Evaluation of biodiversity policy instruments: What works and what doesn't?

Article *in* Oxford Review of Economic Policy · October 2012

DOI: 10.1093/oxrep/grs009

Daniela A. Miteva
The Ohio State University

Subhrendu K Pattanayak
Duke University

Oxford Review of Economic Policy, Volume 28, Number 1, 2012, pp. 69–92

Evaluation of biodiversity policy instruments: what works and what doesn't?

Daniela A. Miteva,* Subhrendu K. Pattanayak,** and Paul J. Ferraro***

Abstract We review and confirm the claim that credible evaluations of common conservation instruments continue to be rare. The limited set of rigorous studies suggests that protected areas cause modest reductions in deforestation; however, the evidence base for payments for ecosystem services, decentralization policies and other interventions is much weaker. Thus, we renew our urgent call for more evaluations from many more biodiversity-relevant locations. Specifically, we call for a programme of research—*Conservation Evaluation 2.0*—that seeks to measure how programme impacts vary by socio-political and bio-physical context, to track economic and environmental impacts jointly, to identify spatial spillover effects to untargeted areas, and to use theories of change to characterize causal mechanisms that can guide the collection of data and the interpretation of results. Only then can we usefully contribute to the debate over how to protect biodiversity in developing countries.

Key words: impact evaluation, payments for environmental services, devolution, community-based natural resource management, deforestation, poverty

JEL classification: Q2, Q23, Q28, Q56, Q57, Q58

I. Introduction

Biodiversity loss results from overharvesting, poaching, the destruction and degradation of habitats, and climate change (Slingenberg *et al.*, 2009; Barnosky *et al.*, 2011). Although much effort is channelled annually towards attempts to halt species loss, we still lack evidence on whether and under what conditions conservation measures can be effective (Ferraro and Pattanayak, 2006). Our paper reviews the most recent evidence on the performance of commonly used conservation measures and identifies gaps in the evaluation of these efforts. As most threatened species and habitats are found in tropical developing countries (Myers *et al.*, 2000; Hoffman *et al.*, 2010), we focus our analysis on the effectiveness of the biodiversity conservation instruments commonly employed there: protected areas (PAs), payments for ecosystem services (PES), and

*Duke University, e-mail: daniela.miteva@duke.edu
**Duke University, e-mail: subhrendu.pattanayak@duke.edu
***Andrew Young School of Policy Studies, Georgia State University, e-mail: pferraro@gsu.edu
doi:10.1093/oxrep/grs009

decentralization of natural resource management.[1,2] Of course, these three instruments are also used in developed countries and many of our conclusions apply equally there. We also briefly overview other common approaches, such as integrated conservation and development projects and forest certification schemes, and find very little empirical evidence on their effectiveness.

PAs are the most commonly used tool for biodiversity conservation in developing countries: about 15 per cent of their combined area falls under PAs (World Database on Protected Areas, 2011, available at www.wdpa.org). PAs place legal restrictions on human access and use within their boundaries and impose penalties on offenders. In contrast, PES schemes are more recent and seem to be concentrated predominantly in Latin America and China.[3] Unlike PAs, which use negative incentives to induce behavioural change, PES aim to promote biodiversity conservation and the provision of ecosystem services through positive incentives in the form of payments to landowners not to convert plots of land with high conservation value (Pattanayak *et al*., 2010). By redistributing management authority to local actors (e.g. municipalities, communities), decentralization measures create positive incentives for sustainable natural resource use as the benefits from the latter are shared among those who bear the costs of protection (Larson, 2002). In principle, local actors can be better monitors of natural resource regulations and hold local governments accountable to otherwise marginalized groups (Larson, 2002; Larson and Soto, 2008; Coleman and Fleischman, 2011).[4]

Although theory from economics and political science suggests that all three instruments can be effective conservation measures, in reality they often fail because of ineffective spatial targeting and dysfunctional institutions. For example, the theory motivating all three instruments assumes that they are applied to important habitats that are threatened with conversion to other land uses. However, PAs are often targeted on lands with the least political resistance to their establishment, and thus typically face the least

[1] In this paper we focus on ecosystem structure and function as a proxy for biodiversity. This focus is consistent with the working definition used by the Convention on Biological Diversity (CBD). Because the scope of the original definition of biodiversity was so broad, and because of the high correlation between the number of species, habitat quality and quantity, and other measures of biodiversity, the CBD has endorsed the ecosystems approach for the implementation and evaluation of conservation policies (CBD, see : http://www.cbd.int/ecosystem/; Slingenberg *et al*., 2009).

[2] Forest decentralization is not a single well-defined policy. The literature has pointed out the multiple connotations of the term (for example, Larson (2002) and Larson and Soto (2008) discuss in detail the multiple definitions; Coleman and Fleischman (2011) discuss the differences in the nature of the forest decentralization measures in Kenya, Uganda, Bolivia, and Mexico). Our emphasis in this paper is on quantifying the impact of changes of the management authority on terrestrial ecosystems. For this reason, we use 'decentralization' as a general term reflecting the redistribution of management authority from a higher to a lower level (communities or local governments).

[3] As of 2010 there are pilot PES programmes in Bolivia, Brazil, Colombia, Costa Rica, Ecuador, Mexico, Nicaragua, Venezuela, Kenya, South Africa, and China (Vincent, 2010). Jack *et al*. (2008) mention a PES-like scheme in Indonesia (Rewarding Upland Poor for Environmental Services (RUPES)).

[4] There is a very large literature (primarily comprised of case studies) that explores under what conditions decentralization can lead to the sustainable management of natural resources. See Larson and Soto (2008) for a review of recent studies. For theoretical arguments for the decentralized provision of local public goods, see Besley and Coate (2003). For research on the broader issues of the relationship between institutions and economic growth, refer to Acemoglu *et al*. (2004) and Besley and Persson (2011).

anthropogenic threat (Andam *et al.*, 2008). Voluntary in nature, PES contracts are also often established on the least profitable and, hence, least threatened lands (Pattanayak *et al.*, 2010; Ferraro *et al.*, 2012). Similarly, decentralization often happens in communities that already have a record of good ecosystem management (Bowler *et al.*, 2011). Furthermore, the theories that motivate the application of PAs, PES, and decentralization assume the existence and effectiveness of institutions and the rule of law (Hayes and Ostrom, 2005; Larson and Soto, 2008). Yet, developing countries are often plagued by uncertain property rights, widespread corruption, and the absence of strong institutions that can effectively coordinate across scales to disseminate information, reduce transaction costs, and monitor and enforce laws (Heltberg, 2001; Vincent, 2010).

The selection bias in the placement of interventions and the lack of effective institutions have cast doubt on the effectiveness of conservation measures and have spurred numerous calls for rigorous empirical evaluation of conservation policies (Kleiman *et al.*, 2000; Pullin and Knight, 2001; Salafsky *et al.*, 2002; Salafsky and Margoluis, 2003; Sutherland *et al.*, 2004; Saterson *et al.*, 2004; Sutherland, 2005; Frondel and Schmidt, 2005; Ferraro and Pattanayak, 2006; Carpenter *et al.*, 2009; Pattanayak *et al.*, 2010). These studies have highlighted the need for policy to be grounded on a firm understanding of *whether*, *under what conditions*, and *how* conservation instruments work. Translating such knowledge into policy can improve the performance, cost-effectiveness, and sustainability of conservation investments.

Emphasizing credible causal inference, we review the evidence for the effectiveness of the three conservation instruments and outline future directions for assessing their performance. In section II, we briefly describe quasi-experimental study designs that can isolate the causal impacts of conservation interventions and contrast these with common designs in the conservation literature. Then we summarize the current evidence for the effectiveness of PAs, PES, and decentralization, and give a brief overview of other common conservation policies. Section III identifies major trends in the existing evidence and highlights the main lessons for biodiversity conservation. Section IV calls for a new programme of research—*Conservation Evaluation 2.0*—that uses better theory, better methods, and better data to fill our knowledge gaps about what works and what does not in protecting biodiversity.

II. What has worked and what hasn't?

(i) Empirical designs and methods

Two common empirical designs employed by natural scientists to assess the performance of conservation measures rely on comparisons of outcomes (e.g. deforestation) in areas (*a*) with and without exposure to a conservation policy instrument, or (*b*) before and after a conservation policy instrument is implemented. 'With–without' analyses implicitly assume that (i) the areas with and without the conservation policy are similar in terms of their expected outcomes in the absence of the conservation policy (i.e. similar in characteristics that affect outcomes, such as accessibility, suitability for agriculture, and proximity to markets) and (ii) there are no spillover effects from the conservation policy to 'unexposed' areas. 'Before–after' analyses assume that the outcome level (or its trend) before a policy is enacted would remain constant after the policy is enacted (Nagendra, 2008) and that there is no selection bias in targeting the policy.

If these assumptions fail, the estimates of conservation policy effectiveness will be biased (Ferraro and Pattanayak, 2006; Ferraro, 2009; Joppa and Pfaff, 2010).[5] Consider the case of PAs in tropical forests. First, deforestation rates may change after the establishment of PAs for reasons other than protection (e.g. commodity prices), thus invalidating a simple before–after comparison of deforestation (Nagendra, 2008; Joppa and Pfaff, 2010). Second, PAs, like other conservation interventions, are not established randomly across the landscape. Instead, they tend to be established in poor locations that are far away from cities and unsuitable for agriculture or urbanization (Pfaff *et al.*, 2009; Joppa and Pfaff, 2009, 2010; Andam *et al.*, 2010). The unfavourable location implies that such lands are less profitable and may not experience deforestation, even in the absence of protection. In such cases, simple inside–outside analysis will yield upwardly biased estimates as the deforestation rates in PAs in the absence of protection are lower than the average deforestation rates of unprotected areas. Finally, the establishment of a protected area may displace the extractive activities to nearby buffer zones (Armsworth *et al.*, 2006; Joppa and Pfaff, 2010). In this case, the estimate of the impact of protection will also be biased upwards, but now because deforestation rates in the unprotected areas would have been lower in the absence of the policy.

Similar violations of key assumptions arise for PES and decentralization. A strong driver of enrolment in PES programmes is the lack of profitable alternative uses for the land (Pattanayak *et al.*, 2010). For example, large forested tracts owned by absentee landlords and with steeper slopes (i.e. low agricultural suitability) have a higher probability of enrolment in a PES programme in Costa Rica (Arriagada *et al.*, 2009, 2012). Likewise, some decentralization policies occur on forested land that is already in a good condition (Ferraro *et al.*, 2012), whereas others emerge because of the degradation of natural resources (Baland *et al.*, 2010). Because the emergence of effective local governance is often attributed to high levels of social capital within communities, such locations may also be better at the enforcement and monitoring of forests in the absence of decentralization (Baland *et al.*, 2010).

To assess the causal impact of a conservation policy, we must establish what would have happened in areas exposed to such a policy if they had not been exposed, i.e. establish the counterfactual outcome in the absence of a conservation policy. As noted above, estimating this counterfactual outcome is difficult because of the non-random assignment of conservation interventions. In econometric terms, non-random assignment induces a correlation between the policy variable (the treatment) and the error term in a regression equation, with the direction of the bias depending on the sign of the correlation between them (Greenstone and Gayer, 2009). Although both experimental and quasi-experimental designs from the programme evaluation literature can be used to isolate the causal impacts of a policy, only the latter have been used in the context of biodiversity conservation (Ferraro, 2009; Greenstone and Gayer, 2009;

[5] For example, comparing the results obtained through conventional and matching methods, Andam *et al.* (2008) found that conventional methods overestimate the effectiveness of protected areas in Costa Rica by more than 65 per cent.

Joppa and Pfaff, 2010).[6] For this reason, here we focus mostly on the commonly used quasi-experimental designs.

The three common quasi-experimental designs include matching, instrumental variables, and difference-in-difference (Pattanayak, 2009). Matching methods break the correlation between the treatment and the error term by matching units affected by the conservation policy (treated units) with observationally similar units that are not affected by the policy (controls) (Imbens and Wooldridge, 2009). Matching assumes that similarity in the observed characteristics translates into similarity in unobservable characteristics, correlated with the outcome and the conservation policy assignment, or that such unobservables are negligible sources of bias (Imbens and Wooldridge, 2009). In contrast, an instrumental variable design breaks the correlation between the error term and the treatment by exploiting a variable that is correlated with the policy assignment, but does not affect the outcome. In this case the causal effect is estimated by measuring how the outcome varies with the portion of the total variation in the treatment explained by variation in the instrumental variable. In other words, if PAs are more likely to be assigned where endemic mammals are present, but the presence of endemic mammals only affects deforestation rates through its effect on the likelihood of a parcel's protection, then the presence of endemic mammals can be used as an 'instrument' to identify a causal effect of PAs on deforestation. In practice, it is often hard to find instruments that are both strong (correlated with the intervention) and valid (uncorrelated with the outcome). Finally, the difference-in-difference (DID) designs measure the impact of a conservation policy by the difference in the before–after change in the outcomes for protected and unprotected areas. DID assumes that any unobserved differences (i.e. systematic biases) are linear and time-invariant and can hence be removed by taking the difference in the outcomes before and after the policy.

The quasi-experimental research designs can successfully address the inference problems associated with time trends and systematic differences between the treated and control observations.[7] Such designs can be used independently or in combination, as well as with other common econometric approaches, such as panel data estimators. They control for time trends in the outcomes because comparisons are made within the same time period. None of these designs, however, is immune to bias from spillovers (leakage) from treated to control units. If such spillovers (leakage) are likely, one needs to either select a control group that is unaffected by spillovers (leakage), or explicitly measure their effects.

(ii) Empirical evidence on conservation policy performance

Protected areas

Table 1 summarizes the studies that use rigorous empirical methods to quantify the impacts of protected areas. Empirical work using credible inference methods has

[6] We are aware of only one proposed study whose design employs a group randomized control trial in which the payments for forest protection are randomly assigned to some villages and not to others: UNEP, National Environment Management Authority (NEMA) Uganda, and International Institute for Environment and Development (IIED) (2010), 'Developing an Experimental Methodology for Testing the Effectiveness of Payments for Ecosystem Services to Enhance Conservation in Productive Landscapes in Uganda', proposal to the Global Environment Facility, Washington, DC, available at http://www.thegef.org/gef/node/2772

[7] See Ferraro and Pattanayak (2006) for a discussion of why quasi-experimental designs are still limited in the context of biodiversity conservation.

Table 1: Protected area studies using rigorous empirical analysis

Study	Location	Unit of analysis	Sample size (protected/ unprotected)	PA type	Methods	Outcome
Andam *et al*., 2008	Costa Rica	pixel	2,711/10,371 2,022/4,724	IUCN I–VI[a]	matching	11 per cent reduction in deforestation
Ferraro and Hanauer, 2011	Costa Rica	pixel	same as in Andam *et al*., 2008	IUCN I–VI[a]	matching	tradeoffs b/w deforestation and poverty reduction
Ferarro *et al*., 2011	Costa Rica	pixel	same as in Andam *et al*., 2008	IUCN I–VI[a]	matching, PLM	11 per cent deforestation reduction; tradeoffs b/w deforestation and poverty reduction
Ferraro *et al*., 2011	Thailand	pixel	same as in Sims, 2010	IUCN I–II[a]	matching	15 per cent deforestation reduction; tradeoffs b/w deforestation and poverty reduction
Joppa and Pfaff, 2010	147 countries[b]	pixel	5 per cent of PA in each country/4x unprotected area	IUCN I–VI	matching	deforestation reduction in over 75 per cent of the countries in the sample
Gaveau *et al*., 2009	Sumatra and Siberut	pixel	463/423	conservation and hydrological PAs	matching, regressions	24 per cent deforestation reduction
Pfaff *et al*., 2009	Costa Rica	pixel	4,229	IUCN I–II	matching, regressions	1–2 per cent deforestation reduction
Haruna, 2010	Panama	pixel	9,467/27,559 8,372/27,121	IUCN I–II	matching	12–16 per cent deforestation reduction 12–15 per cent deforestation reduction
Sims, 2010	Thailand	locality	20,565	IUCN I–II	IV	7–19 per cent deforestation reduction
Schwarze and Jurhbandt, 2010	Indonesia	pixel	10,418/13,888	Lore-Lindu National Park	matching	9.4 per cent deforestation reduction
Nelson and Chomitz, 2011	tropical developing countries	pixel	varies	All PAs[c]	matching, LOESS	some reduction in forest fires, impacts vary by intervention, time period, and distance to major city

Notes: [a] Indigenous reserves and wetlands were excluded; [b] all with >100 sq. km PAs; [c] the study aggregates PAs into strict PAs, multiuse, and unknown, based on the IUCN categories; [d] PLM is partial linear models; [e] LOESS is locally weighted scatterplot smoothing.

focused predominantly on the effectiveness of PAs in preventing deforestation, most often measured as a binary outcome at the pixel level.[8] The results seem to suggest that PAs are effective at stalling deforestation (e.g. Andam *et al.*, 2008; Gaveau *et al.*, 2009; Pfaff *et al.*, 2009; Sims, 2010; Joppa and Pfaff, 2010; Ferraro and Hanauer, 2011), had mostly negligible spillover effects (Andam *et al.*, 2008; Sims, 2010), and reduced the incidence of forest fires (Nelson and Chomitz, 2011). Nevertheless, the estimated effects are much smaller than conventional before–after and with–without methods would imply.

A few studies have suggested that the impacts of the PAs are heterogeneous and vary through time and in space according to the baseline characteristics of the area. For example, Ferraro *et al.* (2011) find that in Costa Rica the impact is greatest on land that has lower slopes, poor population, and is closer to major cities. They also find that in Thailand the impact of PAs on preventing deforestation is highest on land with lower slopes, but far away from major cities. Andam *et al.* (2008) find larger impacts of older PAs compared to newer PAs. Pfaff *et al.* (2011) compare the impacts across federal and state-managed parks and find that the intervention is more successful under the former. Nelson and Chomitz (2011) find that PAs have a positive impact on reducing forest fires, with the magnitude of the effect varying by geographic location, type of PAs (strictly protected versus multiuse), and the proximity to cities.

Decentralization measures

Table 2 summarizes the studies that aim to quantify the causal impact of decentralization measures on environmental outcomes. These studies find that the placement of decentralization interventions is associated with factors that also affect the measured outcomes, thereby invalidating simple comparisons between decentralized and non-decentralized resources. For example, Somanathan *et al.* (2009) observe that state-controlled forest plots had more forest cover at the baseline, were located on north-facing slopes away from roads and villages, and had large nearby forest stocks and low population density. Baland *et al.* (2010) find that community forests were located closer to the villages, while Edmonds (2002) finds that the villages with decentralized forests had higher levels of electricity and piped water access, were close to a local market and forestry offices, and received more agricultural assistance.

In contrast to the PA studies that use deforestation or fire as an outcome, almost all of the decentralization studies use measures of forest degradation (proxied by the amount of fuelwood collected, density of the canopy cover, forest regeneration, and lopping). Overall, they find limited evidence that forest management decentralization policies had a positive impact on forest degradation. Somanathan *et al.* (2009) find a statistically significant impact of decentralization on pine tree forests, but no impact on broad-leaved forests which are more heavily used by households and more likely to be degraded (Baland *et al.*, 2010). In another part of India, Baland *et al.* (2010) found a positive impact on lopping, but not on the tree cover, age of the trees (proxied by the tree diameter at breast height (DBH)), or the presence of saplings. Coleman and Fleischman (2011) find that, on average, the African forests in their sample (parts of Uganda and Kenya) experienced a negative, albeit insignificant, impact from decentralization, while

[8] Gaveau *et al.* (2009), Sims (2010) and Honey-Roses *et al.* (2011) employ a continuous outcome variable (per cent deforestation).

Table 2: Decentralization studies using rigorous empirical analysis

Study	Location	Unit of analysis	Intervention	Sample	Method	Outcome
Burgess *et al*., 2011	Indonesia	Pixel	Number of political jurisdictions	Large number of pixels	Poisson model	decentralization increased deforestation
Coleman and Fleischmann, 2011	Bolivia	Forest user group	National versus municipal institutions	11 treatment, 42 control groups	Probit, matching	(+) forest investments (+) perceived forest quality (not statistically significant)
Andersson and Gibson, 2007	Bolivia	Municipality	National versus municipal institutions	30 observations, 2 period GIS data	IV	No effect of municipal institutions on total or permitted deforestation; (+) impact on illegal deforestation
Pfaff *et al*., 2011	Brazil	Pixel	State versus federal management	40,321 pixels	Matching	Federal PAs reduced deforestation, impact varies by type of PA
Edmonds, 2002	Nepal	Household	State versus local	1,200 households	Matching, IV, RD	14 per cent reduction in fuelwood collection
Baland *et al*., 2010	India	Forest transect	State versus local	83 villages, 399 forest transects	OLS and Clogit w/ village FE,	20–30 per cent reduction lopping no impact on DBH, canopy cover, #saplings or fuelwood collection time
Somanathan *et al*., 2009	India	Pixel	State versus council managed forests	355 treatment, 582 controls for broad-leaved pixels; 318 treatment, 504 controls for pine trees	Regressions, Matching	Forest degradation (per cent crown cover): (+) impact for pine tree forests, no impact for broad-leaved forests, reduced cost of conservation
Heltberg, 2001	India	Village, household	Local institutions	180 households, 37 villages	IV	no impact on degradation (household firewood dependence, state of the forest)
Bandyopadhya and Shyamsundar, 2004	India	Household	Community management	8,307 households in 524 villages	Matching	Fuelwood consumption increase in villages with community management (some concerns with the model, though)
Coleman and Fleischman, 2011	Kenya	Forest user group	Community management	14 treatment, 57 control groups	Probit, matching	(–) forest investments (–) perceived forest quality (not statistically significant)
Coleman and Fleischman, 2011	Mexico	Forest user group	National versus community management	19 treatment, 21 control groups	Probit, matching	(+) forest investments (+) perceived forest quality
Coleman and Fleischman, 2011	Uganda	Forest user group	National versus community management	42 treatment, 102 control groups	Probit, matching	(+) forest investment[a] (–) perceived forest quality (not statistically significant)

Note: local = village or community level. [a] Planting trees, seeds, bushes.

the forests in the Latin American countries (parts of Bolivia and Mexico) were positively affected (only for Mexico were the results statistically significant, however). Using multi-period panel data on the district level in Indonesia, Burgess *et al*. (2011) find that increasing the number of jurisdictions spurs deforestation and that the impact increases immediately before local elections.

Based on the studies summarized in Table 2, the impact of decentralization policies in terms of reducing forest degradation and deforestation seems context-specific; it varies in terms of the scope, the benefits, and the rights transferred to local populations. This is not surprising given that the ambiguous definition of 'decentralization': the studies suggest that it can refer to policies increasing the decision-making authority of lower-level bureaucrats or to increasing the local-level authority of local users (Larson, 2002; Larson and Soto, 2008). In certain cases, it can be associated with the transfers of capital to local users or the establishment of property rights (Coleman and Fleischman, 2011). Thus, we need to understand clearly the mechanisms through which decentralization affects environmental outcomes before we can quantify impacts. For example, Coleman and Fleischman (2011) propose accountability and empowerment as predictors of whether forest decentralization policies can improve forest quality and the welfare of local users.

Payments for ecosystem services

Table 3 summarizes the current causal evidence on the performance of PES schemes. The studies tend to find reduced deforestation and increased reforestation taking place as a result of the participation in the PES schemes. None of the studies considers the impact on forest quality, which may have improved because of better management (Pattanayak *et al*., 2010). All of the causal evidence comes from Latin American countries that have significantly more land under private ownership compared with the rest of the world (Vincent, 2010).

The effectiveness of the PES schemes depends on the programme design (e.g. where, to whom, and by whom the payments are made), the degree of compliance and spatial spillovers (leakage) (Pattanayak *et al*., 2010). Previous studies have pointed out that the small impacts may be due to the poor initial targeting of PES schemes (especially in Costa Rica) because of the inadequate attention to the costs and benefits of the programme (Pfaff *et al*., 2008; Arriagada *et al*., 2012). Because participation in these schemes is voluntary, PES programmes are likely to suffer from moral-hazard and adverse-selection problems (Ferraro, 2008; Pattanayak *et al*., 2010; Ferraro *et al*., 2012).

Other conservation initiatives

Integrated conservation and development projects (ICDP) are widespread project-based interventions that aim to tackle directly the links between natural resource dependence, conservation, and poverty (Blom *et al*., 2010). Forest certification schemes provide financial stimuli for firms and farmers to adhere to defined environmental standards (Blackman and Rivera, 2010). Despite the long history and popularity of ICDP and forest certification schemes, we omit an extensive discussion of them because the number of rigorous impact studies is very small, with the evidence suggesting no impact from the interventions. For example, the only two studies that use rigorous empirical methods find no evidence that ICDPs shifted households away from agriculture toward

Table 3: PES studies using rigorous empirical analysis

Study	Location	Unit of analysis	Sample	Methods	Outcome
Rios and Pagiola, forthcoming	Colombia	farm plots	72 PES contracts, 29 controls	Tobit/OLS	3.6 ecosystem services pts
Alix-Garcia *et al.*, forthcoming	Mexico	farm plots	352 PSAH contracts, 462 controls	Matching and tobit	50 per cent deforestation reduction[b]
Scullion et al., 2011	Mexico	farm plots	38 PES contracts, unspecified # controls	DID	34.8 per cent deforestation reduction (pine-oak forest)18.3 per cent deforestation reduction (cloud forests)
Honey-Roses *et al.*, 2011	Mexico	polygon[a]	425 treatment, 3,778 controls	Matching, DID	3–16 per cent deforestation reduction in high quality habitat 0–2.5 per cent deforestation reduction in lower quality habitat
Sierra and Russman, 2006	Costa Rica	farm plots	30 PES contracts, 30 controls	OLS	0.4 ha fallow –0.25 ha forests
Arriagada *et al.*, 2008	Costa Rica	census tracts	1,050 PSA tracts, 7,138 controls	PSM and regressions	21.2–34.1 ha forest gain
Pfaff *et al.*, 2008	Costa Rica	pixel	40 PSA pixels, 40–240 controls	PSM	<1 per cent deforestation reduction
Arriagada *et al.*, 2012	Costa Rica	farms	50 treated PSA farms, 152 control	Matching and DID regression	gain of 11–17 per cent of the mean contracted forest area
Robalino *et al.*, 2008	Costa Rica	pixel	925 PSA pixels, 925–4,625 controls	PSM	0.4 per cent deforestation reduction

Notes: [a] In contrast to pixels, these are of irregular shape and size. They result from the unique combinations of geospatial layer attributes and may therefore not coincide with farm plot boundaries.
[b] However, total avoided deforestation benefits are modest because the clearing rates without the programme were very low.

sustainable forest use in the Brazilian Amazon (Weber *et al.*, 2011; Bauch *et al.*, 2012). De Lima *et al.*, 2008 find small impacts from forest certification.

We have also omitted from the current discussion: (i) conservation policies that are commonly used to target individual species, such as the US Endangered Species Act, Individual Transferrable Quotas (ITQs), and measures associated with the Convention on International Trade of Endangered Species of the Wild Flora and Fauna (CITES); and (ii) common conservation policies in developed countries (e.g. PAs, easements). Furthermore, our discussion focuses on policies targeting the *symptoms* of unsustainable natural resource management (e.g. deforestation and forest degradation); we exclude from our analysis policies that might impact the underlying causes of biodiversity loss (e.g. international trade, macroeconomic policies, increasing demand for timber, food, ranching) as the causal chain is likely to be longer and more complicated. We discuss the implications for leakage and spillovers from conservation and related policies in section IV. Finally, although our emphasis in this article is on environmental outcomes, we note that there is an even greater paucity of evidence on the socioeconomic effects of conservation policies (e.g. Andam *et al*., 2010; Weber *et al*., 2011).

III. What have we learnt so far?

Protected areas seem to be effective

PAs seem to reduce deforestation consistently. A comparison of the effects of PAs and other interventions is possible for only one country. In Costa Rica, the effects of the PA system seem to be larger than the effects of the PES scheme (cost-effectiveness, however, is unknown). The larger effects may arise from multiple factors. For example, the PAs may have been established during periods of higher deforestation or may have existed for a longer period of time. Nothing is known about the effect of PAs on forest degradation except through the use of fire as its proxy. The evidence base on PES and decentralization in terms of deforestation and forest degradation is smaller and less consistent in its findings. Additionally, studies from the impact evaluation literature have noted the possibility of a large publication bias, with the majority of published articles skewed towards finding the expected statistically significant effect (Duflo, 2004; Greenstone and Gayer, 2009; Ravallion, 2009).

Spillovers from conservation policies tend to be negligible

As noted above, conservation policies may result in changing the patterns of activities outside the targeted areas. Few studies have attempted to control for or measure spillovers. Some studies have tried to control for local spillovers by excluding from the control group areas that fall within a certain radius of the treated observations (Andam *et al*., 2008; Ferraro *et al*., 2011; Pfaff *et al*., 2011; Miteva *et al*., 2012*a*,*b*). Others have attempted to quantify the spillover effects directly by matching the unprotected areas near a protected area to areas unlikely to have been impacted by protection (Andam *et al*., 2008; Gaveau *et al*., 2009). However, only one study has explicitly tested for the presence of spillovers in the context of PES: Alix-Garcia *et al*. (forthcoming) find significant negative spillovers for the poorest quartile of their sample and significant positive spillover effects for the wealthiest quartile. Overall, these studies find small positive or no statistically significant spillovers, possibly because conservation impacts themselves are too small to generate spillovers.

Evidence limited to very few locations

Not only are studies with a credible empirical design rare, but the existing ones are not representative of the biodiversity 'hotspots'. For example, the majority of studies on PAs and PES focus on Costa Rica, which has been an exceptional country in terms of development and biodiversity conservation; very little evidence comes from other biodiversity-rich developing countries. Miteva *et al*. (2012*a*,*b*) provide evidence on the impacts of Indonesian PAs on deforestation, poverty, forest fires, species loss, and water quality. There are two global PA impact studies (Joppa and Pfaff (2010) and Nelson and Chomitz (2011). However, they use few controls for confounding covariates (necessarily because of the global scale of analysis), focus on a limited time period (2000–8), consider a limited fraction of the country (only 5 per cent of the protected area in each country as in Joppa and Pfaff (2011), or use approximations of the PAs where the exact borders are missing (Nelson and Chomitz, 2011). In contrast, the decentralization studies in Table 2 consider policies in East Asia (three countries), Latin America (three countries), and East Africa (two countries). However, almost all of these studies employ data collected from relatively small geographic areas and thus raise concerns about external validity.

No evidence on protecting ecosystem structure and function

The studies presented in Tables 1–3 consider the impact of biodiversity conservation policies on deforestation and forest degradation, which are *assumed* to be good proxies for species richness and ecosystem function. Moreover, the current literature does not consider where the conservation gains take place (with a few exceptions of studies looking at the heterogeneity of impacts according to the baseline characteristics of the area), what the resulting landscape configuration is (e.g. in terms of fragmentation and isolation of the forest patches), and whether the gains meet the threshold for the provision of certain ecosystem services (such as improving water quality).[9] In other words, the degree to which deforestation and forest degradation can proxy for the ecosystem structure and function determines how useful these data are for telling us about the effectiveness of common biodiversity policy instruments (Jack *et al*., 2008).

Impacts of conservation policies are heterogeneous

Baseline: As Tables 1–3 suggest, the research focus has shifted away from quantifying the average impact of a policy to analysing the heterogeneity of policy performance as a function of the bio-physical and socio-economic characteristics of the targeted areas. Nevertheless, such studies are few and most (except for Ferraro *et al*. (2011) and Nelson and Chomitz (2011)) compare the point estimates for the impacts within discrete groups of the data.[10] The two studies that examine impact heterogeneity as a continuous function of the slope, distance to major cities, and poverty, find significantly non-linear impacts, with the PAs being most effective in areas with low baseline poverty, low slope (Ferraro *et al*., 2011), and closer to large cities (Ferraro *et al*., 2011; Nelson and Chomitz, 2011). In these studies, PAs had a negative impact on deforestation when the baseline poverty was high (Ferraro *et al*., 2011, in Costa Rica) or at intermediate distances to major cities (Ferraro *et al*., 2011, in Thailand). The only study considering impact heterogeneity in a PES scheme finds that the programme is more environmentally effective when baseline poverty levels are low (Alix-Garcia *et al*., 2010).

[9] Recently, Sims (2011) has returned to her data set of Thai PAs to examine if the PAs influence habitat fragmentation. She finds that PAs did prevent significant fragmentation overall, increasing average forest patch size by 20–33 per cent and forest patch density by 2–4 per cent. The more strictly protected wildlife sanctuaries appear to have encouraged consolidation of cleared patches and prevented forest fragmentation even in interior areas, consistent with core-focused enforcement patterns.

[10] Most of these studies do not allow us to assess whether there are statistically significant differences between the sub-groups. A notable exception is the study by Ferraro and Hanauer (2011), who use heteroskedasticity-robust variance adjustments to compute the confidence intervals.

Type: In the case of PAs, Pfaff *et al*. (2011) find that federal parks are more successful at reducing deforestation in Brazil compared to state parks. Distinguishing between strictly protected and multi-use parks, Nelson and Chomitz (2011) find that the latter tend to result in reduced forest-fire incidence in Latin America and Asia. In the PES literature, case studies and descriptive approaches have suggested that the impact is likely to vary according to whether government or users provide funds (Engel *et al*., 2008). However, the hypothesis that the effectiveness of PES schemes depends on the funding source has not been evaluated using rigorous quantitative approaches (Pattanayak *et al*., 2010). No clear patterns emerge for the different types of decentralization in Table 3.

Duration: Usually, conservation policies need time to effect changes (e.g. Baland *et al*., 2010; Jack *et al*., 2008). Some studies have circumvented this by focusing on the older policies. For example, Andam *et al*. (2008), Ferraro and Hanauer (2011), and Ferraro *et al*. (2011) consider separately the impacts of the PAs established before 1979 and after 1981; their results suggest that older PAs prevented more deforestation. In contrast, Nelson and Chomitz (2011) find a consistently larger impact of PAs on preventing forest fires when they restrict their treatment group from all PAs protected before 2000 to only the ones established between 1990 and 2000. Somanathan *et al*. (2009) focus on forests that have been decentralized for at least 15 years. By discretizing the age of decentralized forest plots into older and newer groups, Baland *et al*. (2010) find that the impact of community-managed forests on lopping increases over time. None of the studies on PES has examined the heterogeneity of impacts through time. To our knowledge, no study has quantified how conservation effectiveness changes as a continuous function of time since protection.[11]

IV. Towards Conservation Evaluation 2.0

In this section, we draw on the literature in development and environmental economics to highlight what the next generation of conservation impact studies should look like: we call this *Conservation Evaluation 2.0*.

(i) Better theory

One of the major drawbacks of the current literature is that the empirical work is disconnected from theories that describe how the interventions affect outcomes. The studies summarized in Tables 1–3 have been key to understanding *whether* and *where* the conservation instruments cause socio-economic or environmental impacts. These reduced-form estimates are essential steps to understanding *why* or *how* the conservation policy works (Ravallion, 2007; Ferraro *et al*., 2012). However, concerns remain about internal and external validity, scalability, and impacts over the long term and across sub-populations (Pattanayak, 2009; Ravallion, 2009; Deaton, 2010*a*; Heckman, 2010). By discussing each of these concerns, we argue that empirical evaluations would greatly benefit from better

[11] In order to do this, researchers either have to assume selection does not change over time or they need to have panel data that allow selection over time. In other words, constructing the counterfactual is a substantial challenge in this case.

links to theory, not simply because of how models are specified and samples are selected, but also because of how impacts (or the lack thereof) are interpreted.

Internal validity: Explicit models of how the intervention effects change and how the hypothetical counterfactuals are realized help programme evaluations (Deaton, 2010a,b; Heckman, 2010). The lack of mechanisms behind conservation policies, coupled with a limited understanding of the contexts in which they operate, raises concerns that the programme may still be related to the post-programme outcome for reasons other than programme (Ferraro, 2009). Without theory to guide the model specification of quasi-experimental designs, we have to worry that omitted variables, inadequate controls for pre-treatment trends, and misspecification in the treatment selection model may mask the impact of the conservation intervention (Greenstone and Gayer, 2009).

Causal models with explicit assumptions (representing 'theories of change') can remedy this situation by indentifying key variables and appropriate samples to define the counterfactual (Pattanayak *et al*., 2010). For example, theory suggests that local institutions and social capital are hard-to-measure often-omitted variables that are likely to bias programme evaluations (Deaton, 2010*a*). To control for such variables in an evaluation of decentralization performance, Baland *et al*. (2010) use a village-level block design and sample decentralized and non-decentralized forests within a village. Miteva *et al*. (2012*b*) use political economy theory to inform which political variables should be considered (in addition to bio-physical and socio-demographic variables) to evaluate PA impacts in Indonesia.

For sample selection, better theory of change helps identify the scale of the impacts from the conservation policy: valid identification necessitates that the treatment and control be independent. For this reason, observations from the control group that are likely to have been affected by spillover effects should be excluded from the control group. Yet, without a theoretical model of *whom* and *how* the conservation policy impacts, it is hard to know where exactly the spillover effects occur. Current attempts to deal with spillovers take an exploratory strategy by considering buffers of different width from the PA boundary (e.g. <2km, 2–4km, 4–6km, and 6–8km as in Andam *et al*., 2008). Yet, there is no theoretical justification for these particular buffers.

External validity: Under what conditions can we generalize the results to other contexts, given that evaluations of conservation instruments often work with non-representative samples? Theory-based mechanisms, along with the appropriate structural parameters, of *how* and *why* the conservation intervention works (or doesn't) are necessary for out-of-sample predictions to forecast the impacts of conservation policies in new contexts (Heckman, 2010; Deaton, 2010*a*,*b*). The existing literature has taken an inductive strategy towards discovering the contexts that matter, such as biophysical (slope, soil quality) or socio-economic (poverty, market access). Instead, theory could be a better (deductive) guide for identifying the constraints that bind and the contexts that matter, and for generating testable hypotheses. Theory could also help identify key structural parameters needed to forecast impacts in other contexts (Timmins and Schlenker, 2009).

Additionally, theory could help us think more generally about economies or diseconomies of scale. For example, a large-scale conservation policy can cause general equilibrium effects (e.g. impacts on crop prices that in turn affect enrolment in PES schemes), especially if there is high dependence on natural resources in closely coupled human-natural systems (see below). Ross *et al*. (2010) use a dynamic computable general equilibrium (CGE) model to examine the environmental and economic impacts of PES in Costa Rica and find small general equilibrium effects.

Coupled systems: People and their environment are part of dynamic coupled systems: the ecosystem structure and function impacts communities and people, whose use of natural resources in turn impacts ecosystem structure and function (Dasgupta and Mäler, 2003). Perverse links persist and externalities abound because market and non-market signals (e.g. state and community institutions) often fail to emerge. By restricting natural resource extraction, conservation policy instruments will trigger a new dynamic in these coupled systems. Therefore, the coupling should influence how we model causal effects and what data we collect. First, it implies that we should consider joint economic and environmental outcomes. Second, it also suggests that we should collect data on and model the influence of initial conditions (e.g. socio-political and biophysical factors). However, few rigorous evaluations consider the joint outcomes of conservation programmes and model them as a non-linear function of initial conditions (Ferraro *et al*., 2011; Ferraro and Hanauer, 2011). Alternatively, analysts can evaluate programmes by conducting theory-based simulations. For example, Pattanayak *et al*. (2009) apply a dynamic CGE model to examine PA impacts in Brazil. They explicitly model how PAs reduce land available for agriculture and increase labour supply (because of lower levels of mosquito-borne diseases caused by deforestation). These land and labour-market effects in turn impact deforestation.

In developing countries, the people–environment coupling is strong (Barrett *et al*., 2011). Environment–poverty trap theories suggest that small initial differences in the local context (e.g. either prior to or resulting from a conservation intervention) can cause large divergences in wellbeing and ecosystem functioning over time (Dasgupta and Mäler, 2003). Traps emerge partly because persistent poverty and rising disparities in each period make it difficult to generate (*a*) critical levels of investment for growth and (*b*) conditions for good institutions to evolve and succeed (Dasgupta, 2009). Given these complex and multiple causes, the long-term impacts of a conservation programme can be very different from the short-term impacts.[12] Thus, where possible, we should collect data during and beyond the programme/project cycle. If long-run evaluations are impractical, Carvalho and White (2004) suggest using theory to describe a step-by-step sequence of causes and effects, collecting data on the initial steps and then examining how well each step is borne out during the project cycle. For example, researchers could check if social capital and local monitoring improve during the course of decentralization to signal the likelihood of long-run success.

(ii) Better methods

Impact evaluation is a rapidly evolving field, which is reflected by the advances in the methodology employed to establish a causal impact of the conservation measures. Earlier papers on the performance of conservation interventions relied on simple

[12] From a cost-effectiveness perspective, programmes that ignore long-run impacts may be highly cost-ineffective. This is because many of the adverse long-run impacts could be irreversible or sticky (e.g. even if outcomes are somewhat reversible, the coupled system displays hysteresis).

comparisons of the average difference in the outcomes between the matched treated and control observations. In contrast, recent papers illustrate the importance of combining matching methods with bias-adjustment techniques and adjusted variance estimators (Abadie and Imbens, 2006, 2011; Imbens and Wooldridge, 2009). These adjustments help achieve consistency, improve the robustness and the asymptotic efficiency of the estimator. However, only a few studies use (and report results from) better methods described in this sub-section.

Ruling out alternative explanations: Because of the observational nature of the data used in quasi-experimental evaluations, there remain concerns that an important variable has been omitted. Because conditional independence assumption which justifies the validity of the matching estimators is untestable, Rosenbaum's bounds approach is one common method for dealing with omitted variables (DiPrete and Gangl, 2004). This approach accounts for the sampling variability in allowing researchers to assess what the magnitude of the unmeasured covariate (captured by a parameter Γ) would be to alter the conclusions of the analysis (Rosenbaum, 2002). Small values of Γ suggest results more sensitive to the presence of hidden bias. Another way is through thought experiments. For example, Sims (2010) worries that differential migration in regions with and without PAs could be causing changes in deforestation. She shows that migration patterns did not, in fact, change during her study period.

Spillovers (leakage): Validity of the impact estimators also rests on the independence of the treatment and control observations. However, as previous studies have suggested, spillovers from 'treatment' to 'control' areas may violate the assumption. Although the literature has suggested excluding the potentially contaminated observations from the control group and explicitly testing for the presence of spillover effects at various distances, spillover analysis is not the norm. Additionally, as suggested in the discussion of feedback, these spillovers may be interesting phenomena that deserve direct modelling, instead of being treated as a nuisance to be dealt with.

Continuous, not discrete: Most impact evaluation studies have employed *discrete* treatments and covariates to examine how the impacts of protection vary across time, space, and intervention. However, many conservation instrument data such as duration, area covered, and amount, and the covariates that influence their impacts (e.g. slope, poverty rates, distance to markets) are all continuous variables.[13] Future evaluations could shift from answering whether the intervention has an impact, to examining the overall shape of the production function—that is, the shape of the relationship between the impact and the *continuous* treatment. Currently, only two studies have looked at the impacts of a conservation intervention as continuous functions of exposure: Sims (2010) considers the percentage of the locality that is a PA, whereas Arriagada (2008) uses a generalized propensity score method to examine how the density of PES contracts in a region affects deforestation. A similar suggestion applies to baseline covariates that modify the impact of a conservation instrument: the modification may be continuous (as considered by Ferraro *et al*. (2011) and Nelson and Chomitz (2011)), and not discrete.

[13] The current practice employs some *ad hoc* rules and subjective decisions as to what constitutes a treated unit. For example, if only a part of a unit falls within a PA, then it is up to the researchers to decide whether to consider it protected.

(iii) Better data

Previously we have noted the challenges for impact evaluation stemming from data shortage because (*a*) most conservation interventions in poor countries are framed as independent proofs of concepts, (*b*) there is poor infrastructure, training, and history of systematic data collection; and (*c*) of the challenges of combining ecological, socio-economic, and institutional data (Ferraro and Pattanayak, 2006). Here we highlight two specific concerns.

Baselines: The availability of multi-period geospatial data with relatively fine resolution has allowed for deforestation patterns to be examined through time. Unfortunately, the context matters in many ways and we have no such repository of social–political data. Usually, we have lacked good baseline data on formal and informal institutions, the degree of information asymmetries, market access, intrinsic incentives and norms, and previous participation in forestry programmes (Jack *et al*., 2008; Pattanayak *et al*., 2010; Arriagada *et al*., 2012, Ferraro *et al*., 2012). Clearly, we need more and better socio-economic and institutional data from biodiversity-relevant locations. Alternatively, we should be tailoring our sampling in the manner of Baland *et al*. (2010) to address hard-to-obtain baseline characteristics.

Interdisciplinarity: Biodiversity is affected by both the amount and the structure of habitats (availability of food and nesting resources within a patch, the connectivity of patches, edge effects) (Krebs, 2001; Turner *et al*., 2001). Reliance on geospatial data in the conservation evaluation studies begets a persistent disconnect between the outcomes researchers can study (deforestation, forest degradation), and the outcomes researchers often wish to study (ecosystem structure and function).[14] Most current studies rely on a binary geospatial metric (presence or absence of deforestation) to assess the ecological impacts of conservation measures. Micro studies of forest decentralization policies provide the few exceptions: fuelwood collected (Heltberg, 2001; Edmonds, 2002); percentage canopy cover per pixel (Somanathan *et al*., 2009); degree of lopping, presence of saplings, DBH, and canopy cover (Baland *et al*., 2010). While these provide significant improvements over binary geospatial measures, it is not immediately clear how these improved metrics relate to the ecosystem structure and function. For example, number of saplings in Baland *et al*. (2010) can be interpreted as a measure of the degree of forest regeneration or may be used as an indicator for disturbance.[15]

Biodiversity conservation is necessarily an interdisciplinary field because it affects people and ecosystems. For this reason, interdisciplinary partnerships between economists and natural scientists are needed, in order to improve future impact evaluation studies in terms of methodology and data collection. Currently, natural scientists seem

[14] To highlight the importance of establishing interdisciplinary partnerships between economists and natural scientists, here we focus on the ecological significance of the geospatial data rather than the technical quality of the available datasets. The latter can be a significant hurdle to good impact evaluation studies, as well: the presence of clouds, especially in tropical forests, the inability to distinguish between forest degradation and deforestation, on one hand, and between different types of tree species, on the other, as well as the lack of good metadata describing the methodology through which the geospatial datasets were obtained and the land-use categories classified, can significantly lower the reliability of geospatial datasets.

[15] The presence of saplings does not seem sufficient as these can be of invasive species that usually fare very well and grow very fast in disturbed areas; disturbance may actually result in changing the composition of the forest towards something that is no good for biodiversity conservation (Krebs, 2001).

to collect abundant data in research designs that preclude rigorous impact evaluations. While economists are generally well versed in impact evaluation techniques, they are often at a loss about the collection and interpretation of ecological data. For this reason, we highlight the need for interdisciplinary collaborations. Within a rigorous analytic framework, these can help identify the right spatio-temporal scale(s) of the analysis in terms of both the socio-economic and ecological processes; select the appropriate proxies and metrics for biodiversity and ecosystem function; and model connectivity and fragmentation.

V. Conclusion

Our review confirms previous claims that it is rare to find causal evidence on the effectiveness of conservation instruments commonly used in developing countries (Ferraro and Pattanayak, 2006; Carpenter *et al*., 2009). The limited evidence suggests that PAs cause modest reductions in deforestation and, thus, may positively affect biodiversity. However, the evidence base for PES, decentralization, and other interventions is much weaker. Because there is very limited geographic overlap between where PAs, PES, and decentralization are studied, we cannot compare the relative effectiveness of these three instruments. In short, despite progress in the last 6 years in the empirical evaluations of conservation programme impacts, the evidence base—limited to a handful of tables—is simply too shallow to say anything meaningful about the billions invested in protecting biodiversity.

Thus, the key messages of this paper are that the conservation evaluation literature needs (i) a larger number of rigorous studies from biodiversity-relevant locations and (ii) better theory, better methods, and better data. Specifically, we call for a programme of research—*Conservation Evaluation 2.0*—that not only uses theory of change to better characterize the mechanisms that trigger heterogeneous impacts varying with context, but also then conducts evaluations of different instruments in different contexts to permit cost–benefit comparisons of conservation instruments.

Achieving the goals of Conservation Evaluation 2.0 is constrained by the fact that few environmental policies and programmes are designed with evaluation in mind. Rather than hope we can find the relevant data and conditions to understand causal effects, heterogeneity, and mechanisms, we need more policies and programmes that are explicitly designed to allow rigorous evaluation of their environmental and social effects. We thus urge practitioners and scholars to implement more programmes with experimental and quasi-experimental designs. Strong candidates for experimental designs include programmes targeted on individuals, firms, local communities, or municipalities. Particularly appropriate would be pilot programmes or programmes implemented by non-governmental organization partners, which are not subject to the conflicting agendas of the various stakeholders, and may have more flexibility with regard to where and with whom they operate. Whether or not experimental or quasi-experimental designs are used, good baseline data on the relevant socio-economic and environmental factors are important for credible evaluations. Moreover, in order to ensure that such evaluations use the right data at the appropriate scale of analysis and can credibly estimate both socio-economic and environmental impacts, interdisciplinary partnerships between social and natural scientists are needed.

One reason why experimental and quasi-experimental designs are not the norm in conservation science is the perceived high costs of implementation (Ferraro and Pattanayak, 2006). We argue that the benefits exceed the costs because Conservation Evaluation 2.0 will help (i) identify and discontinue programmes for which the desired causal impacts cannot be detected, (ii) improve the impact and cost-effectiveness of existing programmes, and (iii) spur innovation. We also note that a cost–benefit analysis would imply that evaluation is most fruitfully applied to commonly used policies, like the three we review, and to policies and programmes that provide an opportunity to test fundamental behavioural questions such as: How do land-users respond to financial incentives? How do local government decision-makers respond to information or capacity building? It should be noted that evaluations involving fundamental behavioural questions are less about testing whether a specific project 'worked' and more about providing insights about the validity of the implicit and explicit causal models that underlie the global environmental investment portfolio.

Not all studies can be conducted with the research designs we discuss. Neither do we suggest that all studies employ one of these designs. However, conservationists—policy-makers, activists, and academicians alike—need to be aware of the limitations in the existing conservation evidence base as well as of the promise of empirical research designs that emphasize eliminating rival explanations for observed patterns of environmental outcome data and necessitate greater human capital for their application.

In conclusion, our review of studies with credible empirical designs highlights the continuing paucity of causal evidence on the effectiveness of common conservation instruments. Because of the publication bias towards studies that find the large positive impacts (Ravallion, 2009), it would be worthwhile to catalogue and review unpublished working papers, theses, and reports to funding agencies. Nonetheless, we urgently need more plain vanilla evaluations of economic and environmental outcomes from many more biodiversity-relevant locations. In addition, we also emphasize the need for a more advanced Conservation Evaluation 2.0 that seeks to measure how programme impacts vary by socio-political and bio-physical context, to track economic and environmental impacts jointly, to identify spatial spillover effects to untargeted areas, and to use theories of change to characterize causal mechanisms that can guide the collection of data and the interpretation of results. Only then can we usefully contribute to the debate about how to protect biodiversity in developing countries.

References

Abadie, A., and Imbens, G. W. (2006), 'Large Sample Properties of Matching Estimators for Average Treatment Effects', *Econometrica*, **74**(1), 235–67.

— — (2011), 'Bias-corrected Matching Estimators for Average Treatment Effects', *Business and Economic Statistics*, **29**(1), 1–11.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2004), 'Institutions as the Fundamental Cause of Long-run Growth', in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Amsterdam, North Holland.

Alix-Garcia, J., Shapiro, E., and Sims, K. R. E. (forthcoming), 'Forest Conservation and Slippage: Evidence from Mexico's National Payments for Ecosystem Services Program', *Land Economics*.

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., and Robalino, J. A. (2008), 'Measuring the Effectiveness of Protected Area Networks in Reducing Deforestation', *Proceedings of the National Academy of Sciences*, **105**(42), 16089–94.

—— Sims, K. R. E., Healy, A., and Holland, M. B. (2010), 'Protected Areas Reduced Poverty in Costa Rica and Thailand', *Proceedings of the National Academy of Sciences*, **107**(22), 9996–10001.

Andersson, K., and Gibson, C. C. (2007), 'Decentralized Governance and Environmental Change: Local Institutional Moderation of Deforestation in Bolivia', *Journal of Policy Analysis and Management*, **26**(1), 99–123.

Armsworth, P. R., Daily, G. C., Kareiva, P., and Sanchirico, J. N. (2006), 'Land Market Feedbacks Can Undermine Biodiversity Conservation', *Proceedings of the National Academy of Sciences*, **103**(14), 5403–8.

Arriagada, R. A. (2008), 'Private Provision of Public Goods: Applying Matching Methods to Evaluate Payments for Ecosystem Services in Costa Rica', North Carolina State University, available at http://search.proquest.com/docview/304534567?accountid=10598

— Sills, E. O., Pattanayak, S. K., and Ferraro, P. J. (2009), 'Combining Qualitative and Quantitative Methods to Evaluate Participation in Costa Rica's Program of Payments for Environmental Services', *Journal of Sustainable Forestry*, **28**(3–5), 343–67.

— Ferraro, P. J., Sills, E. O., Pattanayak, S. K., and Cordero-Sancho, S. (2012), 'Do Payments for Environmental Services Affect Forest Cover? A Farm-level Evaluation from Costa Rica', *Land Economics*, **88**(2), 382–99.

Baland, J.-M., Bardhan, P., Das, S., and Mookherjee, D. (2010), 'Forests to the People: Decentralization and Forest Degradation in the Indian Himalayas', *World Development*, **38**(11), 1642–56.

Bandyopadhyay, S., and Shyamsundar, P. (2004), 'Fuelwood Consumption and Participation in Community Forestry in India', World Bank Policy Research Working Paper 3331, Washington, DC, *World Bank*.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., and Ferrer, E. A. (2011), 'Has the Earth's Sixth Mass Extinction Already Arrived?', *Nature*, **471**(7336), 51–7.

Barrett, C. B., Travis, A. J., and Dasgupta, P. (2011), 'On Biodiversity Conservation and Poverty Traps', *Proceedings of the National Academy of Sciences*, **108**(34), 13907–12.

Bauch, S., Sills, E. O., Pattanayak, S. K. (2012), *'Have We Managed to Integrate Conservation and Development? ICDP Impacts in the Brazilian Amazon'*, NC State University.

Besley, T., and Coate, S. (2003), 'Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach', *Journal of Public Economics*, **87**(12), 2611–37.

— Persson, T. (2011), *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton, NJ, Princeton University Press.

Blackman, A., and Rivera, J. E. (2010), *'The Evidence Base for Environmental and Socioeconomic Impacts of Sustainable Certification'*, SSRN eLibrary.

Blom, B., Sunderland, T., and Murdiyarso, D. (2010), 'Getting REDD to Work Locally: Lessons Learned from Integrated Conservation and Development Projects', *Environmental Science and Policy*, **13**(2), 164–72.

Bowler, D. E., Buyung-Ali, L. M., Healey, J. R., Jones, J. P. G., Knight, T. M., and Pullin, A. S. (2011), 'Does Community Forest Management Provide Global Environmental Benefits and Improve Local Welfare?', *Frontiers in Ecology and the Environment*, **10**(1), 29–36.

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., and Sieber, S. (2011), 'The Political Economy of Deforestation in the Tropics', National Bureau of Economic Research Working Paper Series, No. 17417.

Carpenter, S. R., Mooney, H. A., Agard, J., Capistrano, D., DeFries, R. S., Díaz, S., Dietz, T., Duraiappah, A. K., Oteng-Yeboah, A., Pereira, H. M., Perrings, C., Reid, W. V., Sarukhan, J., Scholes, R. J., and Whyte, A. (2009), 'Science for Managing Ecosystem Services: Beyond the Millennium Ecosystem Assessment', *Proceedings of the National Academy of Sciences*, **106**(5), 1305–12.

Carvalho, S., and White, H. (2004), 'Theory-based Evaluation: The Case of Social Funds', *American Journal of Evaluation*, **25**(2), 141–60.

Coleman, E. A., and Fleischman, F. D. (2011), 'Comparing Forest Decentralization and Local Institutional Change in Bolivia, Kenya, Mexico, and Uganda', *World Development*, **40**(4), 836–49.

Dasgupta, P. (2009), 'Poverty Traps: Exploring the Complexity of Causation', in J. von Braun, R. Vargas Hill, and R. Pandya-Lorch (eds), *The Poorest and the Hungry: Assessments, Analyses, and Actions*, Washington, DC, International Food Policy Research Institute, 129–46.

— Mäler, K.-G. (2003), 'The Economics of Non-convex Ecosystems: Introduction', *Environmental and Resource Economics*, **26**(4), 499–525.

Deaton, A. (2010*a*), 'Instruments, Randomization, and Learning about Development', *Journal of Economic Literature*, **48**(2), 424–55.

— (2010*b*), 'Understanding the Mechanisms of Economic Development', *Journal of Economic Perspectives*, **24**(3), 3–16.

de Lima, A. C. B., Novaes Keppe, A. L., Correa Alves, M., Fernando Maule, R., and Sparovek, G. (2008), 'Impact of FSC Forest Certification on Agroextractive Communities of the State of Acre, Brazil', Instituto de Manejo e Certificaçao Florestal e Agricola (Imaflora), University of Sao Paulo (USP), Entropix Engineering Company, available at http://www.eco-index.org/search/pdfs/IMAFLORA-FSC.pdf

DiPrete, T. A., and Gangl, M. (2004), 'Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments', *Sociological Methodology*, **34**(1), 271–310.

Duflo, E. (2004), 'Scaling Up and Evaluation', in F. Bourguignon and B. Pleskovic (eds), *Accelerating Development*, Washington, DC, and Oxford, World Bank and Oxford University Press.

Edmonds, E. V. (2002), 'Government-initiated Community Resource Management and Local Resource Extraction from Nepal's Forests', *Journal of Development Economics*, **68**(1), 89–115.

Engel, S., Pagiola, S., and Wunder, S. (2008), 'Designing Payments for Environmental Services in Theory and Practice: An Overview of the Issues', *Ecological Economics*, **65**(4), 663–74.

Ferraro, P. J. (2008), 'Asymmetric Information and Contract Design for Payments for Environmental Services', *Ecological Economics*, **65**(4), 810–21.

— (2009), 'Counterfactual Thinking and Impact Evaluation in Environmental Policy', *New Directions for Evaluation*, **122**, 75–84.

— Pattanayak, S. K. (2006), 'Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments', *PLoS Biology*, **4**(4), e105.

— Hanauer, M. (2011), 'Protecting Ecosystems and Alleviating Poverty with Parks and Reserves: "Win–Win" or Tradeoffs?', *Environmental and Resource Economics*, **48**(2), 269–86.

— — Sims, K. R. E. (2011), 'Conditions Associated with Protected Area Success in Conservation and Poverty Reduction', *Proceedings of the National Academy of Sciences*, **108**(34), 13913–13918.

— Lawlor, K., Mullan, K. L., and Pattanayak, S. K. (2012), 'Forest Figures: Ecosystem Services Valuation and Policy Evaluation in Developing Countries', *Review of Environmental Economics and Policy*, **6**(1), 20–44.

Frondel, M., and Schmidt, C. M. (2005), 'Evaluating Environmental Programs: The Perspective of Modern Evaluation Research', *Ecological Economics*, **55**(4), 515–26.

Gaveau, D. L. A., Epting, J., Lyne, O., Linkie, M., Kumara, I., Kanninen, M., and Leader-Williams, N. (2009), 'Evaluating Whether Protected Areas Reduce Tropical Deforestation in Sumatra', *Journal of Biogeography*, **36**(11), 2165–75.

Greenstone, M., and Gayer, T. (2009), 'Quasi-Experimental and Experimental Approaches to Environmental Economics', *Journal of Environmental Economics and Management*, **57**(1), 21–44.

Haruna, A. (2010), 'Measuring Protected Areas' Impact on Deforestation in Panama', Duke University, available at http://hdl.handle.net/10161/2880.

Hayes, T., and Ostrom, E. (2005), 'Conserving the World's Forests: Are Protected Areas the Only Way?', *Indiana Law Review*, **38**, 595–618.

Heckman, J. J. (2010), 'Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy', *Journal of Economic Literature*, **48**(2), 356–98.
Heltberg, R. (2001), 'Determinants and Impact of Local Institutions for Common Resource Management', *Environment and Development Economics*, **6**(02), 183–208.
Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., . (2010), 'The Impact of Conservation on the Status of the World's Vertebrates', *Science*, **330**(6010), 1503–9.
Honey-Roses, J., Baylis, K., and Ramírez, M. I. (2011), 'A Spatially Explicit Estimate Of Avoided Forest Loss', *Conservation Biology*, **25**(5), 1032–43.
Imbens, G. W., and Wooldridge, J. M. (2009), 'Recent Developments in the Econometrics of Program Evaluation', *Journal of Economic Literature*, **47**(1), 5–86.
Jack, B. K., Kousky, C., and Sims, K. R. E. (2008), 'Designing Payments for Ecosystem Services: Lessons from Previous Experience with Incentive-based Mechanisms', *Proceedings of the National Academy of Sciences*, **105**(28), 9465–70.
Joppa, L. N., and Pfaff, A. (2009), 'High and Far: Biases in the Location of Protected Areas', *PLoS ONE*, **4**(12), e8273.
— — (2010), 'Reassessing The Forest Impacts of Protection', *Annals of the New York Academy of Sciences*, **1185**(1), 135–49.
— — (2011), 'Global Protected Area Impacts', *Proceedings of the Royal Society B: Biological Sciences*, **278**(1712), 1633–8.
Kleiman, D. G., Reading, R. P., Miller, B. J., Clark, T. W., Scott, J. M., Robinson, J., Wallace, R. L., Cabin, R. J., and Felleman, F. (2000), 'Improving the Evaluation of Conservation Programs', *Conservation Biology*, **14**(2), 356–65.
Krebs, C. (2001), *Ecology: The Experimental Analysis of Distribution and Abundance*, San Francisco, Benjamin Cummings.
Larson, A. M. (2002), 'Natural Resources and Decentralization in Nicaragua: Are Local Governments Up to the Job?', *World Development*, **30**(1), 17–31.
— Soto, F. (2008), 'Decentralization of Natural Resource Governance Regimes', *Annual Review of Environment and Resources*, **33**(1), 213–39.
Miteva, D. A., Murray, B. C., and Pattanayak, S. K. (2012*a*), 'Factors Determining the Effectiveness of Protected Areas in Preserving Mangroves in Indonesia: A Quasi-experimental Impact Evaluation', Working Paper, Duke University.
—Pattanayak, S. K., and Murray, B. C, (2012*b*), 'Protected Areas and Ecosystem Services in Indonesia', Working Paper, Duke University.
Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000), 'Biodiversity Hotspots for Conservation Priorities', *Nature*, **403**(6772), 853–8.
Nagendra, H. (2008), 'Do Parks Work? Impact of Protected Areas on Land Cover Clearing', *AMBIO: A Journal of the Human Environment*, **37**(5), 330–7.
Nelson, A., and Chomitz, K. M. (2011), 'Effectiveness of Strict vs Multiple Use Protected Areas in Reducing Tropical Forest Fires: A Global Analysis Using Matching Methods', *PLoS ONE*, **6**(8), e22722.
Pattanayak, S. K. (2009), 'Rough Guide to Impact Evaluation of Environmental and Development Programs', Working Paper 2187, eSocialSciences, available at http://ideas.repec.org/p/ess/wpaper/id2187.html
— Wunder, S., and Ferraro, P. J. (2010), 'Show Me the Money: Do Payments Supply Environmental Services in Developing Countries?', *Review of Environmental Economics and Policy*, **4**(2), 254–74.
— Ross, M. T., Depro, B. M., Bauch, S. C., Timmins, C., Wendland, K. J., and Alger, K. (2009), 'Climate Change and Conservation in Brazil: CGE Evaluation of Health and Wealth Impacts', *B.E. Journal of Economic Analysis and Policy*, **9**(2), 1.
Pfaff, A., Robalino, J., and Herrera, L. D. (2011), 'Decentralization Given Environment–Development Tradeoffs: Federal versus State Conservation and Impacts on Amazon deforestation', Working Paper, available at http://www.webmeets.com/files/papers/AERE/2011/536/PfaffRobalinoHerrera-PoliticalEconomy.pdf

Pfaff, A., Robalino, J., and Sanchez-Azofeifa, G. A. (2008), 'Payments for Environmental Services: Empirical Analysis for Costa Rica', Working Paper, Duke University, available at http://www.nber.org/public_html/confer/2007/si2007/EE/pfaff.pdf

— — — Andam, K. S., and Ferraro, P. J. (2009), 'Park Location Affects Forest Protection: Land Characteristics Cause Differences in Park Impacts Across Costa Rica', *B.E. Journal of Economic Analysis and Policy*, **9**(2).

Pullin, A. S., and Knight, T. M. (2001), 'Effectiveness in Conservation Practice: Pointers from Medicine and Public Health', *Conservation Biology*, **15**(1), 50–4.

Ravallion, M. (2007), 'Evaluating Anti-Poverty Programs', ch. 59 in T. P. Schultz and J. A. Strauss (eds), *Handbook of Development Economics*, **vol. 4**, 3787–846, Elsevier.

— (2009), 'Evaluation in the Practice of Development', *The World Bank Research Observer*, **24**(1), 29–53.

Rios A. R., and Pagiola S. (forthcoming), 'Poor Household Participation in Payments for Environmental Services in Nicaragua and Colombia', in L. Tacconi, S. Mohanty, and H. Suich (eds), *The Livelihood Impacts of Incentive Payments for Reduced Deforestation and Degradation*.

Robalino, J., Pfaff, A., Sanchez-Azofeifa, G. A., Alpizar, F., Leon, C., and Rodriguez M. (2008), 'Changing the Deforestation Impacts of Ecopayments: Evolution (2000–2005) in Costa Rica's PSA Program', Working Paper, CATIE, Turrialba, Costa Rica.

Rosenbaum, P. (2002), *Observational Studies*, 2nd edn, New York, Springer.

Ross, M., Depro, B., and Pattanayak, S. K. (2010), 'General Equilibrium Impacts of Costa Rica's Payments for Ecosystem Services', Working Paper, Duke University.

Salafsky, N., and Margoluis, R. (2003), 'What Conservation Can Learn from Other Fields about Monitoring and Evaluation', *BioScience*, **53**(2), 120–2.

— — Redford, K. H., and Robinson, J. G. (2002), 'Improving the Practice of Conservation: A Conceptual Framework and Research Agenda for Conservation Science', *Conservation Biology*, **16**(6), 1469–79.

Saterson, K. A., Christensen, N. L., Jackson, R. B., Kramer, R. A., Pimm, S. L., Smith, M. D., and Wiener, J. B. (2004), 'Disconnects in Evaluating the Relative Effectiveness of Conservation Strategies', *Conservation Biology*, **18**(3), 597–9.

Schwarze, S., and Juhrbandt, J. (2010), 'How Cost-effective are National Parks in Reducing Deforestation? The Cost-effectiveness of the Lore-Lindu National Park in Indonesia', paper presented at the International Society for Ecological Economics, 'Advancing Sustainability in a Time of Crisis', Oldenburg and Bremen, Germany.

Scullion, J., Thomas, C. W., Vogt, K. A., Perez-Maqueo, O., and Logsdon, M. G. (2011), 'Evaluating the Environmental Impact of Payments for Ecosystem Services in Coatepec (Mexico) Using Remote Sensing and On-site Interviews', *Environmental Conservation,* **38**(04), 426–34.

Sierra, R., and Russman, E. (2006), 'On the Efficiency of Environmental Service Payments: A Forest Conservation Assessment in the Osa Peninsula, Costa Rica', *Ecological Economics*, **59**(1), 131–41.

Sims, K. R. E. (2010), 'Conservation and Development: Evidence from Thai Protected Areas', *Journal of Environmental Economics and Management*, **60**(2), 94–114.

— (2011), 'Do Protected Areas Reduce Forest Fragmentation? A Microlandscapes Approach', submitted working paper, Amherst College.

Slingenberg, A., Braat, L., van der Windt, H., Rademaekers, K., Eichler, L., and Turner, K. (2009), 'Study on Understanding the Causes of Biodiversity Loss and the Policy Assessment Framework', European Commission Directorate-General for Environment, available at http://ec.europa.eu/environment/enveco/biodiversity/pdf/causes_biodiv_loss.pdf.

Somanathan, E., Prabhakar, R., and Mehta, B. S. (2009), 'Decentralization for Cost-effective Conservation', *Proceedings of the National Academy of Sciences*, **106**(11), 4143–7.

Sutherland, W. J. (2005), 'How Can We Make Conservation More Effective', *Oryx*, **39**, 1.

— Pullin, A. S., Dolman, P. M., and Knight, T. M. (2004), 'The Need for Evidence-based Conservation', *Trends in Ecology and Evolution*, **19**(6), 305–8.

Timmins, C., and Schlenker, W. (2009), 'Reduced-form Versus Structural Modeling in Environmental and Resource Economics', *Annual Review of Resource Economics*, **1**(1), 351–80.

Turner, M., Gardner, R., and O'Neill, R. (2001), *Landscape Ecology in Theory and Practice: Pattern and Process*, New York, Springer Verlag.

Vincent, J. R. (2010), 'Microeconomic Analysis of Innovative Environmental Programs in Developing Countries', *Review of Environmental Economics and Policy*, **4**(2), 221–33.

Weber, J. G., Sills, E. O., Bauch, S., and Pattanayak, S. K. (2011), 'Do ICDPs Work? An Empirical Evaluation of Forest-based Microenterprises in the Brazilian Amazon', *Land Economics*, **87**(4), 661–81.