Published in final edited form as:

Glob Environ Change. 2017 March ; 43: 148–160. doi:10.1016/j.gloenvcha.2017.02.002.

Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being

Erin O. Sills[a,b,1], **Claudio de Sassi**[b], **Pamela Jagger**[b,c], **Kathleen Lawlor**[d], **Daniela A. Miteva**[e], **Subhrendu K. Pattanayak**[f], and **William D. Sunderlin**[b]

[a]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695

[b]Center for International Forestry Research, Bogor, Indonesia 16000

[c]Department of Public Policy and Carolina Population Center, University of North Carolina at Chapel Hill, NC 27599

[d]Department of Economics, University of North Carolina at Asheville, NC 28804

[e]Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH 43210

[f]Sanford School of Public Policy and Nicholas School of the Environment, Duke University, Durham, NC 27708

Abstract

Climate change mitigation in developing countries is increasingly expected to generate co-benefits that help meet sustainable development goals. This has been an expectation and a hotly contested issue in REDD+ (reducing emissions from deforestation and forest degradation) since its inception. While the core purpose of REDD+ is to reduce carbon emissions, its legitimacy and success also depend on its impacts on local well-being. To effectively safeguard against negative impacts, we need to know whether and which well-being outcomes can be attributed to REDD+. Yet, distinguishing the effects of choosing particular areas for REDD+ from the effects of the interventions themselves remains a challenge. The Global Comparative Study (GCS) on REDD+ employed a quasi-experimental before-after-control-intervention (BACI) study design to address this challenge and evaluate the impacts of 16 REDD+ pilots across the tropics. We find that the GCS approach allows identification of control groups that represent the counterfactual, thereby permitting attribution of outcomes to REDD+. The GCS experience belies many of the common critiques of the BACI design, especially concerns about collecting baseline data on control groups. Our findings encourage and validate the early planning and up-front investments required to evaluate the local impacts of global climate change mitigation efforts with confidence. The stakes are high, both for the global environment and for local populations directly affected by those efforts. The standards for evidence should be concomitantly high.

[1]To whom correspondence should be addressed. Telephone: 919-515-7784, Fax: 919-515-6193, NC State University, Raleigh, NC, 27695-8008, sills@ncsu.edu.

Keywords

climate change mitigation; conservation and development; impact evaluation; REDD+; well-being

1. INTRODUCTION

While the importance of monitoring and evaluation has long been recognized by the conservation community (Christensen, 2003; Kleiman et al., 2000; Stem et al., 2005), research in the past decade has sharpened the focus on testing attribution and quantifying the causal impacts of conservation interventions (Baylis et al., 2016; Ferraro and Hanauer, 2014). This research focus is motivated by the concern that conservation advocates might have been spending "money for nothing" (Ferraro and Pattanayak, 2006) and is designed to support "evidence-based policy," similar to recent work in other realms of international development (e.g., research supported by 3ie and the Millennium Challenge Corporation). It aligns well with growing interest in results-based financing or "pay-for-performance" approaches in international aid sectors including health (Honda, 2012), education (Slavin, 2010), social protection (Davis et al., 2012), and conservation (Pattanayak et al., 2010). Results-based financing plays a potentially important role in climate change mitigation, including for reducing emissions from deforestation and forest degradation plus conservation, sustainable management of forests, and enhancement of forest carbon stocks in developing countries, or REDD+. The basic concept of REDD+ is to pay governments, communities, and/or individuals for verified reductions in deforestation and degradation (and associated reductions in carbon emissions) below an established 'reference-level' or counterfactual.

There are high hopes that REDD+ will be more effective than previous efforts to conserve tropical forests, because of the promise of relatively large and long-term financial assistance conditional on measured outcomes that are demonstrably "additional," or attributable to REDD+ activities (Venter and Koh, 2012). Although no single global system for REDD+ has emerged, funding has flowed through mechanisms such as the REDD+ Partnership (http://reddpluspartnership.org), the Governors' Climate and Forest Taskforce (www.gcftaskforce.org/), the Green Climate Fund (http://gcfund.org/), and voluntary carbon markets (Hamrick et al., 2015). In 2015, the 21st Conference of the Parties (COP) of the UNFCCC adopted guidelines for REDD+ and called for countries to take action to conserve and enhance, as appropriate, sinks and reservoirs of greenhouse gases.

While reducing carbon emissions is the primary motivation for REDD+, much of the policy dialogue, media coverage, and criticism has focused on potential co-benefits and costs for local people and biodiversity conservation (Agrawal et al., 2011; Burgess et al., 2013; Visseren-Hamakers et al., 2012). Impacts on local well-being - both positive and negative – will affect the feasibility, legitimacy and cost of REDD+ (Fisher et al., 2011; Lubowski and Rose, 2013), its success in achieving long-term reductions in forest carbon emissions (Chhatre et al., 2012; Lawlor et al., 2010), and the continued availability of finance from both public and private sectors (Lawlor et al., 2013). There is widespread concern about potential negative impacts on people who rely on the forests targeted for REDD+

interventions (Sunderlin et al., 2014), due to their historical exclusion from policy-making processes and fears that traditional land rights will not be recognized, and therefore opportunity costs of foregone traditional land uses not compensated. These concerns are exacerbated by the lack of clear evidence on the causal effects of previous forest conservation interventions (Miteva et al., 2012; Pattanayak et al., 2010).

In response to concerns about potential negative social impacts (i.e. direct impacts of interventions on local people), social safeguard policies were promulgated at the 16th COP (Decision 1/CP.16), and certification systems focused on monitoring these impacts - such as the Climate, Community, and Biodiversity (CCB) standards - have been widely adopted in voluntary carbon offset markets (Hamrick et al., 2015; Merger et al., 2011). These standards and safeguards require that REDD+ interventions be designed with local input to meet local needs, and be monitored and evaluated to assess their impacts on the well-being of local populations (Jagger et al., 2014). This has focused attention on how to measure local well-being, including livelihoods (e.g. collection of forest products) and welfare (e.g. household income). There has been less consideration of how to establish attribution (Agrawal et al., 2011; Caplow et al., 2010). Defining counterfactual scenarios that quantify what would have happened without REDD+ in order to assess the causal impacts of REDD+ on carbon emissions has been a key area of research and policy development (Olander et al., 2008; Romijn et al., 2015). We argue that social outcomes should also be compared to counterfactual outcomes in order to distinguish the impacts of interventions from the effects of where those interventions take place and contemporaneous policy and economic changes. However, there are unique challenges involved in designing monitoring and evaluation frameworks and obtaining the data required to apply such counterfactual thinking to the social impacts of conservation interventions, both because they cannot be observed objectively through remote sensing and because of confounding by human behaviors such as self-selection into participation.

Development of safeguard policies and certification standards would benefit from more systematic evidence on the social impacts of REDD+, including how they vary with intervention design and site characteristics. The more than 350 sub-national REDD+ pilot initiatives (Simonet et al., 2014; Sunderlin et al., 2014) offer an opportunity to generate this evidence based on real-world experience with REDD+ as it is being implemented on the ground. Recognizing these initiatives as an important testing ground for a new global system of forest conservation with uncertain impacts on local people, the Center for International Forestry Research (CIFOR) designed and implemented the Global Comparative Study on REDD+ (GCS), a quasi-experimental study including collection of “BACI” (before-after-control-intervention) data from a pan-tropical sample of households in 16 REDD+ sites in Brazil, Cameroon, Indonesia, Peru, Tanzania and Vietnam (Figure 1). In these six countries, CIFOR selected initiatives where it was possible to apply the BACI study design starting in 2010. This meant that the implementing organizations had defined their intervention areas - allowing assignment of villages to ‘control’ or ‘intervention’ status, but had not yet offered performance-based incentives - allowing data to be collected on conditions both ‘before’ and ‘after’ (Sunderlin et al., 2016). We provide a full accounting of the study design and methods employed by the GCS, which has both the broadest scope and largest household sample of any empirical study of REDD+ to date.

Sunderlin et al. (2016) demonstrate that the REDD+ initiatives included in the GCS are representative of the global population of pilot initiatives, using a database of all REDD+ initiatives compiled independently by CIRAD (Simonet et al. 2014). In this database, the means and proportions of initiatives with different characteristics are qualitatively similar in the GCS sample and in the entire population of initiatives. This supports the external validity of the GCS for understanding REDD+ initiatives. However, there remain major challenges to internal validity, including that the locations of these initiatives are not random and that REDD+ is rarely implemented in isolation but rather in the context of many prior and on-going conservation and development interventions. Careful study design is required to overcome these challenges.

CIFOR's Global Comparative Study on REDD+ demonstrates both the challenges and the feasibility of using a quasi-experimental design to evaluate how climate change mitigation and conservation interventions affect the well-being of local populations. In quasi-experimental impact evaluation, the critical methodological choices include when to collect data (i.e. whether to begin before the intervention) and how to define a comparison (or control) group to represent the counterfactual scenario (i.e. what would have happened in the absence of the intervention). These choices fundamentally affect the design of the evaluation, the ability to attribute outcomes to the intervention, and the credibility of impact estimates. We use the GCS baseline (or 'before' intervention) data to demonstrate the possibility of obtaining a balanced sample of control and intervention households through a two-step process of first "pre-matching" villages using information gathered through rapid rural appraisal and then "post-matching" households based on household and village survey data. The multi-country scope of the GCS both makes it a vitally important vehicle for learning about REDD+ and provides a general test of the BACI study design for evaluating the effects of conservation interventions on local well-being in a wide range of biophysical, cultural and economic settings. Comparative studies that apply the same methodology in multiple countries are critical for planning and assessing global conservation efforts such as REDD+.

2. ATTRIBUTION AND THE BACI STUDY DESIGN

When using observational data to evaluate impacts, the key challenge is ruling out alternative explanations such as contemporaneous economic and policy changes and selection bias (Ferraro and Pattanayak, 2006). For example, in the six countries where the GCS studied REDD+ initiatives, there were many other efforts to reduce deforestation and improve human well-being during the same time period as REDD+ implementation. Selection bias is also likely, because factors that influence the designation of REDD+ intervention areas (e.g., forest stock, rate of forest loss, biodiversity, quality of local governance (Lin et al. 2012)), could also affect well-being. Confounding factors can occur at multiple scales, ranging from the REDD+ site to the household. Those factors may be observable (i.e., quantifiable by the researcher, such as the deforestation rate) or unobservable (i.e., not measurable by the researcher, such as quality of local leadership). The effects of both types of confounders must be separated from the effects of the REDD+ interventions in order to estimate impacts. While we only consider quasi-experimental approaches to evaluating REDD+ impacts, Appendix A reviews alternatives.

Conducting a baseline survey before an intervention as part of a BACI study design is the best way to gather high-quality data on the outcomes, confounders, and other determinants of participation (Mullan et al., 2013; Ravallion, 2014). This requires significant advance planning and either a very large random sample of control villages or a sampling strategy to select control villages similar to the intervention villages. The main objective is to ensure enough overlap in the covariate distributions of intervention and control households, either to control for those without extrapolating outside the observed range of the data, or to identify a “balanced sub-sample” through *ex post* statistical matching. Specifically, the control and intervention households in a “balanced sub-sample” would have similar distributions of potentially confounding characteristics and parallel trends in outcomes prior to the intervention. If the survey sample includes such a balanced sub-sample, matching combined with differences-indifferences (DID) regression provides a relatively straight-forward and ‘doubly-robust’ estimation approach that controls for selection bias due to the observed factors (through their inclusion in the matching routine and as regression covariates) and time-invariant unobservables (through differencing out of linear outcome models) (Ho et al., 2007).

BACI data on conservation initiatives are rare in part because of the requirement for careful sample design and data collection *before* the intervention begins. Among the few examples are Pagiola et al. (2005), Wilkie et al. (2006), and Glew et al. (2012). While organizations planning REDD+ initiatives gather information on baseline deforestation rates in intervention and comparison areas in order to model and assess emission reductions against a counterfactual, they typically collect baseline socioeconomic data only in intervention areas. These data are essential for planning interventions and meeting certification requirements (e.g. the CCB Standards), and they may be collected in conjunction with efforts to obtain FPIC (free, prior and informed consent), which also requires REDD+ proponents to visit villages in the proposed intervention area. Such data support an alternative research design: comparing outcomes before and after (BA) the intervention among households subject to the intervention. A second design commonly used by researchers is to collect data after the intervention in both the intervention area and in similar areas not affected by the intervention in order to compare outcomes for control and intervention households (CI).

As explained in Jagger et al. (2010), there are three key reasons that BACI is theoretically a stronger research design than BA or CI. First, baseline data from both control and intervention areas can be used to model the selection mechanism, which supports identification of a matched sample for further analysis and assessment of the external validity of the impact evaluation. Second, outcomes in the intervention group can be compared to those in a matched comparison group, allowing the effects of REDD+ interventions to be disentangled from the effects of contemporaneous policy, market, and social changes (such as economic recessions or booms, unusual rainfall or temperature, anti-poverty initiatives) that could also influence outcomes. Third, the baseline data allow underlying differences in the intervention and control households to be netted out of impact estimates through DID, effectively eliminating bias from time-invariant factors even when those factors are not observable, in linear models of outcomes. Thus, in principle, a BACI study design with matching allows the analyst to isolate the causal effects of any conservation or climate change intervention such as REDD+, even when implemented in carefully selected sites that are subject to multiple other influences. However, this requires that the intervention and control units (or at least a matched sub-sample of those units) have similar distributions of potential confounders and parallel trends in outcomes under the counterfactual state of the world with no intervention (Abadie, 2005; Jagger et al. 2010). As compared to randomized controlled trials, there are likely to be many factors that differ systematically between control and intervention groups in operational REDD+ interventions, and if those influence outcomes, the assumption of parallel trends becomes untenable. Thus, the feasibility of collecting data that actually deliver on the promise of the BACI study design requires validation. As proof of concept, we describe the GCS experience with BACI, including both the process of establishing the study sample and the outcome of that process, i.e. demonstrating that it resulted in comparable sub-samples of intervention and control households with common support on potential confounders. We also show how BACI was integrated into the full study plan for the GCS, from the initial rapid rural appraisal exercise to obtain data for pre-matching villages, to the final use of baseline data for matching and bias-correction in estimation of impacts.

3. METHODS

To implement the BACI study design, the GCS collected data in four stages integrated into the study plan shown in Figure 2. As explained in Jagger et al. (2010), understanding the theory of change for an intervention is the starting point for determining what data need to be collected, especially on potential confounders that may influence both the placement and outcomes of the intervention. The GCS collected data on these factors first through a rapid assessment of observable village characteristics, and then through a baseline (“before”) survey of 2,056 households in 62 intervention villages and 1,994 households in 63 control villages. Details of the sampling, data collection, and data management for these two stages are provided in Sunderlin et al., (2016) and summarized here. We then describe the remaining elements of the GCS study plan and finally, our methods for evaluating whether the baseline data will allow attribution of social impacts to REDD+.

3.1 Sample

The six countries were selected to include a range of national trends and responses to deforestation. In those countries, we identified all sites where sub-national REDD+ initiatives were being planned, i.e. where there were detailed and well documented plans (i) for interventions to reduce net carbon emissions by reducing deforestation and degradation, or by implementing forest conservation, restoration, and management of existing forests, and (ii) for quantifying and reporting those net reductions in forest carbon emissions. For the GCS, we selected the 22 initiatives that had clearly defined intervention areas and intervention villages, but had not yet offered conditional incentives to reduce emissions from deforestation and degradation at the beginning of 2010. In most sites, the proponents had begun the process of obtaining FPIC in the intervention villages. For budgetary and logistical reasons, only village-level data were collected in 5 sites, while both village and household survey data were collected in the other 17. One of the 17 initiatives was in a very sparsely populated region, with no comparable villages outside the intervention area. Thus, a complete set of village and household BACI data were collected from the remaining 16 sites (Figure 1). Table B1 lists the full name, lead implementing organization (the proponent), and the jurisdiction where each REDD+ initiative is located.

As shown in Figure 2, once the initiatives had been selected, the GCS field teams obtained information on the boundaries of intervention areas and villages. To enable matching, they selected and characterized up to 15 villages in each intervention area as candidates for the sample. In initiatives that encompass more than 15 villages, they selected villages where specific interventions were planned and where deforestation rates were equal to or greater than the regional average. They then identified candidate 15 villages outside each intervention area that had similar market access, deforestation pressures, and livelihood strategies, but that were not expected to be affected by direct spillover effects from the REDD+ initiative, due to geographic or other barriers. The field teams characterized these villages using a rapid rural appraisal approach, i.e. by compiling secondary data, reviewing documentation on the initiative, and interviewing key informants. Specifically, they gathered data on potential confounders including population, land area, ethnicity, livelihoods, distances to roads and markets, forest dependence, forest cover, deforestation rate and pressures, local institutions, and experience with forest conservation NGOs (non-governmental organizations). This list of factors was informed by the theory of change for REDD+ (e.g. including site selection as described by (Lin, 2012)) and prior research on the determinants of well-being and forest use among rural communities (e.g. through the Poverty Environment Network described in Angelsen et al. (2014)).

With the rapid rural appraisal data on villages, we implemented covariate matching with a Mahalanobis metric (Gu and Rosenbaum, 1993) to identify a matched sample in each country including four intervention and four control villages per site (except in one site with fewer than four villages in the intervention area). The factors included in the matching represented: (1) deforestation pressures, (2) experience with forest conservation NGOs, (3) forest tenure, (4) village institutions, (5) population, (6) forest cover, (7) forest dependence, and (8) distance to main road. The statistical matching procedure was implemented by a central analytical team, not engaged in field data collection, to ensure procedural consistency across all countries and to minimize the influence of researcher preferences for particular field sites. The target sample size of 8 villages and 240 households per site was the maximum possible given budgetary and logistical constraints.

3.2 Baseline survey

In each village, a field team interviewed a random sample of households (with a target sample size of 30) and completed a village survey in a meeting that was widely publicized and open to any residents of the village. Both survey instruments (available from http://www.cifor.org/library/3286/technical-guidelines-for-research-on-redd-project-sites-with-survey-instruments-and-code-book/) included questions on socio-demographics, land use, forest resources, tenure, well-being, market access and institutions. We implemented the same procedures in intervention and control villages, so any possible bias associated with our methods of collecting and processing the data would not differ between intervention and control and therefore not confound estimates of impact.

Based on the household data, we constructed a measure of annual household income per adult equivalent from crop production, animal husbandry, forest-based activities, wage labor, and business, as well as transfers such as remittances and government payments, over the 12 months prior to the survey. We include both revenues from sales and the imputed values of household consumption of household production, minus the costs of purchased inputs. Thus, our measure of income could be considered "value-added" to household labor and land. Household incomes are converted to USD at 2010 market exchange rates from http://data.worldbank.org/indicator/PA.NUS.FCRF (after deflating to 2010 values in the few sites where the baseline survey was conducted later). Adult equivalents are calculated using the OECD equivalence scale, which assigns a value of 1 to the first household member, 0.7 to each additional adult, and 0.5 to each child (Organisation for Economic Co-operation and Development, 1982).

We identified and excluded "extreme income outliers," which we defined as observations with reported income greater than the 75th percentile or smaller than the 25th percentile by an amount more than 10 times the inner quartile range (IQR), without commensurate household wealth, including both physical and land assets. That is, we screened out observations based on mismatches between income and assets that suggested beyond a reasonable doubt that the extreme values reflected misreporting rather than real heterogeneity.

3.3 Follow-up field work

Before collecting "after" data, it is important to assess whether and which specific interventions have occurred, when, and where. This is similar to assessments of provision and utilization in public health (Habicht et al., 1999). The GCS field teams triangulated based on documentation, interviews with proponents, and village meetings. The resulting information provides the basis for deciding when to collect "after" data, for updating the survey instrument to elicit information on household participation in the interventions realized in the survey villages, and for defining which villages were subject to which interventions in order to estimate the effects of village-level interventions.

The "after" survey of villages and households must cover the same sample and use the same core survey instrument as the "before" phase in order to analyze differences in differences. In the GCS, we expanded the sample to include some households who had arrived after the baseline survey (sampled with the same probability as households present at baseline), and we expanded the survey instrument to capture information on all conservation and development interventions that had occurred between baseline data collection and our return visit.

3.4 Analysis of baseline data

The baseline household survey data, collected before the REDD+ intervention, can be used for matching, post-matching adjustments, and DID analysis of impacts. The proponents selected villages, rather than individual households, and we pre-matched control villages to those selected intervention villages. However, as demonstrated in the next section, the survey data may reveal incidental imbalances in household characteristics. This motivates *ex post* matching to identify a balanced sub-sample of intervention and control households. For example, the data could be used to estimate a household-level propensity score model for each country, including potential confounders at both the village and household levels. The estimated propensity score and the potential confounders could then be used in a covariate matching routine such as genmatch (Diamond and Sekhon, 2013), in order to find the best balanced sample of households possible given the sample established at baseline. Second, the baseline data can be used for post-matching adjustments to control for any remaining imbalance in both observables and correlated unobservables in the matched sample. Possibilities include bias-corrected matching (Abadie and Imbens, 2011), inverse probability weighting (Imbens and Wooldridge, 2009), and semi-parametric DID (Abadie, 2005). Third, for households in the balanced sample, the baseline and follow-up data can be used to calculate differences and estimate DID models.

3.5 Assessment of balance in GCS baseline data

To assess the balance of the household sample, we examine the central tendencies and distributions of potential confounders that reflect different dimensions of village and household characteristics. By design, the samples of intervention and control villages are balanced in terms of the rapid rural appraisal data. However, there may be differences in the samples that only become evident in the village survey data, because of more precise measurement than rapid rural appraisal. In addition to differences in village characteristics, there may also be incidental differences in household characteristics. Thus, we test for balance on a similar set of covariates as measured in the village survey, and similarly test covariates from the household survey. We focus our analysis on household income per adult equivalent at baseline, which is likely to influence most well-being outcomes at follow-up and could confound the effect of REDD+ if it were distributed differently among intervention and control households. Given that we do find differences between the full intervention and control samples, we examine whether and for which portions of the distribution there is sufficient overlap to identify a balanced sub-sample using covariate matching.

In both the pooled pan-tropical sample and each country sample, we compare the intervention and control villages and households first by t-tests for differences in the means of each covariate and a Hotelling test for the vector of covariates. A Hotelling's T^2 statistic less than the critical F value suggests that the intervention and control households are drawn from the same underlying distributions. Second, we calculate normalized or standardized differences. Because the normalized differences are not affected by sample size, they offer a more consistent way to assess balance across villages (N = 125) and households (N=4050). Different authors and software packages recommend slightly different versions of the standardized difference in means. We adopt Eq.1 from Austin (2009) and Abadie and Imbens (2011), where ~ indicates the mean and the subscripts i and c refer to intervention and control. We also follow Austin (2009) in considering normalized differences over 10% to indicate lack of balance.

$$\frac{(\tilde{x}_i - \tilde{x}_c)}{\sqrt{((\sigma^2{}_i - \sigma^2{}_c)/2)}} \quad [1]$$

Third, we compare the distributions of potential confounders in the intervention and control samples by examining box plots showing the median, IQR, and outliers. We use these plots to compare the IQR of the intervention and control samples and to assess whether outliers are concentrated in either the intervention or control samples, which would make it difficult to identify the effect of REDD+ on those households. Fourth, we compute the ratio of variances. This ratio is 1 for identical distributions, so we use an F test to assess whether the ratio falls outside of a 95% confidence interval around 1. Fifth, for continuous variables, we apply the Kolmogorov-Smirnov test, which measures the probability that both the control and intervention samples are drawn from the same underlying population. To allow for non-continuous distributions of our variables, we use bootstrapping to obtain p-values. Large p-values mean that the intervention and control households are likely to have been drawn from the same underlying distribution of observed covariates.

If the full intervention and control samples are not matched, the next step is to consider whether it would be possible to obtain a balanced sub-sample of intervention and control households through statistical matching based on the household and village survey data. To assess this, we examine the distributions of household income and other potential confounders at baseline using empirical quantile-quantile plots. In a balanced sample, the quantiles of a variable in the intervention and in the control group should fall along the 45 degree line in a square plot. In any quantiles with substantially more intervention than control households, balance could only be obtained by heavily weighting the relatively small number of control observations in those quantiles.

4. RESULTS

By covariate matching, we obtained samples of intervention and control villages balanced on their characteristics identified through rapid rural appraisal. This was facilitated by good balance on forest cover and the main sources of deforestation pressure in the candidate intervention and control villages identified by the field teams (Lin et al. 2012). Other factors, such as the count of village institutions, had significantly different means but overlapping distributions in the candidate intervention and control villages. The greatest difference between candidate intervention and control villages was that most candidate intervention villages, but very few candidate control villages, had prior experience with conservation NGOs (Lin et al. 2012). This reflects the fact that REDD+ often builds on prior conservation efforts (Sunderlin and Sills, 2012).

Table 1 confirms that the selected intervention and control villages also have statistically similar means on a range of characteristics measured though the village survey, with the exception of experience with external development interventions. Most villages – both intervention and control - reported that they had received some type of technical assistance, subsidized inputs, or other support from the government or civil society during the 12 months prior to the survey. This was true in all but three (95%) intervention villages. Because we pre-matched on rapid rural appraisal data about prior experience with conservation efforts, 85% of the control villages in our sample also had experience with external interventions. Thus, while the proportions are significantly different, there is substantial overlap. The lack of significant differences in all other village characteristics (according to t-tests) is at least in part because of the small sample size of 125 villages. The standardized or normalized differences, which are widely preferred measures of balance because they are not directly influenced by sample size (Imbens and Wooldridge, 2009), are nearly all greater than the 0.1 cut-off suggested by Austin (2009). As illustrated for Brazil and Indonesia (which together account for half of the REDD+ initiatives in our sample) in Table B2, the even smaller samples of villages in each country demonstrate the same pattern of no statistically significant differences in means but large normalized differences. The implication is that impacts on village-level outcomes would be challenging to estimate, because of the difficulty of controlling for potentially confounding differences across intervention and control villages. However, the key social outcomes of interest are largely defined at the household level, so we turn next to the question of obtaining a balanced household sample.

To estimate the impact of REDD+ on households, we need to rule out rival explanations for any differences in outcomes for the 2,056 intervention and 1,994 control households in the GCS sample, including any systematic differences in the types of households in control and intervention villages. An important potential confounder is baseline socioeconomic status, which we quantify as household income per adult equivalent, shown in Figure 3 for the pan-tropical sample and each country. Income at baseline is very likely to influence income at follow-up, and thus could either mask or exaggerate the effect of interventions if the probability of being included in the REDD+ intervention also varies with income. Figure 3 shows that the distribution of income is similar among intervention and control households in the pan-tropical sample, but suggests that there are differences in the distributions in the individual country samples.

The second and third rows of Table 2 present the same measures of balance as reported in Table 1 (probability values for t-tests of differences in means, and normalized differences) for household income in the pan-tropical sample and in each country sample. By these measures, the pan-tropical sample and all of the country samples except Vietnam are balanced, with p-values larger than 0.1 and normalized differences smaller than 10%. Table 2 also presents two additional measures of balance in the distribution of income: the ratio of variances, which would be 1 for identical distributions, and p-values for the Kolmogorov-Smirnov test, which would be large for control and intervention samples drawn from the same underlying population. These measures suggest that even though the central tendencies of the samples are similar, there are differences in the distributions of intervention and control households in all cases except Peru and Indonesia. Specifically, the ratio of variances falls outside the 95% confidence interval in the pan-tropical and all country samples except Peru and Indonesia. In addition, in the pan-tropical and Cameroon samples, the p-values for the boot-strapped Kolmogorov–Smirnov are small, suggesting that they are unlikely to be drawn from the same distributions.

Many other factors, in addition to initial socioeconomic status, can act as confounders. Balance statistics are reported in the Appendix for the full pan-tropical sample (Table B3), Brazil (Table B4), and Indonesia (Table B5) for 15 variables representing household demographics (age and gender composition of the household), socioeconomic status (home construction, cooking technology, electricity and water connections, perceived sufficiency of income), land holdings (hectares of forest and non-forest land managed), and forest use (income derived from the forest). By selecting variables that measure different dimensions of household well-being, including both livelihoods and living standards, we increase the likelihood that any unobserved confounders would be correlated with at least one of our selected variables, and thus be reflected in the balance statistics.

In contrast to income, Table B2 shows that the means of many other household characteristics are significantly different in the control and intervention samples, based on t-tests. Nearly half (seven out of 15) of the variables have significantly different intervention and control means, and a Hotelling test rejects the equality of means for the set of variables ($F = 6.782$, with an associated p-value of <0.001). While many of these differences are small, the difference in land area managed is large, with control households reporting a much larger area than intervention households on average. The t-tests indicate statistically strong differences in share of family members who are female (higher in control), mean age of family members (higher in control), type of cooking fuel used (intervention more likely to use solid fuels), electricity and piped water connections as well as latrines (less likely among intervention households). These differences suggest that proponents generally selected “poorer” villages, where households have less access to public services (electricity, water) and fewer assets (cookstoves that use modern fuels, latrines). The statistical significance is partly due to the large sample size (4035 to 4062 for each variable), which contributes to the statistically significant t-values even though none of the normalized differences are above the 10% criterion.

A well-balanced household sample would be similar not only in central tendencies but also in the full distributions of characteristics. Among the characteristics considered, only family size, health, and house quality clearly have both similar means and similar distributions among intervention and control households, based on the t-test, ratio of variances, and Kolmogorov-Smirnov test. On the other hand, a diverse set of potential confounders have different distributions among intervention and control households. This also holds true for a variable that we expected to be different among intervention and control households: whether the respondent has heard of REDD+. Respondents living in intervention areas are much more likely to have heard of REDD+, consistent with efforts to obtain FPIC.

Next we consider whether it will be possible to control for potential confounders by identifying a matched sub-sample of intervention and control households (Crump et al., 2009). This requires sufficient overlap in the characteristics of intervention and control households, which we assess with empirical quantile-quantile (eQQ) plots and illustrate for income in Figure 4. The intervention and control households have broadly similar distributions of income (close to the 45 degree line) except in the highest quantiles (Figure 4). This suggests that it will be possible to identify a sub-sample of households that are well-balanced in terms of socioeconomic status, allowing attribution of social impacts to REDD+

interventions. One caveat is that it will be more difficult to estimate impacts on high-income households, because in all country samples except Cameroon, there are more intervention than control households in the highest income quantiles. The implication is that high-income intervention households would be matched to a relatively small pool of high-income control households, increasing the variance of impact estimates. Alternatively, if the high-income intervention households are dropped from the matched sample, estimated average effects will not apply to them.

Another caveat is that we are not able to assess whether the suitability of the control group is compromised by differences in unobservable characteristics (e.g., risk aversion, entrepreneurship, or receptivity to external interventions). This concern has been raised specifically about BACI (Richards and Panfil, 2011), although it applies broadly to any quasi-experimental study design. For example, intervention households may be more receptive to external interventions, as suggested by the higher percentage of intervention villages that reported external support over the past 12 months. With BACI, this concern is alleviated by baseline data that allow matching and post-matching adjustment on variables representing many different dimensions of socioeconomic and forest conditions, at least some of which are likely to be correlated with the unobservables (e.g., education as a proxy for openness to external interventions). Further, we can control for time-invariant unobservables through use of DID, i.e. comparing trends in the intervention and control sites. Thus, this caveat really only applies to unobservable characteristics that (i) affect both selection into REDD+ and the outcome of interest, (ii) have variable effects over time, and (iii) are not correlated with observed characteristics in the survey data. In other words, the scope for concern is significantly narrowed by applying DID within a matched dataset.

5. DISCUSSION

The BACI approach and other quasi-experimental research designs have not been widely adopted in conservation, despite broad consensus that conservation and climate change interventions such as REDD+ should be "evidence-based." We believe that this is in large part due to skepticism about the utility, feasibility, and cost of collecting data outside the intervention area at baseline (prior to the intervention). In this section, we use the GCS experience to "ground-truth" concerns about cost; the relevance of control groups; the ethics and technical capacity to collect and analyze data on those groups; and about potential delays in implementation and in measurable impacts.

Rigorous impact evaluation – using any study design – is costly. For example, 3ie reports on their website (http://www.3ieimpact.org/en/funding/open-window/ow-faq/) that their typical grant for an impact evaluation is just under USD350,000. The cost for all four stages of GCS fieldwork (Figure 2) is USD$12,000 to USD$15,000 per village in Cameroon, Indonesia, and Vietnam, but three times higher in Brazil (largely due to the strong Brazilian economy and labor laws). Thus, the approximate total cost of field work per initiative ranges from USD$100,000 to USD$350,000, with personnel (including enumerators, encoders, drivers) accounting for the bulk of those costs.

Detailed household surveys impose a significant burden on respondents as well as the research team. However, if interventions are targeted to households and impacts are expected at the household-level, then clearly household data will provide the most precise measurement of those impacts, as well as allowing estimation of sub-group effects (e.g., impacts on households that are poor, headed by women, or members of ethnic minorities). Microeconomic heterogeneity is the rule rather than the exception (Heckman, 2001), implying that evaluations should be designed to consider sub-group effects rather than a single average "treatment effect" (Jagger, 2010). This is especially true in the context of recent policy debates surrounding social safeguards for REDD+ (Jagger et al., 2014).

Conservation organizations may be reluctant to incur costs in collecting baseline data from villages where they do not plan to work (Wongbusarakum et al., 2014). This requires field teams to visit additional villages and households, increasing the length of enumerator contracts and their transportation costs (relative to a BA study design). However, in the GCS, the largest component of transport costs (especially in Brazil and Indonesia) was for teams to travel to each region. Once teams were in the region, transportation costs to additional villages were relatively low. Further, adding control villages does not affect the costs of survey development and enumerator training. In sum, collecting survey data on a control group clearly increases costs but not at a linear rate with the increase in sample size.

It would be less expensive to only collect data after an intervention has begun (the CI study design), but in the case of REDD+, baseline data at least on the intervention area are required for certification and sale of carbon offsets (e.g., under Plan Vivo or CCB Standards). Such upfront costs – whether for BACI or certification - present budgetary challenges. As implemented by the GCS, BACI also involved additional costs to collect data (on up to 30 villages per site) for pre-matching. However, across the diversity of sites in the GCS, field teams were able to obtain sufficient information on village characteristics from secondary sources and key informants. This allowed us both to characterize site selection, e.g. identifying prior experience with a forest conservation NGO as a key characteristic of REDD+ project sites (Lin et al., 2012), and to employ statistical matching to identify a sample that was well-balanced on most village-level factors and that was informed but not unduly influenced by the field teams' prior experiences in the sites.

Implementing organizations stand to gain less from rigorous impact evaluation than donors or funders who can use findings to guide the selection and design of future projects (Köhlin et al., 2015; Swartzendruber and Khan, 2015). In that sense, impact evaluations are public goods, and it is not surprising that proponents are reluctant to invest much in their production. Donors could reduce the cost to implementing organizations by funding evaluation budgets and by reassuring implementers that findings will be used to improve rather than criticize their projects. In particular, donors could exert their influence to ensure collection of baseline data for quasi-experimental evaluations and to encourage randomization of interventions where possible (cf., Ferraro and Pattanayak 2006). The key is to encourage rigorous impact evaluation using the most appropriate methods for the most policy relevant interventions. As argued by Köhlin et al. (2015), donors should help close the persistent "know-do" gap in impact evaluation by steering evaluation effort towards the most relevant interventions. Pilot REDD+ initiatives are among the interventions that merit rigorous impact evaluation, because of the anticipated scaling-up of REDD+ to help meet the global <2 degree target for climate change (Zarin et al., 2016).

The goal of the GCS sampling approach was to survey control households similar to intervention households in every way except that they were not subject to the REDD+ intervention. We have demonstrated that the GCS obtained a reasonably balanced sample with sufficient overlap to allow matching. One risk of a well-balanced sample is that the control units are likely to be ‘next in line’ for the intervention, meaning that they could be lost from the control group. To reduce this risk, the GCS only studied REDD+ initiatives that were in the narrow window between finalizing site selection and beginning implementation at the time of baseline data collection. By including only sites where the intervention area had been clearly delineated, we were successful in picking control villages that remained in the control group: none of the 63 villages initially selected as controls have changed status to intervention villages.

Collection of data on a control group can raise ethical questions, because those households do not benefit from the intervention but still bear the burden of responding to the survey (i.e., “no survey without service” principle proposed for randomized control trials (Osrin et al., 2009)). Further, there is a common misconception that quasi-experimental methods require that control groups be actively prevented from participating in the intervention. In fact, these methods are typically used when researchers cannot control who participates in the intervention, and control groups often receive interventions later as part of a phased roll-out. In the GCS, researchers returned to all villages to present baseline results as an intermediate step between the “before” and “after” data collection, and to present follow-up results as the final step in field work. These reports on survey findings from the village and the broader site were positively received, with households in both control and intervention villages welcoming the researchers who arrived to share survey results, thus helping to address the “motivational problem” of control groups (Mosley, 1998).

In order to take full advantage of a BACI design, the research team must have considerable capacity for collection and analysis of data. Impact evaluation relies on high quality survey data, collected from a representative sample of households by a team of well-trained and carefully monitored enumerators. Once collected, data must be processed, analyzed, interpreted and presented by a team with strong skills in data management and analysis, including estimation of matching routines and multivariate regression models. New evaluations can build on previous efforts, e.g. using the GCS survey instruments and guidelines that are available from CIFOR’s website (Sunderlin et al., 2016). More generally, rigorous impact evaluation of REDD+ initiatives has been facilitated by early establishment of research partnerships between implementers and universities (Caplow et al., 2010). International research networks (e.g., the Collaboration for Environmental Evidence, Environment for Development, and the South Asian Network for Development and Environmental Economics) are also helping increase capacity through collaborative studies with in-country partners, training workshops and short-courses (Pattanayak, 2009).

The time-scale of interventions such as REDD+ can also create challenges for the BACI approach. Slow implementation of the pilots has been a challenge for the GCS. Based on assessments of provision and participation, CIFOR has delayed ‘after’ data collection by an average of one year across all sites to allow more time for REDD+ to be implemented on the ground. Of course, this also implies a delay in reporting findings about impacts. More fundamentally, the impacts of conservation interventions may only emerge over the long run. For example, changes in livelihoods strategies (e.g., shifting from dependence on shifting cultivation to a more diversified portfolio) may occur over the course of many years.

This partly explains the greater popularity of the CI study design, which can be implemented after the impacts of an intervention have become evident, allowing immediate estimation of those impacts. However, to control for selection bias, CI studies must rely on historical census data (usually only available in aggregated form), time-invariant factors (which are typically only a sub-set of the relevant confounders), or baseline conditions elicited through retrospective questions (with accurate recall likely only for factors considered most salient by the respondent (Jagger et al., 2012; Mullan et al., 2013)). Given that agriculture is the core livelihood activity of most households in the GCS sample, they may not have been able to accurately recall specifics of forest use and income in previous years. Nor is there likely to be detailed information on forest use in secondary data, e.g. from the population or agricultural census (Bakkegaard et al., 2015). Thus, the CI study design is likely to lead to an incomplete understanding of both the selection process and the processes generating the outcomes of interest in the absence of the intervention

Finally, in addition to being delayed, interventions may also evolve. While all of the initiatives selected for the GCS planned to offer performance-based conditional incentives, they have not all implemented this strategy. This reflects the evolution of REDD+ towards locally adapted bundles of interventions including education about forests and climate, tenure clarification, enforcement of forest regulations, and support for alternative livelihoods, as well as conditional benefits, all of which the GCS has tracked over time through the assessment of provision and participation (Sills et al., 2014). As with any quasi-experimental impact evaluation, the GCS will evaluate the impacts of interventions as they have actually been implemented under operational conditions.

6. CONCLUSION

The BACI study design offers a promising approach to evaluating REDD+ and other conservation and climate change interventions. One of the key challenges is being ready to collect baseline data in the often narrow time window after the boundaries of an intervention are defined but before implementation begins. The alternative is to re-construct baseline information after the intervention has taken place (Bamberger, 2010) but that also presents challenges (Ravallion, 2014). The detailed household survey implemented by the GCS at baseline measured and therefore allows the analyst to control for many potential confounders. At the same time, these data provide a rich description of who lives in and around REDD+ pilots and reveal patterns in site-selection for those pilots (Sills et al., 2014).

In the case of the GCS, the need to collect baseline data for the study design largely defined the study sample of 16 sub-national initiatives in 6 countries across the tropics (Sunderlin et al., 2016). While implementing a common methodology across 16 sites on three continents presented logistical challenges, it provides a more balanced picture than individual case studies. It also generated samples of intervention and control households with enough overlap to apply quasi-experimental methods such as matching with DID to estimate causal impacts. Thus, this approach provides rigorous evidence on the causal impacts of REDD+ on local well-being across the tropics, as needed for developing and implementing REDD+ social safeguards policies.

More specifically, the GCS demonstrates the feasibility and limitations of pre-matching villages to obtain a household sample appropriate for impact evaluation. The delineation of intervention areas for REDD+ initiatives is informed by both deforestation threats and socioeconomic conditions. From the perspective of implementing REDD+, this represents careful planning that should be expected of all conservation initiatives. From the perspective of evaluating REDD+, it represents selection bias that can confound impact estimates, unless the initial differences are controlled through careful sampling and analysis. In the GCS, field teams identified candidate intervention and comparison villages with balanced levels of forest cover and types of deforestation pressures, and they obtained proxy measures of other characteristics such as village institutions. Pre-matching based on those proxy measures led to samples of intervention and control households that are similar on average and that have somewhat different but overlapping distributions of key characteristics like baseline income, access to public services, and forest use (e.g., area and income from forests). This suggests that it will be possible to identify a balanced sub-sample in order to estimate the impacts REDD+.

The GCS experience belies many of the common critiques of the BACI approach as articulated by stakeholders including some of our implementing partners in the conservation and development community. With survey data from a sample of villages carefully pre-matched to ensure overlap in the distributions of confounders, we can determine whether and which outcomes can be attributed to REDD+. The GCS demonstrates that it is possible to design such a sample and collect comparable data across different countries, different size villages, and different forest types. More broadly, our goal is to encourage the early planning and significant up-front investments required to confidently attribute outcomes to conservation interventions. The stakes are high, both for the global environment and for the typically low-income rural populations in the places where these interventions are targeted. The standards for evidence on those interventions should be concomitantly high.

Acknowledgments

We thank the funders of the GCS (Norad, Australian Government Department of Foreign Affairs and Trade, Australian Aid, UKaid, the European Commission, and the CGIAR Fund), and the additional supporters of the Jagger et al. (2010) *Guide to learning about livelihood impacts of REDD+ projects* (ProFor, USAID, USDA Forest Service, Ministry of Foreign Affairs of Finland, and the David and Lucile Packard Foundation). We are indebted to the large team of researchers who implemented the methods and collected the baseline data described in this article. P.J. was supported by the Carolina Population Center and its NIH Center grant [P2C HD050924]. K.L. was supported under the Environmental Protection Agency (EPA) STAR Graduate Fellowship program [FP91714001]. D.M. was supported by the Duke University Graduate School.

References

Abadie A. Semiparametric Difference-in-Differences Estimators. Rev Econ Stud. 2005; 72:1–19. DOI: 10.1111/0034-6527.00321

Abadie A, Imbens GW. Bias-Corrected Matching Estimators for Average Treatment Effects. J Bus Econ Stat. 2011; 29:1–11. DOI: 10.1198/jbes.2009.07333

Agrawal A, Nepstad D, Chhatre A. Reducing Emissions from Deforestation and Forest Degradation. Annu Rev Environ Resour. 2011; 36:373–396. DOI: 10.1146/annurev-environ-042009-094508

Angelsen A, Jagger P, Babigumira R, Belcher B, Hogarth NJ, Bauch S, Börner J, Smith-Hall C, Wunder S. Environmental Income and Rural Livelihoods: A Global-Comparative Analysis. World Dev. 2014; 64:S12–28. DOI: 10.1016/j.worlddev.2014.03.006

Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and monte carlo simulations. Biometrical J. 2009; 51:171–184. DOI: 10.1002/bimj.200810488

Bakkegaard, RK., Agrawal, A., Animon, I., Bosselmann, A., Hogarth, N., Miller, D., Persha, L., Rametsteiner, E., Wunder, S., Zezza, A., Zheng, Y. FAO Forestry Paper no 179. FAO, CIFOR, IFRI, World Bank; Rome, Italy: 2015. National socioeconomic surveys in forestry: Guidance and survey modules for measuring the multiple roles of forests in household welfare and livelihoods. Available from: http://www.fao.org/3/a-i6206e.pdf

Bamberger M. Reconstructing Baseline Data for Impact Evaluation and Results Measurement. World Bank PREM Note No 4. 2010

Baylis K, Honey-Rosés J, Börner J, Corbera E, Ezzine-de-Blas D, Ferraro PJ, Lapeyre R, Persson UM, Pfaff A, Wunder S. Mainstreaming Impact Evaluation in Nature Conservation. Conserv Lett. 2016; 9(1):58–64. DOI: 10.1111/conl.12180

Burgess ND, Mwakalila S, Munishi P, Pfeifer M, Willcock S, Shirima D, Hamidu S, Bulenga GB, Rubens J, Machano H, Marchant R. REDD herrings or REDD menace: Response to Beymer-Farris and Bassett. Glob Environ Chang. 2013; 23:1349–1354. DOI: 10.1016/j.gloenvcha.2013.05.013

Caplow S, Jagger P, Lawlor K, Sills E. Evaluating land use and livelihood impacts of early forest carbon projects: Lessons for learning about REDD+ Environ Sci Policy. 2010; 14:152–167. DOI: 10.1016/j.envsci.2010.10.003

Chhatre A, Lakhanpal S, Larson AM, Nelson F, Ojha H, Rao J. Social safeguards and co-benefits in REDD+: a review of the adjacent possible. Curr Opin Environ Sustain. 2012; 4:654–660. DOI: 10.1016/j.cosust.2012.08.006

Christensen J. Auditing Conservation in an Age of Accountability. Conserv Pract. 2003; 4:12–18. DOI: 10.1111/j.1526-4629.2003.tb00065.x

Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. Biometrika. 2009; 96:187–199. DOI: 10.1093/biomet/asn055

Davis B, Gaarder M, Handa S, Yablonski J. Evaluating the impact of cash transfer programmes in sub-Saharan Africa: an introduction to the special issue. J Dev Eff. 2012; 4:1–8. DOI: 10.1080/19439342.2012.659024

Diamond A, Sekhon JS. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. Rev Econ Stat. 2013; 95:932–945. DOI: 10.1162/REST_a_00318

Ferraro PJ, Hanauer MM. Advances in Measuring the Environmental and Social Impacts of Environmental Programs. Annu Rev Environ Resour. 2014; 39:495–517. DOI: 10.1146/annurev-environ-101813-013230

Ferraro PJ, Pattanayak SK. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS Biol. 2006; 4:482–488. DOI: 10.1371/journal.pbio.0040105

Fisher B, Lewis SL, Burgess ND, Malimbwi RE, Munishi PK, Swetnam RD, Turner RK, Willcock S, Balmford A. Implementation and opportunity costs of reducing deforestation and forest degradation in Tanzania. Nat Clim Chang. 2011; 1:161–164. DOI: 10.1038/NCLIMATE1119

Glew, L., Mascia, M., Pakiding, F. Field Manual (version 1.0). World Wildlife Fund and Universitas Negeri Papua; Washington D.C. and Manokwari, Indonesia: 2012. Solving the Mystery of MPA Performance: monitoring social impacts.

Gu X, Rosenbaum P. Comparison of multivariate matching methods: Structures, distances, and algorithms. J Comput Graph Stat. 1993; 2:405–420. DOI: 10.2307/1390693

Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. Int J Epidemiol. 1999; 28:10–18. DOI: 10.1093/ije/28.1.10 [PubMed: 10195658]

Hamrick, K., Goldstein, A., Peters-Stanley, M., Gonzalez, G. Forest Trends. Washington DC: 2015. Ahead of the curve: State of the Voluntary Carbon Markets 2015. Available from: http://forest-trends.org/releases/p/ahead_of_the_curve_state_of_the_voluntary_carbon_markets_2015

Heckman JJ. Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture. J Polit Econ. 2001; 109:673–748. DOI: 10.1086/322086

Ho DE, Imai K, King G, Stuart EA. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Polit Anal. 2007; 15:199–236. DOI: 10.1093/pan/mpl013

Honda A. 10 best resources on … pay for performance in low- and middle-income countries. Heal Policy Plan. 2012; doi: 10.1093/heapol/czs078

Imbens GW, Wooldridge JM. Recent Developments in the Econometrics of Program Evaluation. J Econ Lit. 2009; 47:5–86. DOI: 10.1257/jel.47.1.5

Jagger, P. Forest sector reform, livelihoods and sustainability in western Uganda. In: German, LA.Karsenty, A., Tiani, A-M., editors. Governing Africa's Forests in a Globalized World. Earthscan; 2010. p. 103-125.

Jagger P, Brockhaus M, Duchelle A, Gebara M, Lawlor K, Resosudarmo I, Sunderlin W. Multi-Level Policy Dialogues, Processes, and Actions: Challenges and Opportunities for National REDD+ Safeguards Measurement, Reporting, and Verification (MRV). Forests. 2014; 5:2136–2162. DOI: 10.3390/f5092136

Jagger P, Luckert MK, Banana A, Bahati J. Asking Questions to Understand Rural Livelihoods: Comparing Disaggregated vs. Aggregated Approaches to Household Livelihood Questionnaires. World Dev. 2012; 40:1810–1823. doi:http://dx.doi.org/10.1016/j.worlddev.2012.04.030.

Jagger, P., Sills, EO., Lawlor, K., Sunderlin, WD. Occasional Paper 56. Cifor. CIFOR; Bogor, Indonesia: 2010. A guide to learning about livelihood impacts of REDD+ projects.

Kleiman DG, Reading RP, Miller BJ, Clark TW, Scott JM, Robinson J, Wallace RL, Cabin RJ, Felleman F. Improving the Evaluation of Conservation Programs. Conserv Biol. 2000; 14:356–365. DOI: 10.1046/j.1523-1739.2000.98553.x

Köhlin, G., Pattanayak, SK., Sills, EO., Mattsson, E., Ostwald, M., Salas, A., Ternald, D. EBA Report. Vol. 2015. Stockholm: 2015. In Search of Double Dividends From Climate Change Interventions Evidence From Forest Conservation and Household Energy Transitions; p. 09Available from: http://eba.se/wp-content/uploads/2016/03/Report201509climate.pdf

Lawlor K, Madeira EM, Blockhus J, Ganz DJ. Community Participation and Benefits in REDD+: A Review of Initial Outcomes and Lessons. Forests. 2013; 4:296–318. DOI: 10.3390/f4020296

Lawlor K, Weinthal E, Lander L. Institutions and Policies to Protect Rural Livelihoods in REDD plus Regimes. Glob Environ Polit. 2010; 10:1–11. DOI: 10.1162/GLEP_a_00028

Lin, L. Geography of REDD+ at Multiple Scales: Country Participation and Project Location. North Carolina State University; 2012. Available from: https://repository.lib.ncsu.edu/handle/1840.16/8200

Lin, L., Pattanayak, SK., Sills, EO., Sunderlin, WD. Site selection for forest carbon projects. In: Angelsen, A., editor. Analysing REDD+: Challenges and Choices. CIFOR; Bogor, Indonesia: 2012. p. 209-232.

Lubowski RN, Rose SK. The Potential for REDD+: Key Economic Modeling Insights and Issues. Rev Environ Econ Policy. 2013; 7:67–90. DOI: 10.1093/reep/res024

Merger E, Dutschke M, Verchot L. Options for REDD+ Voluntary Certification to Ensure Net GHG Benefits, Poverty Alleviation, Sustainable Management of Forests and Biodiversity Conservation. Forests. 2011; 2:550–577. DOI: 10.3390/f2020550

Miteva DA, Pattanayak SK, Ferraro PJ. Evaluation of biodiversity policy instruments: what works and what doesn't? Oxford Rev Econ Policy. 2012; 28:69–92. DOI: 10.1093/oxrep/grs009

Mosley, P. The use of control groups in impact assessments for microfinance, Social Finance Unit Working Paper No 19. Geneva: 1998.

Mullan K, Sills E, Bauch S. The Reliability of Retrospective Data on Asset Ownership as a Measure of Past Household Wealth. F. Methods. 2013; doi: 10.1177/1525822X13510370

Olander LP, Gibbs HK, Steininger M, Swenson JJ, Murray BC. Reference scenarios for deforestation and forest degradation in support of REDD: a review of data and methods. Environ Res Lett. 2008; 3doi: 10.1088/1748-9326/3/2/025011

Organisation for Economic Co-operation and Development. The OECD list of social indicators. OECD; Paris: 1982.

Osrin D, Azad K, Fernandez A, Manandhar DS, Mwansambo CW, Tripathy P, Costello AM. Ethical challenges in cluster randomized controlled trials: experiences from public health interventions in Africa and Asia. Bull World Health Organ. 2009; 87:772–779. DOI: 10.2471/BLT.08.051060 [PubMed: 19876544]

Pagiola S, Agostini P, Gobbi J, de Haan C, Ibrahim M, Murgueitio E, Ramírez E, Rosales M, Ruíz JP. Paying for Biodiversity Conservation Services. Mt Res Dev. 2005; 25:206–211. DOI: 10.1659/0276-4741(2005)025[0206:PFBCS]2.0.CO;2

Pattanayak, SK. SANDEE Working Paper 40-09. Kathmandu, Nepal: 2009. Rough guide to impact evaluation of environmental and development programs (No. 2187). Available from: https://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/4417/847_PUB_Working_Paper_40.pdf?sequence=1

Pattanayak SK, Wunder S, Ferraro PJ. Show Me the Money: Do Payments Supply Environmental Services in Developing Countries? Rev Environ Econ Policy. 2010; 4:254–274. DOI: 10.1093/reep/req006

Ravallion M. Can We Trust Shoestring Evaluations? World Bank Econ. Rev. 2014; 28:413–431. DOI: 10.1093/wber/lht016

Richards M, Panfil S. Towards cost-effective social impact assessment of REDD+ projects: meeting the challenge of multiple benefit standards. Int For Rev. 2011; 13:1–12. DOI: 10.1505/ifor.13.1.1

Romijn E, Lantican CB, Herold M, Lindquist E, Ochieng R, Wijaya A, Murdiyarso D, Verchot L. Assessing change in national forest monitoring capacities of 99 tropical countries. For Ecol Manage. 2015; 352:109–123. DOI: 10.1016/j.foreco.2015.06.003

Sills, EO.Atmadja, S.de Sassi, C.Duchelle, AE.Kweka, D.Resosudarmo, IAP., Sunderlin, WD., editors. REDD+ on the ground: A case book of subnational initiatives across the globe, REDD+ on the ground: A case book of subnational initiatives across the globe. CIFOR; Bogor, Indonesia: 2014.

Simonet, G., Karsenty, A., de Perthuis, C., Newton, P., Schaap, B. REDD+ projects in 2014: an overview based on a new database and typology; Les Cah la Chaire Econ du Clim Inf debates Ser. 2014. p. 34Available from: http://www.chaireeconomieduclimat.org/en/publications-en/information-debates/id-32-redd-projects-in-2014-an-overview-based-on-a-new-database-and-typology/

Slavin RE. Can financial incentives enhance educational outcomes? Evidence from international experiments. Educ Res Rev. 2010; 5:68–80. DOI: 10.1016/j.edurev.2009.12.001

Stem C, Margoluois R, Salafsky N, Brown M. Monitoring and Evaluation in Conservation: a Review of Trends and Approaches. Conserv Biol. 2005; 19:295–309. DOI: 10.1111/j.1523-1739.2005.00594.x

Sunderlin WD, Larson AM, Duchelle AE, Resosudarmo IAP, Huynh TB, Awono A, Dokken T. How are REDD+ Proponents Addressing Tenure Problems? Evidence from Brazil, Cameroon, Tanzania, Indonesia, and Vietnam. World Dev. 2014; 55:37–52. DOI: 10.1016/j.worlddev.2013.01.013

Sunderlin, WD., Larson, AM., Duchelle, AE., Sills, E., Luttrell, C., Jagger, P., Pattanayak, SK., Cronkleton, P., Ekaputri, AD., de Sassi, C., Aryani, R. Technical guidelines for research on REDD+ subnational initiatives. Bogor; Indonesia: 2016.

Sunderlin, WD., Pratama, CD., Bos, AB., Avitabile, V., Sills, EO., de Sassi, C., Joseph, S., Agustavia, M., Pribadi, UA., Anandadas, A. REDD+ on the Ground: A Case Book of Subnational Initiatives across the Globe. Center for International Forestry Research (CIFOR); Bogor, Indonesia: 2014. REDD+ on the ground: The need for scientific evidence; p. 2-21.

Sunderlin, WD., Sills, EO. REDD+ projects as a hybrid of old and new forest conservation approaches. In: Angelsen, A., editor. Analysing REDD+: Challenges and Choices. Center for International Forestry Research (CIFOR); Bogor, Indonesia: 2012. p. 177-191.

Swartzendruber F, Khan MN. Evaluation of Natural Resource Management Interventions Linked to Climate Change: A Scoping Study. Climate-Eval Community of Practice. 2015; doi: 10.1017/CBO9781107415324.004

Venter O, Koh LP. Reducing emissions from deforestation and forest degradation (REDD+): game changer or just another quick fix? Ann N Y Acad Sci Annals of the New York Academy of Sciences. 2012; 1249:137–150. DOI: 10.1111/j.1749-6632.2011.06306.x

Visseren-Hamakers IJ, McDermott C, Vijge MJ, Cashore B. Trade-offs, co-benefits and safeguards: current debates on the breadth of REDD+ Curr Opin Environ Sustain. 2012; 4:646–653. DOI: 10.1016/j.cosust.2012.10.005

Wilkie DS, Morelli GA, Demmer J, Starkey M, Telfer P, Steil M. Parks and people: Assessing the human welfare effects of establishing protected areas for biodiversity conservation. Conserv Biol. 2006; 20:247–249. DOI: 10.1111/j.1523-1739.2005.00291.x [PubMed: 16909679]

Wongbusarakum, S., Madeira, EM., Hartanto, H. The Nature Conservancy. Arlington, VA: 2014. Strengthening the Social Impacts of Sustainable Landscapes Programs: A practitioner's guidebook to strengthen and monitor human well-being outcomes. Available from: http://www.conservationgateway.org/ConservationPractices/PeopleConservation/SocialScience/Pages/strengthening-social-impacts.aspx

Zarin DJ, Harris NL, Baccini A, Aksenov D, Hansen MC, Azevedo-Ramos C, Azevedo T, Margono BA, Alencar AC, Gabris C, Allegretti A, Potapov P, Farina M, Walker WS, Shevade VS, Loboda TV, Turubanova S, Tyukavina A. Can carbon emissions from tropical deforestation drop by 50% in 5 years? Glob Chang Biol. 2016; 22doi: 10.1111/gcb.13153

REFERENCES

Andam KS, Ferraro PJ, Sims KRE, Healy A, Holland MB. Protected areas reduced poverty in Costa Rica and Thailand. Proc Natl Acad Sci. 2010; 107:9996–10001. DOI: 10.1073/pnas.0914177107 [PubMed: 20498058]

Angelsen A. REDD models and baselines. Int For Rev. 2008; 10:465–475. DOI: 10.1505/ifor.10.3.465

Angelsen A, Jagger P, Babigumira R, Belcher B, Hogarth NJ, Bauch S, Börner J, Smith-Hall C, Wunder S. Environmental Income and Rural Livelihoods: A Global-Comparative Analysis. World Dev. 2014; doi: 10.1016/j.worlddev.2014.03.006

Angelsen A, Rudel TK. Designing and Implementing Effective REDD + Policies: A Forest Transition Approach. Rev Environ Econ POLICY. 2013; 7:91–113. DOI: 10.1093/reep/res022

Arriagada RA, Sills EO, Ferraro PJ, Pattanayak SK. Do Payments Pay Off? Evidence from Participation in Costa Rica's PES Program. PLoS One. 2015; 10:e0131544.doi: 10.1371/journal.pone.0131544 [PubMed: 26162000]

Bauch SC, Sills EO, Pattanayak SK. Have We Managed to Integrate Conservation and Development? ICDP Impacts in the Brazilian Amazon. World Dev. 2014; 64:S135–S148. DOI: 10.1016/j.worlddev.2014.03.009

Bottrill M, Cheng S, Garside R, Wongbusarakum S, Roe D, Holland MB, Edmond J, Turner WR. What are the impacts of nature conservation interventions on human well-being: a systematic map protocol. Environ Evid. 2014; 3:1–11. DOI: 10.1186/2047-2382-3-16

Clements T, Suon S, Wilkie DS, Milner-Gulland EJ. Impacts of Protected Areas on Local Livelihoods in Cambodia. World Dev. 2014; 64:S125–S134. DOI: 10.1016/j.worlddev.2014.03.008

Herold, M., Angelsen, A., Verchot, LV., Wijaya, A., Ainembabazi, JH. A stepwise framework for developing REDD+ reference levels. In: Angelsen, A., editor. Analysing REDD+: Challenges and Choices. CIFOR; Bogor, Indonesia: 2012. p. 279-299.

Jagger, P. Forest sector reform, livelihoods and sustainability in western Uganda. In: German, LA.Karsenty, A., Tiani, A-M., editors. Governing Africa's Forests in a Globalized World. Earthscan; 2010. p. 103-125.

McNally CG, Uchida E, Gold AJ. The effect of a protected area on the tradeoffs between short-run and long-run benefits from mangrove ecosystems. Proc Natl Acad Sci. 2011; 108:13945–13950. DOI: 10.1073/pnas.1101825108 [PubMed: 21873182]

Miteva DA, Loucks CJ, Pattanayak SK. Social and Environmental Impacts of Forest Management Certification in Indonesia. PLoS One. 2015; 10:e0129675.doi: 10.1371/journal.pone.0129675 [PubMed: 26132491]

Potvin C, Tschakert P, Lebel F, Kirby K, Barrios H, Bocariza J, Caisamo J, Caisamo L, Cansari C, Casamá J, Casamá M, Chamorra L, Dumasa N, Goldenberg S, Guainora V, Hayes P, Moore T, Ruíz J. A participatory approach to the establishment of a baseline scenario for a reforestation Clean Development Mechanism project. Mitig Adapt Strateg Glob Chang. 2007; 12:1341–1362. DOI: 10.1007/s11027-006-9056-3

Richards M, Panfil S. Towards cost-effective social impact assessment of REDD+ projects: meeting the challenge of multiple benefit standards. Int For Rev. 2011; 13:1–12. DOI: 10.1505/ifor.13.1.1

Robalino, J., Sandoval, C., Villalobos, L., Alpízar, F. Local Effects of Payments for Environmental Services on Poverty - Discussion Paper Series (No. DP 14-12). Costa Rica: 2014.

Scherr SJ. A downward spiral? Research evidence on the relationship between poverty and natural resource degradation. Food Policy. 2000; 25:479–498. DOI: 10.1016/S0306-9192(00)00022-1

Wunder S. Poverty Alleviation and Tropical Forests - What Scope for Synergies? World Dev. 2001; 29:1817–1833. DOI: 10.1016/S0305-750X(01)00070-5

APPENDIX A

The key requirement for attributing outcomes to REDD+ is to understand what would have happened in the absence of REDD+, i.e. the counterfactual to REDD+. The major alternatives for establishing the counterfactual are (1) simulation models, (2) experiments (randomized controlled trials), and (3) quasi-experimental methods such as adopted by the GCS.

The first alternative is widely used to establish carbon reference levels by drawing on theory, prior empirical evidence, and/or stakeholder views.[*] Modelers rely on historical data about land-use change, often assuming that prior trends would continue in a future without REDD + (Angelsen and Rudel, 2013; Herold et al., 2012). Models may be informed by economic theory about supply and demand and empirical evidence on the relationships between deforestation and factors such as commodity prices, agricultural suitability of land, infrastructure (roads), and population density (Angelsen, 2008). Such modeling approaches could potentially also be used to establish counterfactual socio-economic outcomes. However, the empirical evidence is more limited because time-series data on households in the rural landscapes likely to be affected by REDD+ are often lacking, and where they exist, datasets typically omit key aspects of forest-based livelihoods (Angelsen et al., 2014). Theory regarding poverty-environment relationships is also underdeveloped, partly because the relationships between natural resources degradation and rural poverty in developing countries are inherently complex and often endogenous, and partly because researchers have not uncovered any general principles describing how changes in land-use affect changes in human well-being (Bottrill et al., 2014; Scherr, 2000; Wunder, 2001). The lack of historical

[*]Stakeholder views have been incorporated through participatory methods that elicit stakeholders' predictions about land-use change in a "no-REDD" scenario (Potvin et al., 2007). Participatory visioning exercises (e.g., using the "Participatory Theory of Change" method described in (Richards and Panfil, 2011)) have also generated important insights, especially on unexpected impacts and impact pathways, although they depend entirely on stakeholders' willingness and ability to make clear predictions about complex scenarios.

data, robust theory, and empirical evidence on the poverty-environment relationships makes it difficult to simulate counterfactual welfare outcomes under REDD+.

The second alternative is the ‘experimental’ approach of randomizing interventions. Rather than relying on statistical techniques to identify or construct appropriate controls, villages or households would be randomly assigned to the intervention or control status, potentially as part of a phased roll-out in which the intervention will eventually encompass all study units. When implementing organizations are willing to cede control over the placement, timing, or bundle of interventions, they can be randomly allocated in order to test the impact of different approaches. Because such experiments are likely be limited to a small sub-set of the project design space, they may have low external validity. In the specific context of REDD+, randomization is not likely to be accepted by implementing organizations whose primary objectives are to deliver cost-effective carbon offsets (for sale in voluntary markets) and to demonstrate the feasibility of REDD+ (to support its inclusion in the UNFCCC agenda), even if testing and generating lessons about REDD+ are secondary objectives.

Third, quasi-experimental strategies to minimize bias from confounding include instrumental variables, regression discontinuity, matching, DID, and matching combined with DID. All of these have been employed to evaluate the welfare impacts of conservation interventions (Andam et al., 2010; Arriagada et al., 2015; Bauch et al., 2014; Clements et al., 2014; Jagger, 2010; McNally et al., 2011; Miteva et al., 2015; Robalino et al., 2014). Use of the first two methods is limited to cases where either an instrumental variable or a quantitative treatment assignment variable is available. The last three methods require data on conditions before the intervention, including data on outcomes for DID and data on confounders for matching. These data can be drawn from secondary sources (Clements et al., 2014; Miteva et al., 2015), retrospective questions (Arriagada et al., 2015), or baseline surveys such as used in the GCS.

APPENDIX B

Table B1

Subnational REDD+ initiatives

	Initiative name	Proponent	Location
BRAZIL	Acre State System of Incentives for Environmental Services (SISA)	Instituto de Mudanças Clímaticas (IMC)	State of Acre
	Sustainable Landscapes Pilot Program in São Félix do Xingu	The Nature Conservancy Brazil	State of Pará
	Cotriguaçu Sempre Verde	Instituto Centro de Vida (ICV)	State of Mato Grosso
	Sustainable Settlements in the Amazon	Instituto de Pesquisa Ambiental da Amazônia (IPAM)	State of Pará
PERU	REDD Project in Brazil Nut Concessions in Madre de Dios, Peru	Bosques Amazonicos (BAM)	Department of Madre de Dios
	Valuation of Environmental Services in Managed Forests of Seven Indigenous Communities in Ucayali, Peru	Asociación para la Investigación y Desarrollo Integral (AIDER)	Department of Ucayali

	Initiative name	Proponent	Location
CAMEROON	Community Payments for Ecosystem Services (PES) in the South and East Regions of Cameroon	Centre pour l'Environnement et le Développement (CED)	South and East Regions
	Mount Cameroon REDD Project	GFA-Envest	Southwest Region
TANZANIA	Community Based REDD Mechanisms for Sustainable Forest Management in Semi-Arid Areas	Tanzania Traditional Energy Development and Environment Organization (TaTEDO)	Shinyanga
	Making REDD Work for Communities and Forest Conservation in Tanzania	Tanzania Forest Conservation Group (TFCG)	Kilosa, Morogoro
INDONESIA	Reducing Carbon Emissions from Deforestation in the Ulu Masen Ecosystem	Government of Aceh (Task Force REDD Aceh)	Aceh
	Ketapang Community Carbon Pools (KCCP)	Fauna and Flora International Indonesia (FFI)	West Kalimantan
	Kalimantan Forests and Climate Partnership (KFCP)	Australian Aid/Kalimantan Forests and Climate Partnership	Central Kalimantan
	The Nature Conservancy within Berau Forest Carbon Program	The Nature Conservancy (TNC)	East Kalimantan
	Katingan Peatland Restoration and Conservation Project	PT. Rimba Makmur Utama (RMU)	Central Kalimantan
VIETNAM	Cat Loc Landscape – Cat Tien National Park Pro-Poor REDD+ Project	The Netherlands Development Organisation (SNV)	Lam Dong Province

Table B2

Characteristics of Intervention and Control Villages in GCS Samples in Brazil and Indonesia

	Brazil					Indonesia				
	N	Means		p>|t|	norm diffs	N	Means		p>|t|	norm diffs
		Treated	Control				Treated	Control		
Village land (hectares)	30	46652.9	84987.8	0.596	−0.201	38	39552.1	26523.5	0.458	0.245
Village forest land (hectares)	30	29553.8	22479.1	0.754	0.114	35	19029.7	10543.1	0.250	0.394
Population of village (people)	31	701.5	402.5	0.101	0.608	39	959.9	1201.2	0.436	−0.252
Distance to nearest road usable by cars during all seasons (km)	32	3.8	18.6	0.122	−0.574	41	16.1	4.4	0.117	0.501
Distance to nearest market (km)	32	48.3	39.4	0.471	0.258	41	27.9	67.9	0.269	−0.357
Received external support in past 12 months (0/1)	32	0.9	0.7	0.078	0.655	41	1.0	1.0	0.329	−0.309
Village decision-making institutions (count)	32	2.5	1.3	0.002	1.230	41	3.8	3.9	0.896	−0.041

Table B3

Balance of household characteristics in pan-tropical GCS sample

		Mean					
Variable	N‡	Treated	Control	T-test p-value	Norm diffs	var ratio (Tr/Co)	KS p-value (bootstrap)
household size	4062	4.826	4.876	0.523	−0.007	0.999	0.186
share of household members who are female	4062	0.459	0.473	0.032	−0.024	0.939	0.070
mean years of schooling per household member	4052	4.785	4.851	0.441	−0.009	0.958	0.016
mean age of household members	4060	28.223	29.105	0.039	−0.023	0.886 *	0.266
mean days ill during past year per household member	4056	10.610	11.092	0.593	−0.006	1.000	0.886
household belongs to largest ethnic group (binary)	4062	0.781	0.784	0.821	−0.002	1.009 *	
roof of house made from low quality materials (binary)	4052	0.244	0.261	0.204	−0.014	0.955	
household cooks with biomass fuel (fuelwood, charcoal, or other vegetation) (binary)	4060	0.864	0.811	<0.001	0.050	0.769 *	
household does not have electricity connection (binary)	4060	0.457	0.370	<0.001	0.062	1.064	
household does not have piped drinking water from well or water system (binary)	4060	0.895	0.849	<0.001	0.049	0.735 *	
household does not have private latrine (binary)	4060	0.481	0.434	0.003	0.033	1.016	
household perceives income as insufficient (binary)	4054	0.393	0.420	0.080	−0.019	0.979	

Variable	N‡	Mean		T-test p-value	Norm diffs	var ratio (Tr/Co)	KS p-value (bootstrap)
		Treated	Control				
total HA of land used by household, excluding mature forest	4035	17.007	17.115	0.965	−0.001	0.144 *	0.004
total HA of mature forest used by household	4035	61.944	127.770	0.118	−0.018	60.663 *	0.180
income from forest, per adult equivalent in USD	4050	586.103	552.832	0.590	0.024	1.117 *	0.312
heard of REDD+ (binary) #	3573	0.246	0.136	<0.001	0.099	1.572 *	

‡Sample sizes smaller than 4062 reflect missing values and exclusion of obvious errors, such as years of schooling greater than age and implausible reports of land assets.

#Households in intervention villages where the concept of REDD+ had not yet been introduced were not asked whether they had heard of REDD+, in order to avoid raising concerns or generating confusion about the intentions of implementing organizations.

*Variance ratios significantly different from 1 at 95% confidence level, based on F test

Table B4

Balance of household characteristics in GCS sample in Brazil

Variable	N‡	Mean		T-test p-value	Norm diffs	var ratio (Tr/Co)	KS p-value (bootstrap)
		Treated	Control				
household size	996	4.400	4.272	0.380	0.079	0.909	0.132
share of household members who are female	996	0.421	0.424	0.812	−0.021	0.844 *	0.508
mean years of schooling per household member	993	3.618	3.292	0.032	0.193	0.872	<0.001
mean age of household members	996	29.571	32.829	<0.001	−0.319	0.697 *	0.022
mean days ill during past year per household member	996	13.273	16.991	0.078	−0.159	0.563 *	0.014
household belongs to largest ethnic group (binary)	996	0.988	0.992	0.575	−0.050	1.424 *	
roof of house made from low quality materials (binary)	988	0.246	0.236	0.727	0.031	1.027	

Variable	N‡	Mean		T-test p-value	Norm diffs	var ratio (Tr/Co)	KS p-value (bootstrap)
		Treated	Control				
household does not use electricity (binary)	994	0.508	0.298	<0.001	0.620	1.196 *	
household cooks with biomass fuel (fuelwood, charcoal, or other vegetation) (binary)	994	0.804	0.696	<0.001	0.354	0.745 *	
household does not have piped drinking water from well or water system (binary)	994	0.943	0.884	<0.001	0.211	0.524 *	
household does not have private latrine (binary)	994	0.400	0.403	0.926	−0.004	0.998	
total HA of land used by household, excluding mature forest	995	1.847	20.709	<0.001	−0.609	0.044 *	<0.001
total HA of mature forest used by household	994	51.740	280.619	0.168	−0.125	0.001 *	0.010
income from forest, per adult equivalent in USD	994	226.010	349.039	0.018	−0.214	0.390 *	0.004
household perceives income as insufficient (binary)	996	0.325	0.305	0.477	0.064	1.037	
heard of REDD+ (binary) #	507	0.231	0.148	<0.001	0.300	1.407 *	

‡ Sample sizes smaller than 996 reflect both missing values and exclusion of obvious errors, such as years of schooling greater than age and implausible reports of land assets.

Households in intervention villages where the concept of REDD+ had not yet been introduced were not asked whether they had heard of REDD+, in order to avoid raising concerns or generating confusion about the intentions of implementing organizations.

* Variance ratios significantly different from 1 at 95% confidence level, based on F test.

Table B5

Balance of household characteristics in GCS sample in Indonesia

		Mean					
Variable	**N‡**	**Treated**	**Control**	**T-test p-value**	**Norm diffs**	**var ratio (Tr/Co)**	**KS p-value (bootstrap)**
household size	1349	4.431	4.620	0.067	−0.141	1.013	0.168
share of household members who are female	1349	0.482	0.495	0.211	−0.096	1.027	0.262
mean years of schooling per household member	1349	4.758	5.379	<0.001	−0.371	0.962	<0.001
mean age of household members	1349	27.621	28.675	0.124	−0.119	0.947	0.084
mean days ill during past year per household member	1347	9.637	8.937	0.663	0.034	1.331 *	0.318
household belongs to largest ethnic group (binary)	1349	0.834	0.842	0.677	−0.032	1.042	
roof of house made from low quality materials (binary)	1349	0.211	0.263	0.026	−0.172	0.857 *	
household does not use electricity (binary)	1349	0.300	0.166	<0.001	0.454	1.518 *	
household cooks with biomass fuel (fuelwood, charcoal, or other vegetation) (binary)	1349	0.848	0.761	<0.001	0.311	0.701 *	
household does not have piped drinking water from well or water system (binary)	1349	0.973	0.799	<0.001	0.809	0.158 *	
household does not have private latrine (binary)	1349	0.687	0.597	<0.001	0.267	0.893	
total HA of land used by household, excluding mature forest	1349	0.824	0.640	0.664	0.033	1.343 *	0.004
total HA of mature forest	1324	1.110	0.908	0.836	0.016	1.488 *	0.412

		Mean					
Variable	**N‡**	**Treated**	**Control**	**T-test p-value**	**Norm diffs**	**var ratio (Tr/Co)**	**KS p-value (bootstrap)**
used by household							
income from forest, per adult equivalent in USD	1348	131.959	104.622	0.162	0.108	0.758 *	0.022
household perceives income as insufficient (binary)	1348	0.230	0.250	0.402	−0.065	0.946	
heard of REDD+ (binary)	1349	0.077	0.020	<0.001	0.379	3.656 *	

‡Sample sizes smaller than 1349 reflect both missing values and exclusion of obvious errors, such as implausible reports of land assets.

*Variance ratios significantly different from 1 at 95% confidence level, based on F test.

Highlights

- REDD+ pilots should be rigorously evaluated for impacts on local well-being.
- We demonstrate the before-after-control-intervention design for evaluating impacts.
- Baseline data on intervention and control households allow attribution.

Figure 1.
Study sites: REDD+ Pilot Initiatives

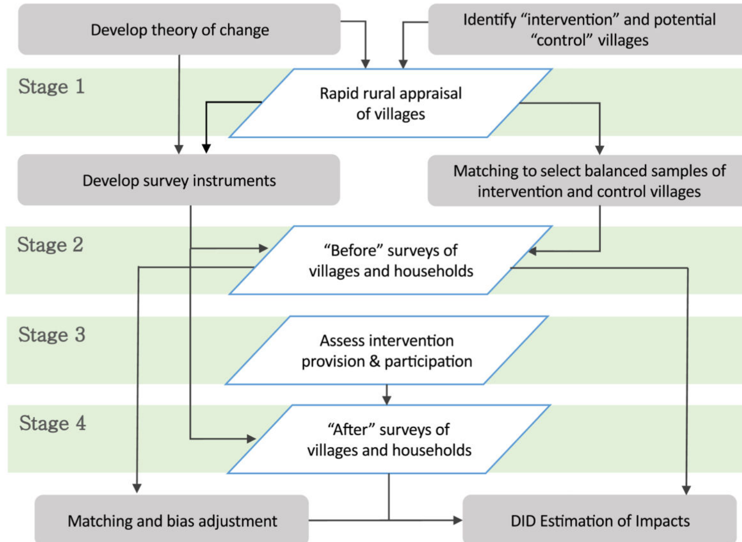

Fig. 2.

GCS Study Plan: White quadrilaterals show data collection in four stages of field work. Gray rectangles show preparatory steps and data analysis. Arrows indicate dependencies on prior steps.

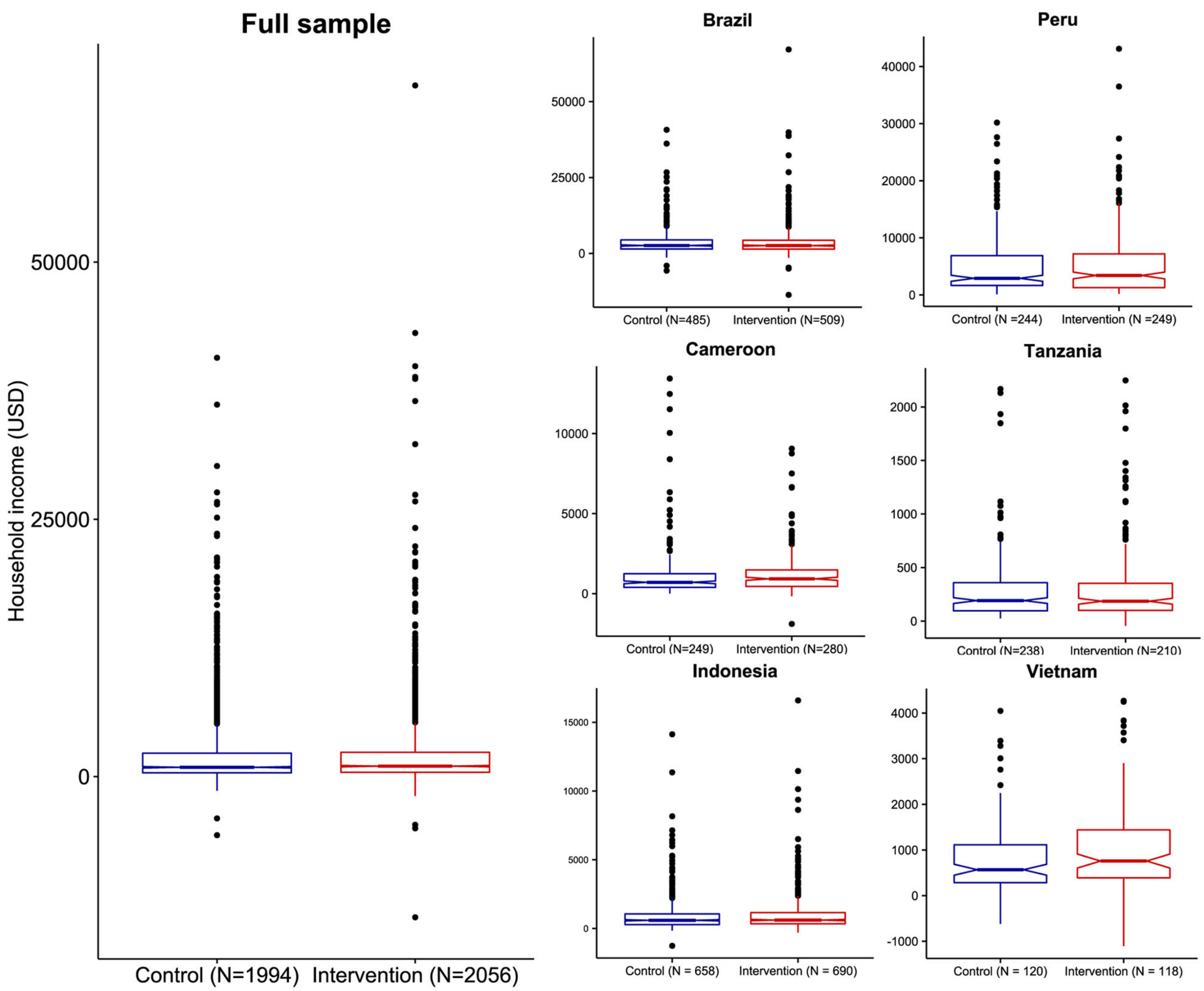

Fig. 3.

Box plots of household income in 2010 USD per adult equivalent in the pan-tropical and each country sample: Boxes show the inner quartile range (IQR) and median; lines span 1.5 IQR from the 75th and 25th percentiles and dots show observations outside this range.

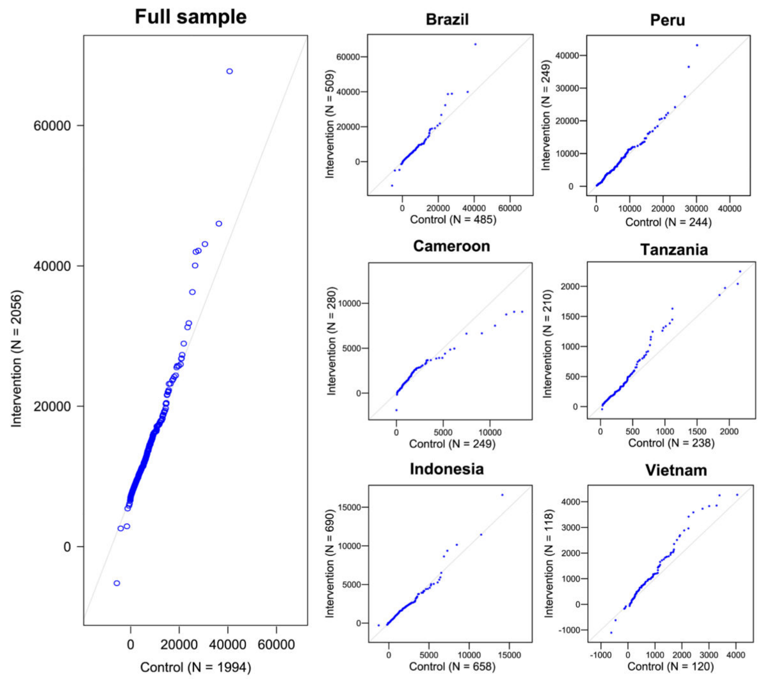

Fig. 4.

Empirical quantile–quantile plots of household income in 2010 USD per adult equivalent in the pan-tropical and each country sample, after removing extreme outliers: A sample balanced between intervention and control observations would fall along the 45 degree line. Extreme outliers are observations with reported income greater than the 75th percentile or smaller than the 25th percentile by an amount more than 10 times the inner quartile range (IQR), without commensurate household wealth, including both physical and land assets.

Table 1

Characteristics of Intervention and Control Villages in Pan-tropical GCS Sample

		Means			
	N‡	Treated	Control	p>|t|	norm diffs
Village land (hectares)	113	34283	35574	0.95	−0.01264
Village forest land (hectares)	109	20515	13769	0.36	0.17574
Population of village (people)	121	868	1134	0.15	−0.26416
Distance to nearest road usable by cars during all seasons (kilometers)	121	8.5	9.8	0.74	−0.06062
Distance to nearest market (kilometers)	120	34.0	43.7	0.49	−0.13896
Received external support in past 12 months (0/1)	125	95%	85%	0.07	0.32868
Count of village decision-making institutions (number)	125	3.7	3.2	0.21	0.22495

‡Sample sizes smaller than 125 reflect missing answer(s) in one or more villages.

Table 2

Differences in household income per adult equivalent between intervention and control households in pan-tropical and each country sample

	Pan-tropical	**Brazil**	**Peru**	**Cameroon**	**Tanzania**	**Indonesia**	**Vietnam**
N	4050	994	493	529	448	1348	238
Income-C	2036.84	3677.00	4983.76	1175.81	279.52	920.47	809.30
Income-I	2313.70	3993.24	5281.16	1267.68	309.26	962.59	1034.67
p>|t|	0.380	0.984	0.552	0.503	0.347	0.542	0.055
Norm Diff	0.028	−0.001	0.054	0.059	0.087	0.034	0.250
Variance Ratio	1.319*	1.609*	1.258	0.625*	1.362*	1.071	1.646*
KS P-value (boot-strapped)	0.058	0.749	0.628	0.018	0.824	0.147	0.113

*Variance ratios significantly different from 1 at 95% confidence level, based on F test.