Chapter 15
Missing the Forests for the Trees? Assessing the Use of Impact Evaluations in Forestry Programmes

Jyotsna Puri and Bharat Dhody

Abstract In this chapter, we examine how impact evaluations can contribute to measuring and understanding the overall effectiveness, efficiency and sustainability of forestry programmes. In most cases we find that impact evaluations have used quasi-experimental methods rather than experimental methods to identify and measure change that can be causally attributed to forestry programmes. We conclude that in measuring the change that be attributed to these programmes, impact evaluation methods help to measure the overall effect, deal with sources of potential bias and mitigate confounding factors while undertaking these measurements. Impact evaluations also hold enormous potential because they are able to leverage the potential held by big and open data. However caution must also be exercised in using these methods. Impact evaluation methodologies must also incorporate causal pathways and methods of implementation research if they are to be relevant to policy and programme managers.

Keywords Forests • Impact evaluation • Experimental methods • Quasi-experimental methods • Adaptation • Sustainable development • Big data

15.1 Introduction

Forests contribute to greater resilience and reduced vulnerability of ecosystems (UNEP 2011). They provide important ecosystem services, influence micro weather systems, are an important carbon sink that in the long run mitigates the risk of climate change and constitute an important part of an ecosystem themselves (World Bank 2015; Wunder et al. 2014). Therefore knowing what policies and programmes are important in preserving the health of forestry systems is an important question for anyone wanting to discuss a sustainable economy.

J. Puri (✉) • B. Dhody
International Initiative for Impact Evaluation (3ie), New Delhi, India
e-mail: jpuri@3ieimpact.org

J.I. Uitto, R. Shaw (eds.), *Sustainable Development and Disaster Risk Reduction*, Disaster Risk Reduction, DOI 10.1007/978-4-431-55078-5_15

The chapter lays out a paradigm for evaluating forestry projects, policies and programmes. It examines how experimental methods and quasi-experimental methods can be used to understand the effectiveness of the forestry sector and discusses their limitations and possible solutions. We then make some recommendations for methodology.

Overall we find that, first, not surprisingly there have been few studies that have used robust attribution methods to assess the impacts of programmes on how well forests are adapting and are sustainable; second, we find that there are big contributions that impact evaluation methodologies can make to the field. However there are also many limitations in traditional experimental and quasi-experimental methods that can limit the understanding of impacts in these multi-intervention and multi-sectoral contexts. Finally, we present possibilities in methodology and data that represent an important way forward.

The chapter is laid out as follows. 15.2 lays out the objectives including a brief exposition of the forestry sector. 15.3 lays out the definition of attribution methods and discusses briefly what impact evaluations are and why they can be powerful for evidence informed policy. 15.4 discusses the main contributions that impact evaluation methodologies and studies can make to this sector. 15.5 discusses some limitations of current approaches. 15.6 discusses some possible solutions and opportunities and 15.7 concludes.

15.2 Objectives

The objective of the chapter is to examine the experience of impact evaluations in the field of forestry. While examining these methods we examine the strengths and limitations of impact evaluation methodologies used in this field. Impact evaluation evidence can be potentially used to inform the extent to which forestry programmes are relevant, effective, replicable and scale-able. They thus respond to the needs of programme managers who are being increasingly asked to provide high quality evidence of whether their programmes are working or not, why and for whom and how much. They also respond to the needs of policy makers to help them assess the effectiveness of forestry programmes in general and to assess in a more robust manner the trade-offs they are required to consistently make (UNEP 2011). In examining impact evaluation studies we also take a step back to see objectively if these are indeed meeting the needs of the sector.

Additionally we also discuss how current impact evaluation practices and methodologies can be extended to attend to the needs of the forestry sector in particular, and climate change discipline in general, to better assess questions of timing, threshold effects, internal validity, complexity and external validity. Since this field is rapidly evolving, the chapter examines how current methods of impact evaluations may be modified to cater to the different needs of the sector. Indeed an overall and practical consideration is cost – impact evaluations are expensive and clearly it is neither required nor possible to undertake impact evaluations of all programmes in this area, given that the average cost of impact evaluations is $450,000 (Puri et al. 2015).

The questions that this study answers are the following:

- What can impact evaluations of forestry programmes tell us?
- What additional value do impact evaluation studies for evaluating forestry programmes provide over traditional process evaluations?
- What are the limitations of current impact evaluation approaches and how can they be made more relevant to the forestry sector?

In this chapter we use a shortlist of impact evaluations that examine forestry programmes, policies or projects to illustrate some of the contributions and limitations of impact evaluations (see Annex II). The list of these studies is not exhaustive. Indeed it is purposive. However they are all impact evaluation studies – they all use experimental or quasi-experimental methods to measure changes caused by forestry programmes.

15.3 Impact Evaluations and Definitions

In recent times the use of experimental and quasi-experimental methods to understand changes that can be attributed to programmes and policies has become popular also see Bowler et al. (2010).

The reason impact evaluations have gained a lot of traction in recent times is because 'identification strategies' allow us to clearly measure the amount of change that can be attributed to the programme, while dealing with confounding factors and potential endogeneity, that may have bias results (see for example Gertler et al. 2011). They thus make it possible to measure this effect but also therefore truly compare programmes across different settings (see Table 15.1). Impact evaluations use either experimental or quasi-experimental methods to identify and measure these effects and also deal with selection bias and programme placement bias (programmes are placed in areas where they are likely to be most successful, which means that what you are measuring when measuring effectiveness is the conflated effect of the programme *and* the context. Since context is not something programme managers typically have control over, this gives biased measures of the success of the programme or its failure).[1] In Annex I, we define some important terms.

[1]This is the sense in which we use the term 'impact evaluations'. Other definitions also exist. Different definitions of impact evaluations emphasize different aspects of the causal chain. The OECD-DAC defines it as the 'The positive and negative, primary and secondary long term effects produced by a development intervention, directly or indirectly, intended or unintended.' The World Bank defines it as 'assessing changes in the well-being of individuals, households, communities or firms, that can be attributed to a particular programme, policy or work'. We use the meaning of impact evaluations as the one used by 3ie: 'Rigorous impact evaluation studies are analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific programme using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context.'

Table 15.1 Types of identification methods that may be used in impact evaluations

Methods	Description	Pros	Cons
I. Experimental design all of which require randomized assignment			
(a) The basic randomized control trial (RCT)	A sample of eligible subjects (areas, people, communities) are randomly assigned into those who receive the programme or intervention and those who do not. Impact is the difference in outcomes between the two groups (i.e. between the treatment group and the comparison group).	Straight forward estimation (difference in statistical averages)	Requires a comparison group;
		It is commonly argued that these designs do not require baseline data but these are usually desired so that balance between different arms can be checked.	Requires check of balance (i.e. whether randomization was successful). If randomization is not successful, then the results are not valid
			Usually requires large sample sizes.
(b) Factorial designs	These should be considered as a variation to the basic RCT (see above where randomized assignment is a one time event.) In this design the comparison group gets the standard package that everyone is getting. The treatment arms get the standard package *plus* small changes or additions that in turn need to be tested and compared with the comparison group. So it requires at least four groups: the comparison arm, treatment group A, treatment group B and a treatment group A+B.	Different new interventions can be tested.	The maintained assumption is that there is little doubt about the efficacy and effectiveness of the 'standard package' This method is used to assess the effectiveness of additional and possibly 'innovative' methods. It requires careful planning and a large eligible sample.
		Cost effectiveness comparisons can be done.	
		Is easily done in contexts where interventions are new.	
(c) Pipeline or phased designs	The treatment or policy is implemented in phases where units are assigned to different phases *randomly*.	It overcomes the traditional resistance to random assignment that the comparison group is being denied the benefits of a policy or an intervention.	It requires that the units are assigned to different phases in a random manner. The phased design and the random assignment usually require a keen understanding and close engagement by the project team that is implementing the programme.

II. Quasi-experimental designs			
Regression discontinuity	A cut-off determines who is eligible to participate. Outcomes of beneficiaries and non-beneficiaries close to the cut-off line are compared.	All quasi-experimental designs require baseline data. Additional data on attributes is also required.	The credibility of the technique hinges strongly on establishing that the discontinuity is random.
			Requires data on many variables that are unrelated to the programme.
			Precision of impact is limited to those participants most close to the allocation cut-off.
Matching (including propensity score matching.)	Programme beneficiaries are compared to a group of non-beneficiaries. The match is constructed by finding people whose observable characteristics are similar to those of the people in the treatment group	Can be designed after the programme has been implemented and rolled out. However requires high quality baseline data that is relevant to the programme on many variables.	Requires a comparison group and credibility hinges strongly on establishing that there is a good match and balance between matched arms/groups.
		Can use previously existing datasets such as censuses, DHS, etc. *if* they are high quality to establish baseline and balance in level and trends.	Requires data on "matching variables" and knowledge of important covariates. The maintained assumption is that a match on 'observed' characteristics is also accounting for any differences in *un*observed characteristics.
Instrumental variables	Participation in a programme can be predicted by an incidental factor, or "instrumental" variable, that is uncorrelated with the outcome (other than by predicting participation)	Can be undertaken after the programme has been rolled out.	Requires strong assumption that the instrument affects the outcome only through one specific channel, which is usually hard to establish.
		The counterfactual is determined by the program	

15.4 What Do Impact Evaluation Studies Show?

In this section, we discuss the contributions that impact evaluations of forestry programmes have been able to make in assessing their effectiveness.

Measure change: Impact evaluations help to understand the direction and measure the magnitude of change of forestry programmes. Thus for example Somanathan et al. (2005) show that in India forest cover increased by 12–16 % as a result of community management. Similarly other studies have measured the effect caused by drivers of deforestation. Andersson et al. (2011) examine the effect of safety net transfers on forestry cover in Ethiopia and find that there is an increase in livestock but no effect on tree plantation and Alix-Garcia et al. (2013) examine the effect of cash transfers on forestry management in Mexico and find a positive effect. Bensch and Peters (2011) examine the effect of improved cookstoves in Senegal and find that charcoal use reduced by 25 % depending on the extent of use of these stoves. They infer the effect on deforestation.

In many cases, these are likely to have repercussions for policy. Pfaff et al. (2008) find that in Costa Rica, the programme for payment of ecosystem services (PSA) had little effect on the 1997–2000 forest clearing. They find that despite a large amount of resources devoted to this programme, PSA prevented deforestation in the first few years of contracting by only 0.21 % in the land that was enrolled. The main reason for the small change is that the overall national deforestation rate was reducing during this period and the incremental contribution made by this programme was minimal. In the same country, Andam et al. (2007) measure the effect of protected areas and conclude that between 1967 and 1997, protected areas helped to reduce deforestation by 10 %.

Deal with bias: In many cases non-impact evaluation studies are unable to disentangle the effect of inherent characteristics of areas/people receiving forestry programmes on one side and the effect caused by the programmes themselves (irrespective of context). Consequently they erroneously attribute all changes in forest cover to the forest programme or policy (e.g. programmes that manage forests differently or offer incentives to reduce deforestation). This is the programme placement bias problem. In other cases, people that select themselves into being beneficiaries of programmes have inherently different characteristics from people that don't. This is the selection bias problem. Both these need to be accounted for, if we want to measure results of a forestry programme in an unbiased way. So Cropper et al. (2001), and Nelson and Chomitz (2009), account for the fact that protected areas that aim to protect forests, are likely to be placed in areas that have low agricultural productivity and profitability. All increases in forest cover in these areas, compared to other areas, cannot be then rightly attributed to the effectiveness of protected areas. This is because many of these areas would have remained forested even in the absence of these policies.[2] In these cases, once programme placement

[2] See Cropper et al. (2001).

bias and selection bias is accounted for, the change in outcomes is usually quite small.

Similarly for PES (payment for ecosystem services) programmes, the effectiveness of these systems is likely to be jeopardized by the potential for adverse self-selection and poor administrative targeting which may result in low effectiveness of land being under forests (Alix-Garcia et al. 2013).

Assess effectiveness of targeting: Impact evaluations can also help understand the effectiveness of targeting, i.e. whether those that are most likely to clear forests are the ones that are being targeted by forestry programmes and policies. Therefore Alix-Garcia et al. (2013) examine whether the PES programme in Mexico was targeting the most vulnerable areas. They find that the country wide programme, with a budget of more than US$5 million, was quite successful in targeting eligible households. Assessments such as these also help to inform how programmes should be designed and some of the pitfalls to be aware of when designing PES programmes for example. In Costa Rica Pfaff et al. (2008) find that the PSA programme did not target locations that were most likely to change land use. Hence there were very small changes in forest cover *caused* by the PES programme.

Estimating impacts on sub-groups: Impact evaluations also can help to address questions of equity and heterogeneous impacts. Somanathan et al. (2005) show that after accounting for potential selection bias and placement bias, community managed forests in India performed better in raising crown cover by 12–16 % compared to unmanaged commons, but only for forests comprising of broadleaf trees but not pine trees. (Understanding the effects on sub-groups however requires that sample sizes are so selected to be representative for the sub-groups of interest.)

Compare different forestry programmes: Many studies examine programmes that engage communities and compare their effectiveness with the status quo such as government managed systems or unmanaged systems (Somanathan et al. 2005; Tachibana and Adhikari 2009; Scullion et al. 2011; Edmonds 2002).[3] In India Somanathan et al. (2005) show that in community managed forests performed no worse and perhaps better than state managed forests. Tachibana and Adhikari (2009) examine this question in Nepal and show that community co-managed forests are more successful in helping deforested areas recover forests than forests that are solely managed by communities. They also find that *co*-managed forests are able to especially protect forests better where large visible extraction is taking place (such as of timber). Similarly Cropper et al. (2001) find that after accounting for selection biases, protected areas as a whole are less effective in protecting forests than specially designated wildlife sanctuaries most likely because the latter have more resources devoted to them.

Unintended consequences and spill-overs: Impact evaluations can also help to measure unintended effects, spill-over effects (see for example Andam et al. (2007) and Arriagada et al. (2012)) as well as strength of these neighbourhood effects and

[3] Three of the four studies (one is ongoing) show that community based management systems did reduce forest clearing (see Table 15.2). The time periods over which these interventions are also examined are very short.

peer effects (Bravo-Ureta et al. 2011; Chibwana et al. 2013). Bravo-Ureta et al. (2011) show that plots *close* to households that received natural resource management training were as likely to benefit from resource management techniques as those much further away concluding that the evidence for peer effects in training programmes is very sparse.

Assess Trade-offs: In the context of forestry projects, it's clearly important to know whether a change in policy is reducing deforestation. Almost as importantly, it's also important to know whether the changes planned for and foreseen by the interventions and the policy changes would in fact have occurred *without* the intervention. Knowing whether and how *much* these effects were is important. Thus Sims (2008) shows that in Thailand, protected areas have prevented forest clearing that otherwise would have occurred: sub-districts with more land in protected areas have between 9 and 32 % more forest cover than typical sub-districts. These programmes have also reduced land available for agriculture. However, social costs have been minimal: consumption levels in sub-districts with more land in national parks was higher by 2 and 7 % respectively and poverty levels 4 and 12 % lower than for comparison sub-districts. On the other hand, inequality measures were higher on average for communities near national parks, indicating a disproportionate share of these gains went to higher income households.[4] In another case, Alix-Garcia et al. (2013) examine the effect ofcash transfers under the aegis of Opportunidades and find that forests were affected detrimentally as a consequence of a cash transfer programme. They use an RCT to measure the magnitude of this effect. This is important because programme managers and policy makers can make choices once they know the magnitudes of change. Similarly in the Ghana (Burwen and Levine 2012) and Senegal (Bensch and Peters (2011) studies of fuel wood use, the studies *measured* the impact on fuel wood and charcoal.

15.5 What Do Impact Evaluation Studies of Forest Adaptation Projects *Not* Show and Challenges

Although impact evaluation studies in forestry have been successful in helping to measure various policy and programmatically relevant topics, there are several areas that impact evaluations have fallen short of.

First, it is also important to see what kinds of forestry programmes these impact evaluations are *not* evaluating. Most forestry programmes are **complex, multi-intervention, multi-sectoral** programmes with livelihoods, health, agriculture and income poverty as their primary outcomes. These sorts of complex programmes that have multiple arms, have large possibilities for additional programmes being tagged on, are implemented by different agencies on the ground and are usually scaled up in a slow, organic manner, are not traditionally evaluated by impact evaluations.

[4] The most probable mechanism for the positive income effects of national parks is increased income from tourism.

Additionally, if programmes have not built in an impact evaluation plan into them from the start, these are usually not impact evaluated either. Last but not least, in all the programmes listed here, researchers use clever and innovative ways to construct counterfactuals or comparison groups. Although with randomized experiments, this is easier to do, programmes also use matching methods to 'construct' these counterfactuals. Other programmes and policies that do not traditionally get impact evaluated are: large national or regional policies, programmes that aim to change institutions, and programmes that are very small (and have no clear possible counterfactual).

Secondly, forestry projects are also frequently projects with high '**causal density**'. As defined by Woolcock (2013), this means that forestry programmes typically have different interventions that start at different times, and are implemented with different intensities. This means that it is not always clear what the 'treatment' group receives and what the comparison group receives. Many studies choose to deal with this by making the treatment a 'package' so that *variations* within the package are ignored. If a package for instance has two types of interventions, frequently an impact evaluation assumes that the treatment group has received both interventions equally (while the comparison group doesn't usually get any part of the package). In vector terminology, an impact evaluation this can be represented as testing (1,1) vs (0,0) where '1' represents the full intervention being implemented and '0' means the intervention is not implemented. The '( )' term represents a full 'package'. But frequently real-world implementation means that equal intensities and amounts across different treatment subjects is not possible. Furthermore,in many cases it is clear that the relevant and more useful impact evaluation question is to evaluate what would happen with different 'dosages' of the same intervention i.e. (1,1) vs. (0.5,1) vs. (1,0.5) vs. (0,0) vs. (0.5,0.5) for instance. (This route has many repercussions for design, roll-out of programmes and sampling and costs of evaluations which we don't discuss here.)

Another gap we found in impact evaluations was the lack of **implementation research**. Impact evaluations presume that the programmes they evaluate have high 'efficacy' i.e. if implemented correctly and completely, will work on the ground, as the programmes were envisioned in the laboratory.This is not borne out in the real world and we maintain that it will be very useful to additionallyexamine what is required to *implement* programmes better: an important question for programme managers and also for policy makers. None of the impact evaluation studies that we reviewed had an implementation research component. Understanding for example what possible methods of community based management programmes are most effective (in the relative sense), what makes forest officials more efficient, what types, frequencies and magnitudes of payments for ecosystem services makes them most effective in ensuring better current and future forest cover are all important questions that most impact evaluation studies shy away from answering.

Fourthly, there is the question of **assumed trends**. In many cases the timelines of programmes that are evaluatedare different and keep changing. They are also different at different points in time. Impact evaluations presume that past trends in comparison and treatment areas will be mirrored during the period of the impact evaluation. As Woolcock (2013) points out, this is incorrect.

The fifth point relates to **data and capacity**. Impact evaluation studies require a lot of data: they require high quality relevant baseline and endline quantitative and qualitative data not just for programme areas/beneficiaries but also for non-programme areas/beneficiaries. Additionally, they also require good and timely implementation data. This has also constrained the application of impact evaluation methodologies. Indeed in most cases unless impact evaluations are planned for, at the inception of the project, it is hard for these to be done at the end. We also found in our assessment that some impact evaluations used randomized control trials as their 'identification' strategy (or methods that allow them to measure the strength of the causal relationship). These included studies by Burwine and Levine (2012) and Hafashimana et al. (forthcoming). But most studies used quasi-experimental methods (Arriagada et al. 2012; Edmonds 2002; Sims 2008; Cropper et al. 2001; Andersson et al. 2011). The technical expertise required to specify, estimate, analyse and understand quasi-experimental methods is, arguably, greater than those for randomized control trials. Indeed this might also account for the small number of impact evaluations in the area. A recent systematic review (Samii et al. 2014) indeed found only 12 studies in developing countries that could reliably be measuring changes in outcomes of forestry programmes (after accounting for inclusion and exclusion criteria).

Last but not least, impact evaluations can be robust but they can also provide **contradictory results** for the effectiveness of programmes. Therefore in Thailand, Sims (2008) and Cropper et al. (2001) differ in their conclusions about the effectiveness of protected areas; in Costa Rica, Andam et al. (2007) and Pfaff et al. (2008) differ in their conclusions about the effectiveness of deforestation related policies. Although clearly not the objective, one of the consequences of using impact evaluation studies is that they usually require a registration of protocol, clear pre-analysis plans and in many cases now, replicability of results, the robustness of these studies and therefore their implications for policies and programmes can be assessed very easily. Systematic reviews (see for example Waddington et al. 2014) also usually incorporate meta-analysis of data (if it is possible) to see what the overall result from a collection of studies is likely to be.[5]

15.6 A Discussion of Some Ways Forward

Overall there are two types of programmes that can directly or indirectly increase forest cover. The first is through development projects where the development programmes affect the intensive or extensive frontier of forests indirectly – by affecting

[5] Although this requires that several fairly stringent conditions are fulfilled – for example the intervention needs to be the same, the outcome needs to be the same and the assumption that the different datasets are coming from the *same* underlying statistical population which has the same underlying distribution, can be a strong one.

the pressure on livelihoods where the substitution effect is hypothesized to be greater than the income effect (see for example Puri 2006).

The second type of programmes are those that aim to affect forests directly, usually through the way they are managed. Following Samii et al. (2014) these can be divided into the following categories: (i) community forestry management or joint forestry management programmes; (ii) protected area programmes; and (iii) payment for ecosystem services programmes.

We argue that this understanding of possible **causal pathways** is critical in any impact evaluation. Causal pathways (also called theories of change or impact pathways) help to identify nodes that are critical for realizing changes in outcomes or impact indicators but also help identify possible unintended consequences and spillover effects. Discussion of causal pathways can also allay some doubts that have been raised about the uses and limitations of impact evaluation methodologies (see Stern et al. 2012).

The second critical limitation affecting the use of impact evaluations is the extensive demands on data. We argue that the methodologies used in impact evaluations allow researchers to be creative in understanding many questions of bias, placement and measure effectiveness in a robust manner. For example, Nelson and Chomitz (2009) assess the impact of tropical protected areas on forest fires, which they argue is the best available global proxy for deforestation at a fine spatial scale. Using forest fires as a proxy for depleted forest cover, they conclude that in Latin America and the Caribbean, protected areas reduce forest loss – by 4.3 %. Cropper et al. (2001), Sims (2008), Andam et al. (2007), Arriagada et al. (2012), Scullion et al. (2011), Edmonds (2002), Somanathan et al (2005), Pfaff et al. (2008), Alix-Garcia et al. (2014), all use **GIS data in different ways and in combination with survey data** to understand attributable impact. In Table 15.2 below we discuss the various data sources that can be used in different ways to understand and measure attributable impact.

15.7 Conclusions

There are several good and other not so good reasons for why impact evaluations are not routinely undertaken in the forestry sector. Capacity and costs are the two most cited reasons for the limited number of impact evaluations in this area.[6] Both these problems can be dealt with. Clearly impact evaluations should not be done in all cases but in some cases their use is critical.

Woolcock (2013) in his excellent exposition on exploring various dimensions of external validity lays out three domains which are likely to influence the external validity of results. He categorizes these domains as, first, causal density ('the extent to which an intervention or its constituent elements are complex'), second, implementation capability (the extent to which any other organization can faithfully

[6] Personal conversations.

Table 15.2 Sources of big and small data for understanding attributable impact in forestry studies

No.	Type	Helps to inform	Other notes
1.	Satellite images	Forest cover, change in forest cover, density of forests, crops grown, land use, land cover.	Landsat high resolution images.
			Need interpretation and ground truthing.
2.	Aerial photographs	Forest cover, surveillance, cartography and drawing maps,	Can be combined easily with GPS data.
			Is unobtrusive and depending on resolution can be used easily in GIS.
3	Other remote sensing images	Forest fires, deforestation, land use, conversion, conservation, altitudes, elevations, topographic maps	Helps to collect data on inaccessible or dangerous objects.
			Replaces expensive on the ground data and ensures that areas are not disturbed in the bargain.
4.	Other maps	Soil, roads, access, use, boundaries of properties including villages and states and other administrative units, altitude, population density, climate maps, ethnicity, migration patterns etc. Maps showing indices such as disaggregated poverty have also become common.	Include topographic maps, soil maps. Usually require other methods for construction such as remote sensing but also careful census type data collection (e.g. for property rights and boundaries.)
5.	Social and household surveys	Livelihoods, behaviour reasons, behaviour patterns, eligibility for programmes, perceptions, socio economic indicators, physical/economic/social access, use, income, assets and impacts on welfare	Requires careful qualitative work beforehand and afterwards to interpret. It also requires a lot of training for data collectors, careful piloting of instruments to ensure questions are conveying what is being asked. Also requires careful data entry and algorithmic checks which can, if done on mobile phones or PDAs be done on the survey itself. May also require data cleaning. Cross-sectional and repeated time series or panel data have attributes that are specifically useful for understanding changes over time. It is important to understand attrition in this data and reasons for it.

6.	Administrative data	Eligibility, socio-economic processes, administrative processes, forest management processes, laws and changes in legislation.	Typically is broad brushstrokes data that can be used for sub-populations and illustrates main changes or attributes without explaining motivations or behaviour change of populations
7.	Individual, structured or semi-structured interviews	Help to explain perceptions, beliefs, customs, reasons for behaviour change, determinants of actions, social status, processes and exchange. Also help to explain unobservable selection bias. As well as explain participation and non-participation.	Usually required before, during and after most quantitative data collection so quantitative data can be anchored, collected well and interpreted clearly. Also required for questionnaire design and sampling.
8.	Case study	Qualitative understanding of local drivers and dynamics including processes.	Case studies are especially important to understand processes and behaviours and get deeper insights into what quantitative data may be telling us.
9.	GPS data	Can help determine locations for cities, markets, hospitals, schools as well as boundaries for properties and areas.	Can be combined into a GIS with other data and helps to combine aerial data with on the ground data. GPS data can provide spatial coordinates that may be combined with mosaics of satellite images to then make these useful with other types of data layers as well.
10.	Management information systems	This data can help capture delivery, implementation fidelity, extent to which project targets have been met	These usually are used with process data to understand implementation fidelity.

implement the type of intervention or programme), and finally, reasoned expectation (the extent to which claims of actual or potential are understood within the context of an evidence based theory of change, which also in turn specifies what can be achieved by when). Using this typology, it is clear that the external validity of impact evaluations is limited when one considers that forestry projects are usually complex and have many intervention arms.

One way to deal with these is to critically incorporate discussions of causal pathways into impact evaluations.

It is not easy to randomize or generate counterfactuals in this area. In our review we recognize that few impact evaluations have usedrandomized assignment. Impact evaluations that use quasi-experimental methods such as regression discontinuity, matching, switching regressions, propensity score matching techniques all require technical expertise that is much more academic than randomized control trials.

Another question that is important to answer is: are impact evaluations necessary and sufficient to answer the big questions? Not unlike in other fields, forestry programmes are complex, are implemented with multiple arms, at different times with different intensities. There are clearly many possible confounding factors that can affect the ultimate outcomes that programmes in this area seek to influence. With caution, we argue that impact evaluations are in fact necessary to answer some of the questions, especially if we want to measure effects robustly that take into account the various sources of bias and are still able to measure the magnitude of these effects.

Impact evaluations require a large amount of data, which is highly disaggregated and has layers that are able to deal with endogeneity and account for confounding effects. Spatially explicit data that use satellite imagery, aerial photographs and other GIS can clearly be combined with traditional sources of data such as household and individual surveys to make these possible. Clearly it is important to be creative here.

We recommend multidisciplinary teams that are able to measure the different variables around the implicit causal chain. We also recommend the use of different sources of data. High resolution data has become available and can be used with real time socio-economic data to understand and also account for socio-economic effects.

Impact evaluation should clearly not be undertaken in all cases. We recommend that **impact evaluations should be undertaken in three cases**.[7] First, there are programmes and policies that are innovative where there is no previous evidence that they even work. Much like pharmaceutical medical phased trials that typically undergo four phases of efficacy testing (see for example Meinert and Tonascia 1986) these are really efficacy trials that just need the proof of concept. So new media campaigns, awareness raising campaigns, programmes that introduce new technologies or new technical ways to save forests, should be tested in this type of impact evaluations. This is where implementation control is critical and these are

[7] This is not new. Several other agencies have adopted this nomenclature. Innovations for Poverty Action (IPA) also uses a similar nomenclature.

laboratory to field experiments. Second, there are programmes where the original programme design has been tweaked to fit the context or have been implemented in a slightly different way. These external replication programmes can and should be impact evaluated because not only does it tell us whether the tweak in the design has the same effectiveness as originally envisioned but also because, impact evaluations can help throw up good lessons for implementation. The third type of projects and policies that this should be done for are large scale programmes where accountability to stakeholders such as donors or the people funding the programme (ordinary people through their taxes) want to know just how much difference the programme is making (so Oportunidades is an example or the Mexico PES programme) and whether it's really important to know if there is value for money.

Acknowledgements The authors thank Raag Bhatia, research assistant, 3ie for excellent support.

Annex I: Definition of Important Terms for Impact Evaluations

Intervention	Is used here interchangeably with the programme or the policy implemented or planned to increase the resilience and reduce the vulnerability of forests.
Treatment group	The stakeholders that receive or are beneficiaries of the programme or the intervention. Can be individuals, households, plots of land, communities, villages, districts etc.
Comparison group	Is the group that is typically compared with the treatment group and (at least for some time) does not get the treatment.
Identification design	Are methods that can help *identify* and help to attribute changes in measured effects to a programme/policy/project. Usually these require that implicit or explicit counterfactuals (also called comparisons) to understand what would not have occurred had the programme not occurred.
Unit of assignment	Is the level or unit at which a programme is implemented.
Unit of measurement	Is the unit for which measurement is undertaken and the units for which the measurement of the effects is important.
Mixed methods	Is the collection of methods that are interdisciplinary and use qualitative and quantitative methods in an integrated manner, informing each other and supporting and assisting each other to provide and supplement each other to provide a more wholistic understanding and measure of the effects of a policy, programme or project.

Annex II: List of Studies, Locations, Interventions and Identification Methods Reviewed in this Chapter

No.	Location of study (author)	The main intervention	Intended outcome	Identification method and data used
1.	Ethiopia (Andersson et al. 2011)	A productive safety net (food for work)	Changes in livestock and tree holdings	Regression with propensity score matching.
2.	Mexico (Alix-Garcia et al. 2013)	Conditional cash transfer programme	Deforestation	Method: Regression discontinuity along with IV discontinuity.
3.	Ghana (Burwen and Levine 2012)	Distribution and use of improved cookstoves	Fuel use	A randomized trial.
4.	Thailand (Cropper et al. 2001)	Road building and protected areas	Deforestation in protected areas	Instrumental variables
5.	Senegal (Bensch and Peters 2011)	Improved cookstoves	Demand for charcoal	propensity score weighted regression approach.
6.	India (Somanathan et al. 2005)	Managed forests	Deforestation measured by crown cover.	Difference in difference
7.	Nepal (Tachibana and Adhikari 2009)	Community co-management of forests	Deforestation.	A switching regression model

8.	Tanzania (Scullion et al. 2011)	Management of forests	Governance, forest conditions and local livelihoods	Quasi experimental methods
9.	Nepal (Edmonds 2002)	Management of forests	Forest cover	Instrumental variables and regression discontinuity approach
10.	Costa Rica (Arriagada et al. 2012)	Protected areas	Avoided deforestation	Mahalanobis weighting with propensity score matching.
11.	Thailand (Sims 2008)	Wildlife Sanctuaries and National Parks	Forest clearing	Quasi-experimental matching techniques
12.	Developing countries (Nelson and Chomitz 2009)	Tropical protected areas.	Deforestation fires which proxy deforestation	Differences in differences with matching
13.	Uganda (Hafashimana et al. forthcoming)	Payment for ecosystem services		The study is on-going (randomized assignment)
14.	Mexico (Alix-Garcia et al. 2014)	Payment for ecosystem services	Forest cover and socio-economic outcomes.	Difference in difference with matching
15.	Costa Rica (Arriagada et al. 2012)	Payment for ecosystem services	Forest cover	Difference in difference with matching
16.	Costa Rica (Andam et al. 2007)	Payment for ecosystem services	Forest clearing	Difference in difference with matching

References

Alix-Garcia J, Mcintosh C, Sims KRE, Welch JR (2013) The ecological footprint of poverty alleviation: evidence from Mexico's oportunidades program. Rev Econ Stat 95:417–435

Alix-Garcia J, Aronson G, Radeloff V, Ramirez-Reyes C, Shapiro E, Sims K, Yañez-Pagans P (2014) Environmental and socioeconomic impacts of Mexico's payments for ecosystem services program

Andam KS, Ferraro PJ, Pfaff ASP, Sanchez-Azofeifa GA (2007) Protected areas and avoided deforestation: a statistical evaluation. Final report. Global Environment Facility Evaluation Office, Washington, DC

Andersson C, Mekonnen A, Stage J (2011) Impacts of the productive safety net program in Ethiopia on livestock and tree holdings of rural households. J Dev Econ, Elsevier 94(1):119–126

Arriagada RA, Ferraro PJ, Sills EO, Pattanayak SK, Cordero-Sancho S (2012) Do payments for environmental services affect forest cover? A farm-level evaluation from Costa Rica. Land Econ 88:382–399

Bensch G, Peters J (2011) Alleviating deforestation pressures? Impacts of improved stove dissemination on Charcoal consumption in urban Senegal. Land Econ 89(4):676–698

Bowler D, Buyung-Ali L, Healey JR, Jones JPG, Knight T, Pullin AS (2010) The evidence base of community forest management as a mechanism for supplying global environmental benefits and improving local welfare. Environmental Evidence: www.environmentalevidence.org/SR48.html

Bravo-Ureta BE, Almeida AN, Solís D, Inestroza A (2011) The economic impact of Marena's investments on sustainable agricultural systems in Honduras. J Agric Econ 62(2):429–448

Burwen J, Levine DI (2012) A rapid assessment randomized-controlled trial of improved cookstoves in rural Ghana. Energy Sustain Dev 16:328–338

Chibwana C, Jumbe CBL, Shively G (2013) Agricultural subsidies and forest clearing in Malawi. Environ Conserv 40(01):60–70

Cropper M, Puri J, Griffiths C (2001) How the location of roads and protected areas affects deforestation in North Thailand

Edmonds EV (2002) Government-initiated community resource management and local resource extraction from Nepal's forests. J Dev Econ 68:89–115

Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CM (2011) Impact evaluation in practice. World Bank, Washington, DC

Hafashimana D, Jayachandran S, Stanton C, de Laat J, Kalenscher T (forthcoming) Testing the effectiveness of payments for ecosystem services to enhance conservation in productive landscapes in Uganda; a prospective randomized evaluation. The International Initiative for Impact Evaluation (3ie)

List JA, Levitt SD (2005) What do laboratory experiments tell us about the real world. NBER working paper

Nelson A, Chomitz KM (2009) Protected area effectiveness in reducing tropical deforestation: a global analysis of the impact of protection status. Eval Brief 7:31

Pfaff A, Robalino JA, Sanchez-Azofeifa GA (2008) Payments for environmental services: empirical analysis for Costa Rica. Terry Sanford Institute of Public Policy, Duke University, Durham

Puri J (2006) Factors affecting agricultural expansion in forest reserves of Thailand: the role of population and roads. Ph.D. dissertation, University of Maryland

Puri J, Khosla A, Oldenbeuving M (2015) Grants for real-world impact evaluations: What are we learning? A statistical overview and a process analysis of 3ie Open Window Grants. 3ie working paper 23. International Initiative for Impact Evaluation (3ie) (forthcoming), New Delhi

Samii C, Lisiecki M, Kulkarni P, Paler L, Chavis L (2014) Effects of Decentralized Forest Management (DFM) on deforestation and poverty in low and middle income countries: a systematic review. Campbell Syst Rev 2014:10

Scullion J, Thomas CW, Vogt KA, Perez-Maqueo O, Logsdon MG (2011) Evaluating the environmental impact of payments for ecosystem services in Coatepec (Mexico) using remote sensing and on-site interviews. Environ Conserv 38:426–434

Sims K (2008) Evaluating the local socio-economic impacts of protected areas: a system level comparison group approach. Global environment facility impact evaluation information document, Washington, DC

Somanathan E, Prabhakar R, Mehta BS (2005) Does decentralization work? Forest conservation in the Himalayas. Planning unit discussion paper (05-04), Indian Statistical Institute, Planning Unit, New Delhi Discussion Papers, Indian Statistical Institute, New Delhi

Stern E, Stame N, Mayne J, Forss K, Davies R, Befani B (2012) Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development, working paper 38. DFID report

Tachibana T, Adhikari S (2009) Does community-based management improve natural resource condition? Evidence from the forests in Nepal. Land Econ 85:107–131

UNEP (2011) Forests in a green economy, a synthesis. United Nations Environment Programme

Waddington H, Snilstveit B, Hombrados J, Vojtkova M, Phillips D, Davies P, White H (2014) Farmer field schools for improving farming practices and farmer outcomes: a systematic review. Campbell Syst Rev 2014:6. doi:10.4073/Csr.2014.6

Woolcock M (2013) Using case studies to explore the external validity of 'Complex' development interventions, Wider working paper no. 2013/096. UNU/WIDER, Helsinki

World Bank (2015) World development report 2015: mind, society, and behavior. World Bank, Washington, DC. doi:10.1596/978-1-4648-0342-0. License: Creative Commons Attribution CC BY 3.0 IGO

Wunder S, Angelsen A, Belcher B (2014) Forests, livelihoods, and conservation: broadening the empirical base. World Dev 64:S1–S11