Chapter 22

Learning while doing

Evaluating impacts of REDD+ projects

Pamela Jagger, Stibniati Atmadja, Subhrendu K. Pattanayak,
Erin Sills and William D. Sunderlin

- REDD+ projects require an impact assessment approach to estimate emissions and removals; for REDD+ to succeed we need information on this and the associated 3E+ outcomes.
- There are few examples of rigorous impact assessment in the conservation, avoided deforestation and payments for environmental services (PES) literature. REDD+ impact assessment could contribute tremendously to our understanding of successful environment and development policy initiatives.
- We will learn more rapidly and effectively by sharing evaluation designs and findings across REDD+ projects.

How will learning from projects improve REDD+?

We have a narrow, but critical, window of opportunity to evaluate and learn from the experience of first generation REDD+ projects. By gathering evidence on processes and outcomes, we will learn what causes REDD+ projects to succeed or fail. REDD+ is a unique opportunity to share the lessons we learn, because of the global distribution and relatively coordinated timing of projects, significant allocation of financial resources, and clear objectives and explicit mandate set by international negotiators. This chapter directs donors, regulators, and project proponents and developers to ways in which we can learn from evaluating projects. We advocate for serious attention and financial resources to be committed to independent process and impact assessment of first generation REDD+ projects.[1] By definition, REDD+ projects are performance based and therefore evaluate their effect on changes in carbon stock in comparison to a reference level. In this early phase of developing REDD+ policy, it is also crucial to examine, evaluate and share findings on the effects and distribution of co-benefits and costs, i.e., according to the effectiveness, efficiency and equity plus co-benefits (3E+) principle (Chapter 1). This broad assessment of project outcomes and processes is critical for learning how changes in forest carbon happen and what causes them.

Identifying and designing methods to facilitate learning from the hundreds of REDD+ projects expected to be implemented over the next few years is not easy. Projects take different approaches, operate at different scales, and are implemented across diverse settings, as clearly spelled out in Chapter 21. Nevertheless, if we invest time and resources in evaluating a representative sample of REDD+ projects using state-of-the-art methods, and if we share our findings among projects and regions, we will learn lessons that will help ensure the success of REDD+.

This chapter makes the case for rigorous empirical evaluations of REDD+ projects, so that we can learn if and how they reduce emissions or increase removals and deliver 3E+ outcomes. We discuss how REDD+ evaluations can contribute to our empirical knowledge and give examples of rigorous impact assessments of natural resource and conservation policies (e.g., payments for environmental services (PES) schemes, avoided deforestation polices, decentralisation reforms and protected area (PA) management). We conclude that the success of REDD+ rests crucially on sharing evaluation designs and findings across REDD+ projects so that we learn more rapidly and effectively.

Why do we need to evaluate REDD+ projects?

Intergovernmental Panel on Climate Change (IPCC) guidelines and voluntary certification standards require that REDD+ projects rigorously evaluate their effect on net carbon emissions (see Chapter 7 on monitoring, reporting and verification (MRV)). This provides a starting point for assessing the impacts of REDD+ projects, not only on carbon, but also on socio-

1 In this chapter, we use the term 'evaluation' to refer broadly to the analysis of public policies. The term 'impact assessment' refers to a specific set of research designs and methods for assessing and understanding outcomes of public policies.

economic and environmental outcomes. While collecting baseline data and regularly monitoring projects, as required for MRV, we have the opportunity to collect data to help us understand the underlying causes of these outcomes. The requirement to verify outcomes makes REDD+ projects inherently different from traditional forest sector development projects. The design of REDD+ projects, combined with the allocation of large sums of money for monitoring and evaluation, is a unique opportunity to significantly improve our knowledge, not only on REDD+, but on development and environment interventions more generally.

There are four reasons to evaluate REDD+ projects using impact assessment methods:

1. REDD+ projects have to assess impact. The Bali Action Plan requires REDD+ projects to measure changes in net carbon emissions that result from project activities;
2. Project proponents and donors need to know what the 3E+ outcomes are, and what tradeoffs between conservation and livelihoods are associated with the outcomes;
3. For REDD+ to gain broad acceptance, it has to work on the ground. Impact assessment can deliver hard evidence on whether or not projects are meeting their goals, and allow project to make adjustments as they go along;
4. While we can learn a lot from individual projects, a common, systematic approach to evaluating REDD+ projects will facilitate learning, and allow comparison of the various factors that influence 3E+ outcomes across projects. A common, systematic approach to evaluation will allow:
   - Site conditions and project design elements associated with 3E+ outcomes to be identified;
   - Rigorous evaluation to inform the design of national policies and processes that enable and guide REDD+; and
   - Practitioners and academics to learn about the effectiveness of alternative conservation instruments, including PES.

Learning tools

Process assessment and impact assessment are tools for understanding causal mechanisms underlying observed outcomes. These tools help us extract timely, persuasive and relevant lessons from projects to inform the policy process. They can and should be part of the mix of monitoring and evaluation methods (Margoulis *et al.* 2009). Table 22.1 shows research designs and data collection requirements for assessing process and impact.

Table 22.1. Options for assessing REDD+ projects

Level of effort and resources	When to design assessment strategy	When to collect data			Process learning
		Baseline	Post-intervention	Controls	
High	Before project implementation	Yes	Yes	Yes	Throughout
Medium	Before project implementation	Yes	Yes	Yes	Some
	Before or in early stages of implementation	Yes	Yes	No	
	In early stages of implementation	No	Yes	No	
Low	After project implementation	No	Yes	No	Limited or none

Process assessment

Process assessment involves documenting and analysing project implementation. Since implementation often deviates from project plans, process assessment is essential to track actual activities, their sequence, course corrections and the actors involved. Process assessment for REDD+ projects is likely to document: how proponents engage with local communities and other forest stakeholders; land, forest and carbon tenure arrangements; stakeholder power relations; logistical aspects, including budgeting; baseline data collection; verification and audit processes; and the direct costs of project implementation. Collecting data at the start and throughout the project is fundamental for evaluating processes, and for understanding why the project did or did not attain its objectives. In cases where rigorous evaluation designs are not possible due to logistics, political considerations or cost, process assessment can provide important data for evaluation based on conceptual models of how interventions generate outcomes.

Impact assessment

The main components of impact assessments are: 1) measuring outcomes after an intervention (e.g., a REDD+ project), and 2) comparing the observed outcomes with the counterfactual, i.e., what would the situation have been without the intervention. To learn from impact assessments, we must understand why we observe particular outcomes. In other words, impact assessments should measure and interpret what causes the effects of interventions. Impact assessments are increasingly used to evaluate social policies and development projects (Leeuw and Vaessen 2009; World

Bank 2009f) and researchers have called for the same approach to evaluate environmental and natural resource policies (Bennear and Coglianese 2005; Frondel and Schmidt 2005; Ferraro and Pattanayak 2006). An ideal impact assessment has four steps: 1) identifying key parameters; 2) collecting data; 3) rigorous evaluation of the data (beyond the scope of this chapter, but see Box 22.1 for references); and 4) disseminating and acting upon the findings.

Box 22.1. Web resources for learning state of the art evaluation techniques

Process assessment

Wageningen University has a website devoted to tools and methods for participatory planning, monitoring and evaluation, http://portals.wi.wur.nl/ppme/content.php?Tools_%26_Methods.

The National Science Foundation has produced a user friendly handbook for mixed method evaluations, http://www.nsf.gov/pubs/1997/nsf97153/start.htm.

The Conservation Measures Partnership and Benetech have developed adaptive management software for conservation projects, www.miradi.org.

Outcomes assessment

The Network of Networks Impact Evaluation Initiative (NONIE) of the World Bank has a series of publications that provide guidance on impact evaluation, http://www.worldbank.org/ieg/nonie/guidance.html.

The International Initiative for Impact Evaluation provides discussion and suggested methods for impact evaluation, http://3ieimpact.org/page.php?pg=resources.

The website of the Independent Evaluation Group at the World Bank provides overviews of evaluation methodology and examples of state of the art evaluations, http://www.worldbank.org/oed/.

Evaluation of conservation and natural resource interventions

Pattanayak (2009) has produced the 'Rough Guide to Evaluation of Environmental and Development Programs', http://www.sandeeonline.com/uploads/documents/publication/847_PUB_Working_Paper_40.pdf.

A special issue of *New Directions for Evaluation* focuses on Environmental Program and Policy Evaluation: Addressing Methodological Challenges, http://www3.interscience.wiley.com/journal/122445950/issue.

Here, we focus on the before-after-control-impact (BACI) design for impact assessment, which estimates impacts using data collected before and after, and from both control and intervention sites.

Regardless of the design, an impact assessment can only provide clear answers if the key questions, variables and outcomes of interest are clearly formulated. Evaluators need to identify:

- The intervention to be evaluated (e.g., REDD+ project activities, excluding any national policy changes in support of REDD+);
- Specific outcomes of interest (e.g., changes in carbon emissions and income derived from the forest);
- Observable indicators of those outcomes (e.g., changes in forest cover and household wealth);
- Observable process indicators that characterise how the intervention is implemented (e.g., maps of tenure and forest use, number of field visits to monitor compliance); and
- Confounding factors that vary within the site and control areas and influence the outcomes of interest (e.g., market access, population density, average annual rainfall).

Collection of baseline[2] data 'before' project implementation facilitates a rigorous impact assessment, because it allows the changes in outcomes before and after the intervention to be estimated. Over a short time, and when there are relatively few other policy, economic or environmental changes, the baseline could be considered to be the counterfactual. This means that nothing would have changed without the intervention. Much of the existing literature on avoided deforestation relies on extrapolating historical trends (e.g., past 5–10 years) or projections that modify historical trends by including other variables. However, the ideal evaluation design is to collect baseline data on key outcome variables and their determinants from both project (treatment) and control sites (see also Figure 22.1).

Advance planning, in addition to allowing collection of baseline data before the project begins, can add to the rigour of the impact assessment by identifying or even creating 'control' groups that are similar to the treatment group but not directly affected by the intervention. Evaluators can scope for areas that are similar to the project site in terms of biophysical, demographic and socio-economic characteristics to serve as control areas. Scoping can also identify areas outside project boundaries that may be affected by leakage.

2 The term 'baseline' has several meanings in the REDD+ debate. In line with common use in the evaluation literature, in this chapter, we use the term 'baseline' for the 'pre-intervention site conditions', not in the sense of a prediction about the future.

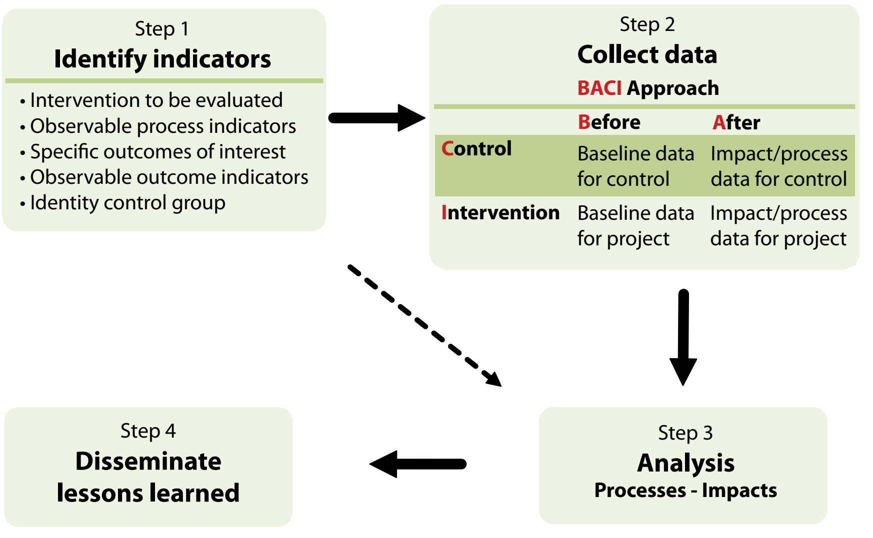

Figure 22.1. The BACI design for assessing REDD+ projects

The final and critical stage of impact assessment is disseminating and acting on the findings. Independent evaluators should ensure that they share the results with project proponents and other stakeholders in supportive ways, and with other projects through networks for joint learning. Project proponents who share lessons from both the successes and the failures transparently with donors, national governments and the global community will encourage widespread acceptance and implementation of REDD+ as a climate-change mitigation strategy. Effective dissemination means producing a range of products for different audiences. These would include reports, in appropriate format and language, for communities, policy makers and peers via the internet and peer-reviewed journal articles.

The BACI design has pitfalls. A key assumption is that it is possible to find control sites that are close enough to the project site to be similar, yet far enough away to ensure that the project has no influence on forest users' behaviour. Failure to find such control sites undermines the basic premise of the BACI design. Further, even the most rigorous impact assessment, using an ideal BACI design for a single site, will not necessarily provide insights into the reasons for the observed outcomes. To learn about the factors that influence outcomes, it is important to 1) compare findings across REDD+ projects evaluated using similar BACI designs, and 2) triangulate findings using contextual information to understand the processes that lead from project implementation to project outcomes. Quantitative-data collection cannot fully capture information on the context; qualitative process learning, using techniques such as participatory rural appraisal and key informant interviews, throughout project implementation is crucial. Methods for process and impact evaluation have been documented in numerous guides to evaluation (see Box 22.1).

Learning from previous impact assessments

The literature on evaluating natural resource management and conservation policy reforms provides important lessons for assessing REDD+ projects. There are few rigorous evaluations of designs and methods to assess the causal effects of conservation investments (Ferraro and Pattanayak 2006). For example, most evaluations of PES schemes are qualitative case studies drawing on records of government and non-governmental organisations, reviews of grey literature, key informant interviews and rapid field appraisals (Pattanayak *et al.* 2009). The most common quantitative empirical assessments use *ex post* data on both treated and control units (e.g., households or watersheds inside and outside the REDD+ project boundary). If the sample is large enough and there is sufficient variation in the data, this kind of assessment allows for multivariate regression of outcomes on treatment status (e.g., whether there was participation in the REDD+ project) to control for potential confounding factors. This research design does not require the analyst to explicitly identify an appropriate control group and thus the results may rely on extrapolation across very different treated and non-treated units.

'Matching' methods, developed to address these issues, are increasingly being used to evaluate the outcomes of natural resource and conservation related policies. They have been used to study the causal impact of individual, transferable quotas on the collapse of fisheries worldwide (Costello *et al.* 2008); moratoria on development in the USA (Bento *et al.* 2007); the effect of protected areas on forest cover in Costa Rica (Andam *et al.* 2008), Sumatra (Gaveau *et al.* 2009) and globally (Nelson and Chomitz 2009); payments for ecosystem services on forest cover in Costa Rica (Arriagada 2008; Pfaff *et al.* 2008); decentralised management on forest cover in India (Somanathan *et al.* 2009); and devolution of forest management on household income from forests in Malawi (Jumbe and Angelsen 2006). The most rigorous of these evaluations apply matching methods to changes in outcomes (before and after the intervention), sometimes reconstructed through secondary or recall data (which can be difficult). This emphasises the importance of collecting baseline data. Even when considering changes in outcomes, matching methods assume that all factors influencing both programme participation and outcomes (e.g., determinants of participation in a REDD+ project and deforestation rates) are observed, measured and used in the matching process. In fact, it can be very difficult to reconstruct the process of selecting sites and recruiting participants *ex post*. Thus, even if the evaluation plan is to apply matching methods *ex post* to measures of final outcomes, process assessment early in the project is critical.

Box 22.2. Examples of state of the art evaluations relevant to REDD+ projects

Measuring the effectiveness of protected area networks in reducing deforestation (Andam *et al*. 2008)

Andam *et al*. (2008) evaluate the effect of Costa Rica's protected-area (PA) system on deforestation using matching methods that compare outcomes on very similar protected and unprotected forest plots. They match forest plots using a technique called covariate balancing of baseline variables (i.e., covariates include high, medium or low land productivity; distance to forest edge, road and city). They find 10% of protected forests would have been deforested had they not been protected. Without controlling for covariates through covariate matching, the result would have been 44%. The difference in findings is because protected areas are generally less accessible, and have lower agricultural productivity.

Evaluating whether protected areas reduce tropical deforestation in Sumatra (Gaveau *et al*. 2009)

Gaveau *et al*. (2009) examine the effect of PAs on deforestation. They combine an analysis of remote sensing images with field-based methods to assess changes in forest cover in Sumatran PAs arising from agricultural encroachment, large-scale mechanised logging, and forest regrowth. They match PAs (i.e., treatment groups) and areas around PAs (i.e., control groups), before and after PAs were established, based on the 'propensity score' of protection (which essentially is based on a statistical model of pre-establishment forest cover, slope, elevation, roads and size of forest edge). The matched comparison suggests that PAs reduced deforestation by 24% from 1990 to 2000, whereas a naïve (i.e., simple mean differences) comparison of PAs and adjacent areas would suggest that PAs reduced deforestation by 59%. As in the case of Andam *et al*. (2008), the overestimation stems from not accounting for the non-random location of PAs in Sumatra ('passive protection').

Income after Uganda's forest sector reform: are the rural poor gaining? (Jagger 2008)

Jagger (2008) uses data from households living adjacent to three major forest sites in western Uganda to assess the effect of Uganda's forest sector decentralisation reform on rural livelihoods. Detailed income portfolio data collected immediately prior to the reform are compared with data collected four years after reform implementation. The decentralisation reform did not affect forest management in one of the forest sites; this site serves as a control in the design. The difference-in-difference method is used to estimate the effect of the reform. Changes in control sites are subtracted from changes in treatment sites. Covariates used in regression models allow for the control of exogenous factors that influence outcomes. The findings demonstrate that the reform has had a limited effect on livelihoods overall, but that the relative importance of forest income has declined for poor households and increased for relatively wealthy households.

Figure 22.2. Reporting research findings to the community, western Uganda
(Photo by: Paul Sserumaga)

The small but growing literature on evaluating various policy reforms related to natural resource management and conservation provides important lessons for those assessing the impact of REDD+ projects:

- Rigorous methods and traditional case study methods often deliver different results;
- Different (potentially complementary) ways to identify control groups include 1) random selection of intervention and control groups; 2) matching and other quasi-experimental methods; and 3) selecting non-treated groups with purposive criteria (i.e., market access, population density and forest type);
- Although baselines can be constructed retrospectively, collecting baseline data before the project begins is much more reliable than informant recall or secondary data;
- Ground-truthing and collecting household data give important insights into project outcomes that remote sensing methods cannot measure.

In addition to being useful for evaluation, data collected at intervals on the same units – or panel data – are critical for understanding dynamic processes such as poverty, migration and the evolution of land use on tropical forest frontiers. Recognising this, an increasing number of research initiatives and studies are collecting panel data for both biophysical and socio-economic indicators in tropical forest zones (see examples in Box 22.3). Some REDD+ projects

Box 22.3. Examples of global and local or regional scale datasets with environmental and socio-economic baselines

Global scale

International Forestry Resources and Institutions: Data from over 300 forest sites throughout the developed and developing world. Data on biophysical indicators of forest conditions and community forest institutions. http://www.sitemaker.umich.edu/ifri/home.

Poverty Environment Network: Detailed, quarterly household data on forest use and income portfolios from about 9000 households in 40 sites in 26 countries throughout the low- to medium-income tropics. http://www.cifor.cgiar.org/pen/_ref/home/index.htm.

Local or regional scale panel data studies of livelihoods and environmental change

Nang Rong Projects, University of North Carolina at Chapel Hill: Demographic, social and land use and land cover data for the past 20 years from Nang Rong, Thailand. http://www.cpc.unc.edu/projects/nangrong.

Ouro Preto do Oeste, University of Salisbury in collaboration with North Carolina State University and UC Santa Barbara: Socioeconomics and land use and land cover data in four waves from 1996 to 2009 from an old frontier in the Brazilian state of Rondônia. http://facultyfp.salisbury.edu/jlcaviglia-harris/NSF/NSF-SES-0452852.htm.

TAPS, Brandeis University: Socioeconomic, cultural, environmental and multidimensional indicators of well-being among the Tsimane in the Department of Beni, Bolivia. http://www.tsimane.org/.

may be able to use baseline and control-group data from these studies. More importantly, these research initiatives offer research tools (e.g., socio-economic household survey instruments, methods for ground-truthing land use cover change findings from remote sensing analysis) and lessons for evaluating REDD+ projects. For example, some initiatives have tracked households for many years and have tested ways to reduce attrition and to systematically update research instruments to reflect new activities and concerns. Studies that collect data across multiple sites, such as those conducted by International Forestry Resources and Institutions and the Poverty Environment Network, have had to balance collecting data consistently (to enable global comparisons) with adapting survey instruments and procedures to local circumstances.

These studies have also had to demonstrate their external validity, that is, to show that the sites they monitor are representative and that their results are generally applicable.

Learning as we move forward with REDD+ projects

Learning from REDD+ projects has payoffs in improving the projects themselves, improving the national policies and processes that guide REDD+, and in laying the foundation for effective, efficient and equitable implementation of REDD+ post-2012. Policy makers and donors should bear this in mind to get REDD+ off to a good start.

Our recommendations to project donors, regulators, proponents, developers and researchers are:

- Collect basic forest and socio-economic data before starting projects and after project implementation;
- Identify how outcomes will be measured and what variables are important to explain outcomes;
- Collect data at regular intervals during project implementation to help understand process and progress;
- Include control sites where possible;
- Invite and collaborate with independent or third-party evaluators and researchers; and
- Strive to make the design and findings of REDD+ project evaluations transparent for all stakeholders.

We recognise that the cost of our proposed mode of learning is potentially high, but we argue that the payoffs (and the costs of not learning) are large, both for project proponents and the global community. Those funded to generate international public goods that identify lessons from the first generation of REDD+ projects should also be funded to do rigorous evaluation research. Suppose that the global REDD+ effort in its first few years costs US $10 billion and that a concerted research and learning effort on REDD+ projects improves the efficiency by a very modest 5%, then the saving of US $500 million far exceeds the cost of learning. Such investment opportunities are rare!