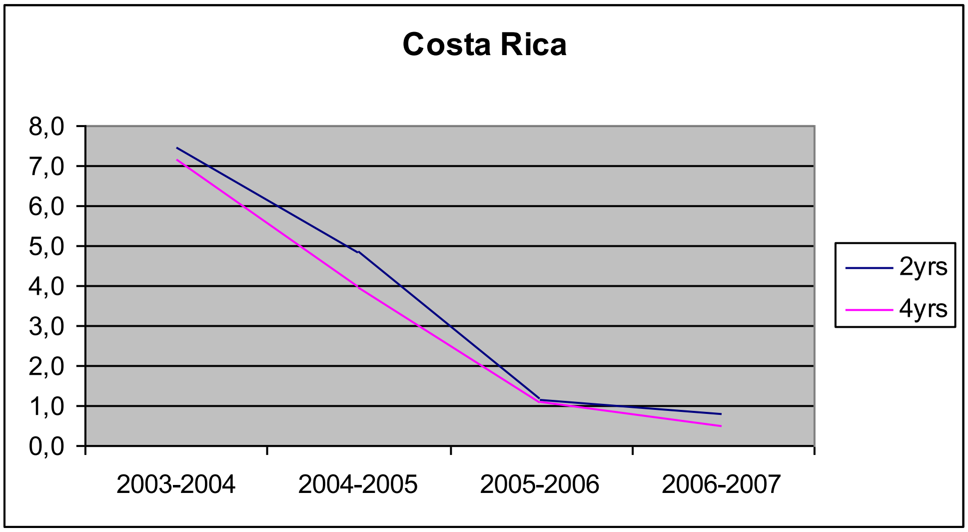
As for selection bias, the general impression is that randomization principles were more successfully implemented in both Costa Rica and Colombia, reducing the probability of selection bias.

With respect to contagion, in Costa Rica geographical proximity between farmers has probably facilitated processes of farmer to farmer diffusion. In Colombia, there was a clearer geographical separation between PES and CG farmers, resulting probably in less farmer to farmer diffusion (contagion). In addition, apart from CIPAV no other institutions in the region were providing technical assistance on silvopastoral practices or related issues. In the Costa Rican site there were a few other institutional actors, but not as many as in the Nicaraguan case.

To sum up, group comparisons for the Costa Rican and Colombian case are likely to provide a clearer picture of the effect of project incentives on adoption behavior of silvopastoral practices than in the Nicaraguan case. However, given the evidence on design implementation problems in the Nicaraguan case, using the experimental design for group comparisons in the other two countries without further insight into possible design implementation issues would seriously undermine the credibility of findings.

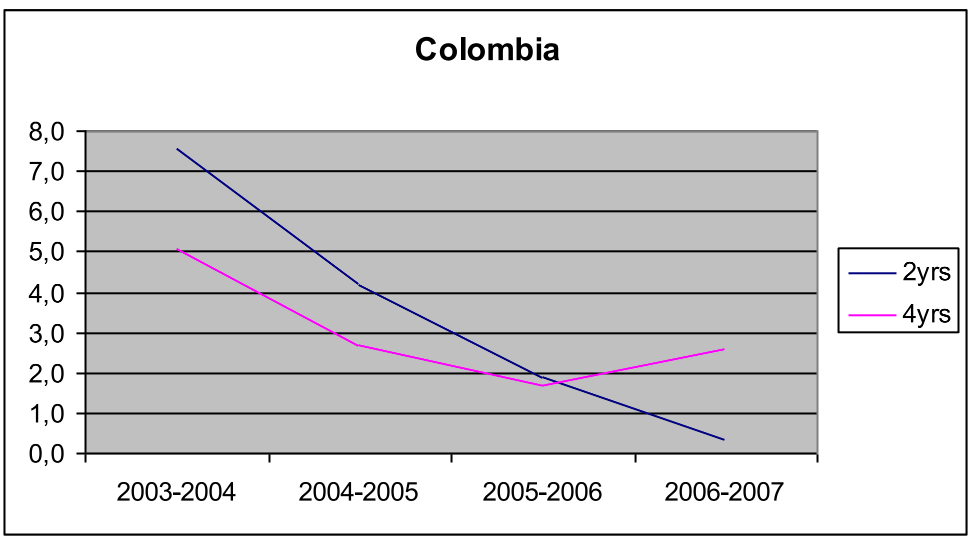
Annex 6. Group comparisons between farmers receiving 2 years versus 4 years of PES, Costa Rica and Colombia

Figure A6.1. Costa Rica



Source: own calculations based on RISEMP project data, January 2008

Figure A6.2. Colombia



Source: own calculations based on RISEMP project data, January 2008