Received: 10 September 2018 | Revised: 29 November 2018 | Accepted: 3 December 2018

DOI: 10.1002/csp2.8

WILEY Conservation Science and Practice
A journal of the Society for Conservation Biology Open Access

CONTRIBUTED PAPER

Experimental evaluation of the impact of a payment for environmental services program on deforestation

Emma Wiik[1] | Rémi d'Annunzio[2] | Edwin Pynegar[1] | David Crespo[3] | Nigel Asquith[4,5] | Julia P. G. Jones[1]

[1]School of Environment, Natural Resources and Geography, Bangor University, Bangor, UK

[2]Forestry Department, FAO, Rome, Italy

[3]Independent

[4]Charles Bullard Fellowship Program, Harvard Forest, Harvard University, Petersham, Massachusetts

[5]Sustainability Science Program, Harvard Kennedy School, Harvard University, Cambridge, Massachusetts

Correspondence
Julia P. G. Jones, School of Environment, Natural Resources and Geography, Bangor University, Bangor, UK.
Email: julia.jones@bangor.ac.uk

Funding information
Leverhulme Trust, Grant/Award Number: RPG-2014-056 ; Natural Environment Research Council, Grant/Award Number: NE/ L001470/ 1

Despite calls for greater use of randomized control trials (RCTs) to evaluate the impact of conservation interventions; such experimental evaluations remain extremely rare. Payments for environmental services (PES) are widely used to slow tropical deforestation but there is widespread recognition of the need for better evidence of effectiveness. A Bolivian nongovernmental organization took the unusual step of randomizing the communities where its conservation incentive program (*Watershared*) was offered. We explore the impact of the program on deforestation over 5 years by applying generalized additive models to Global Forest Change (GFC) data. The "intention-to-treat" model (where units are analyzed as randomized regardless of whether the intervention was delivered as planned) shows no effect; deforestation did not differ between the control and treatment communities. However, uptake of the intervention varied across communities so we also explored whether higher uptake might reduce deforestation. We found evidence of a small effect at high uptake but the result should be treated with caution. RCTs will not always be appropriate for evaluating conservation interventions due to ethical and practical considerations. Despite these challenges, randomization can improve causal inference and deserves more attention from those interested in improving the evidence base for conservation.

KEYWORDS
deforestation, effectiveness, efficacy, experimental evaluation, forest conservation, impact evaluation, intention-to-treat, land use change, payments for ecosystem services, PES

1 | INTRODUCTION

Following calls for improvements in the quality of evidence underpinning conservation interventions (Ferraro & Pattanayak, 2006), there are a rapidly growing number of robust conservation impact evaluations. Impact evaluation seeks to establish the extent to which an outcome can be attributed to the intervention itself, rather than to confounding factors (Baylis et al., 2016; Ferraro & Hanauer, 2014). Careful statistical analysis is increasingly used for constructing counterfactuals (what would have happened in the absence of the intervention). For example, statistical matching is now quite widely used (e.g., Eklund et al., 2016;

Data accessibility: Data and code to reproduce analysis in this paper are available here: doi.org/10.6084/m9.figshare.7418264. The full details of a baseline and endline social survey of participants and non-participants in *Watershared* from control and treatment communities (a small amount of data from this is used in the paper) is publically archived (Bottazzi et al., 2017).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2019 The Authors. Conservation Science and Practice published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology

Rasolofoson, Ferraro, Jenkins, & Jones, 2015; Sills et al., 2017) while other quasi-experimental methods which require particular conditions such as instrumental variables (Sims, 2010) or regression discontinuity (Alix-Garcia, Mcintosh, Sims, & Welch, 2013), have spread more slowly. Randomized control trials (RCTs) where units are experimentally allocated to treatment or control reduce the influence of confounding factors (Ferraro & Hanauer, 2014) and therefore, at least in theory, greatly improve the quality of causal inference. RCTs at the field scale have been the mainstay of applied ecology for decades, however are vanishingly rare at the landscape scale despite calls for wider use (Ferraro, 2011; Miteva, Pattanayak, & Ferraro, 2012; Pattanayak, Wunder, & Ferraro, 2010; Samii, Lisiecki, Kulkarni, Paler, & Chavis, 2014).

The rarity of RCT in evaluating the impact of large-scale conservation interventions can be attributed to the numerous practical and ethical considerations involved (Baylis et al., 2016; Pynegar, Jones, Gibbons, & Asquith, 2018). One of these is scale itself: it clearly would not be feasible to randomly allocate Protected Areas in a landscape. Furthermore, despite the enthusiasm with which RCTs have been promoted in some fields such as development economics, interpretation is not always simple and randomization does not relieve one of the needs to consider covariates and confounding factors (Deaton & Cartwright, 2018). Finally, RCTs require involvement of researchers throughout the implementation phase; they cannot be conducted post-hoc. All are likely to be important reasons for the limited number of RCTs evaluating large-scale conservation interventions.

A useful distinction in any impact evaluation is between "effectiveness" and "efficacy" (how interventions work in real-world practice versus under ideal implementation; Pullin & Knight, 2001). Effectiveness may be low not because the intervention lacks efficacy but because implementation, uptake and adherence are imperfect (Glennerster & Takavarasha, 2013). When analyzing RCTs, including the outcomes for individuals *as randomized* in "intention-to-treat" (ITT) estimates is widely considered most appropriate for evaluating real world effectiveness (Gupta, 2011). Where uptake is incomplete, examining outcomes according to uptake and adherence can be informative, especially for exploring the potential efficacy of new approaches (Glennerster & Takavarasha, 2013; Ten Have et al., 2008). For example an "as-treated" impact estimate (where units are analyzed as they were treated rather than as they were randomized) can be useful (McNamee, 2009).

Payments for environmental services (PES, also known as Payments for Ecosystem Services; Wunder, 2015), which incentivize land managers to provide ecosystem services, have been promoted to slow tropical deforestation since the late 1990s (Landell-Mills & Porras, 2002; Sánchez-Azofeifa, Pfaff, Robalino, & Boomhower, 2007). While strong evidence on PES impacts is limited (Börner et al., 2017; Miteva et al., 2012, Samii et al., 2014), approaches such as statistical matching have been quite widely used to evaluate deforestation impacts for example in Costa-Rica (Robalino & Pfaff, 2013) and Cambodia (Clements & Milner-Gulland, 2015). Regression discontinuity was recently used to evaluate the impact of payments on land management actions in Mexico (Alix-Garcia et al., 2018). The only RCT to evaluate PES (Jayachandran et al., 2017) suggested—for high forest pressure, low opportunity cost, and the requirement to enroll all of one's forest land—that a PES scheme in Uganda cost-effectively reduced deforestation over a two-year period. Given the heterogeneity of PES impacts across varied settings, and few evaluations relative to the exploding number of programs (Salzman, Bennett, Carroll, Goldstein, & Jenkins, 2018), more such RCTs would be valuable.

In 2010, the Bolivian nongovernmental organization *Fundación Natura Bolivia* (*Natura*) and five municipal governments initiated an RCT of their conservation incentive program known as *Watershared* (Pynegar et al., 2018). *Watershared* makes in-kind compensations to incentivize landowners to cease deforestation and cattle grazing on enrolled parcels. A total of 129 communities were randomly allocated to treatment or control (offered agreements or not). We investigate the effectiveness and efficacy of *Watershared* at reducing deforestation, over 5 years, by applying generalized additive models (GAMs) to global forest change (GFC) data (Hansen et al., 2013). We undertake a standard ITT evaluation to explore effectiveness at the level of randomization regardless of uptake of *Watershared* agreements in individual communities. We further quantify efficacy by evaluating the effect of uptake on deforestation (c.f. "as-treated" analysis). Throughout, we control for factors that can relate to both uptake and deforestation, including propensity to enroll (endogeneity), and consider the potential influence of unobserved confounding factors.

2 | METHODS

2.1 | Study context

Since 2003, *Natura*'s *Watershared* program in the Bolivian Andes has used in-kind incentives to encourage land owners to conserve forests, to preserve exceptional biodiversity, store carbon, and ensure locally valued ecosystem services (Asquith, 2016). *Watershared* is not a PES scheme according to the original definition involving buyers and sellers of services (Wunder, 2007), however it does involve "voluntary transactions between service users and service providers that are conditional on agreed rules of natural resource management for generating offsite services" (Wunder, 2015). Therefore the *Watershared* scheme is relevant to those interested in the design of conservation incentive schemes such as PES. In exchange for enrolling parcels of land in *Watershared* agreements, farmers receive varied forms of support

(including fruit trees, bee boxes, irrigation material and barbed wire) to help shift away from swidden agriculture and improve livestock management (Bottazzi, Wiik, Crespo, & Jones, 2018). More than 210,000 ha belonging to 4,500 families are under agreements (Asquith, 2016).

The study region: The Río Grande Valles Cruceños Natural Integrated Management Area (Spanish acronym ANMI) is a 734,000-ha protected area in the Santa Cruz valleys of Bolivia, created in 2007 (Figure 1a). There are regional differences in rainfall which contribute to the existence of five ecoregions which we simplified to three (Appendix S1, Supporting Information): Tucuman-Bolivian Forest; Chaco; and the dry valleys. The area is home to approximately 20,000 people scattered across small towns and hamlets. Most people farm using a mixed system of staple crops including maize and potato, small-scale vegetable cultivation, and livestock rearing. Cattle are grazed in the forests for at least part of each year.

RCT: In 2010 *Natura*, motivated by a desire to know if their intervention was effective, decided to roll out *Watershared* in 129 communities in the ANMI as an RCT to facilitate impact evaluation (Pynegar et al., 2018). Following baseline data collection, including a socioeconomic survey (Bottazzi et al., 2017), communities were randomly allocated to control ($n = 64$) or treatment ($n = 65$), stratified by cattle ownership and population density. However, when our team later constructed community boundaries using national data (National Institute of Agrarian Reform) and field validation we found that two neighboring control communities were in practice considered as one and did not have separate boundaries. Thus, we examine 128 communities (control $n = 63$ and treatment $n = 65$).

The randomization was consented to by municipal leaders on the grounds that the program would subsequently be implemented in all communities (this occurred in 2016 and the program now runs in both treatment and control communities). *Watershared* agreements were offered to households in treatment communities. There were three levels of agreement with slightly different conditions and incentives (SI 2). For example, the strictest level (level 1) only applied to forest within 100 m of a stream and cattle had to be excluded as well as deforestation stopped. The other two levels did not require cattle exclusion (SI 2). While the analysis looking at the impact of *Watershared* on water quality (Pynegar et al., 2018), considered only level 1 agreements, in this paper investigating the impact of *Watershared* on deforestation we include all levels. Compliance for level 1 and 2 agreements was monitored annually by *Natura* technicians walking transects within the parcels under agreement. Level 3 agreements did not receive active monitoring. In cases of gross noncompliance, in-kind incentives (such as irrigation tubing or bee hives) have been redistributed to the community. As with many such schemes, not all land enrolled represented additional conservation (additionality was ca. 13%; Bottazzi et al., 2018) and there were barriers to entry leading to higher uptake by households with formal land title, larger homes, cattle, and stronger social connections (Grillos, 2017). Uptake (percentage of a community area under *Watershared* agreements) was highly variable across the treated communities (Figure 1b), varying from 3 to 80% (median = 14%).

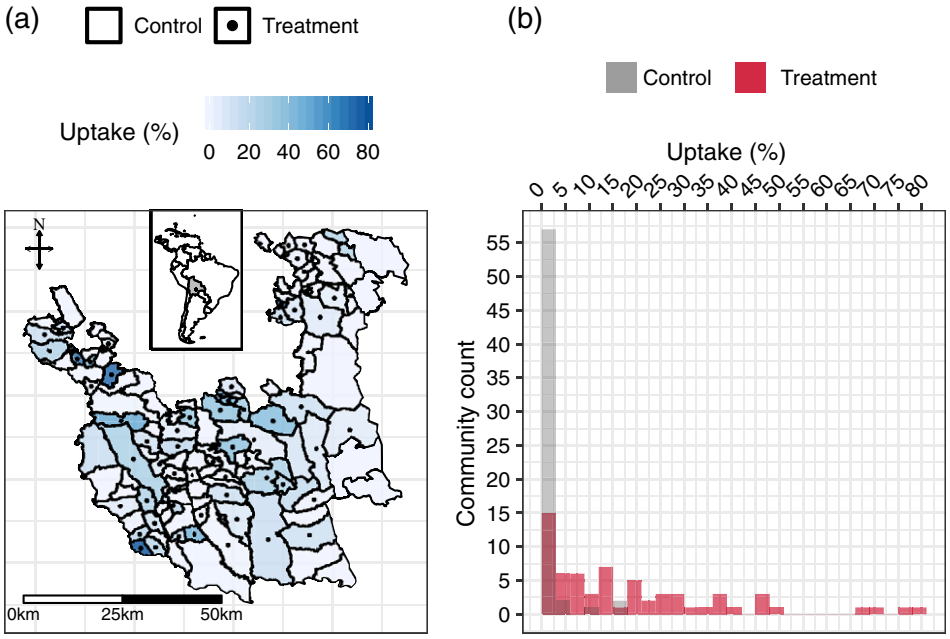

FIGURE 1 (a) Map of study area (the Río Grande Valles Cruceños Natural Integrated Management Area, Spanish acronym: ANMI) showing the uptake of the intervention (as % of community surface area). (b) The distribution of uptake between control ($n = 63$) and treatment ($n = 65$) communities

Consent to randomization was granted by community leaders in the area on the understanding that the intervention would subsequently be implemented in all communities (this general roll-out was conducted in 2016). The consent forms used in baseline and endline are archived alongside the data (Bottazzi et al., 2017). The endline social survey data used in part of this analysis was assessed under the Bangor University Research Ethics Framework.

2.2 | Deforestation data and data validation

Deforestation data were extracted from the GFC product (Hansen et al., 2013) that provides spatially explicit tree-cover percentage for 2000 and annual tree-cover change for 2000 to 2016. Thus, "Treecover2000" and "lossyear" layers were downloaded for tile 10S_070W and projected into UTM zone 20S. A threshold of 30% of tree cover was applied to generate a Forest/Non-Forest mask and then applied to the lossyear layer to select loss occurring on that mask only. The layers were combined into a deforestation map, with the resulting pixels classified into four groups: Forest stable; Non-Forest stable; Loss in the baseline period (2000–2010); Loss in the RCT period (2011–2016). This map was validated following (Olofsson et al., 2014) using visual checks on a stratified random sample of 426 points (see SI 3). Twenty-two points were excluded as poor-quality time series imagery made validation impossible. Accuracy of the remaining points ($n = 404$) was 94% (Table S3.1) with user's accuracy ranging from 63% (for the loss in the RCT period) to 97% (stable forest).

Spatial data processing (all code is available at https://github.com/lecrabe/res2_bolivia_rct) was implemented through R (R Core Team, 2017), OFGT (http://www.openforis.org/tools), and the GDAL library (http://www.gdal.org/). Accuracy assessment components (SI 3) were developed using dedicated tools (Olofsson et al., 2014).

2.3 | Statistical analysis

2.3.1 | Analytical approach

Although *Watershared* agreements are individual (a farmer will agree to enroll land or not), our deforestation analysis is at community level for three reasons. First, the randomization unit is the community. Second, although there are shapefiles for enrolled parcels, we do not have shapefiles for unenrolled parcels (either in control or in treatment communities), making finer-resolution comparison impossible. Finally, an analysis looking at whether deforestation was lower in enrolled parcels than other land would be highly vulnerable to confounding by on-farm leakage (Pfaff & Robalino, 2017).

Three further considerations impacted the analysis of *Watershared*'s effect on deforestation. First, while randomization exogenously allocated treatments, the voluntary nature of uptake yielded nonrandom variation in uptake. We controlled for factors that might influence both participation and the outcome as much as possible by controlling for uptake propensity (see below). Second, owing to this variation in uptake, there is a distinct difference between the randomization, which is binary (control/treatment), and the intervention, which is continuous (% area under agreements). We therefore have two models: an ITT model evaluating the effectiveness of *Watershared* as-implemented and a "continuous-treatment" (CT) model to explore the potential efficacy. Third, due to implementation error, a few households living in treatment communities enrolled land they own in control communities. Therefore there were control communities with enrolled land (see Figure 1b). We included all communities in our analysis despite this contamination of the control, accepting that it may introduce noise.

2.3.2 | Modeling the propensity for uptake of *Watershared* agreements

We modeled uptake propensity by regressing % land area enrolled in a treatment community against socioeconomic predictors aggregated to community scale as means (We also tested a model using medians, which explained 10% less variation; this is not shown). We selected predictors based on an analysis of household-level participation (Grillos, 2017), derived from a baseline survey by *Natura* in 2010 in all communities (Bottazzi et al., 2017). The predictors were: wealth (land, cattle available); education of household head (years); social embeddedness (generations a household has been present, frequency of involvement in community work); environmental attitudes (perceptions of forest value and local water quality); and remoteness (travel time to the nearest market—see SI 4 for more details). We used the predictions to create a propensity score for treatment and control communities, and used this score as a control variable in our deforestation analysis. One community which lacked baseline socioeconomic data and therefore a propensity score, had to be discarded from the analysis.

An important assumption in our deforestation models is that deviation from uptake propensity (i.e., uptake that cannot be explained by predicted uptake) is independent of confounding factors. For this to be the case, some of the unexplained variation in the uptake model would need to be related to variation affecting uptake but not deforestation. We suggest that such variation may be due to differences in how the offer of the program was experienced across the communities, for example by the timing of *Natura*'s visits to certain communities, the relationship between *Natura* technicians and communities, or the willingness of the community leader to spread the word about *Natura's* visit. We support our interpretation of our propensity model results using households' answers to the question (asked of those who did not take up agreements) "Why did you not join the scheme" ($n = 513$).

2.3.3 | Leakage

Leakage is a common concern in Payments for Ecosystem Services schemes as pressures may be displaced rather than eliminated (Alix-Garcia, Shapiro, & Sims, 2012; Börner et al., 2017). It is well known that leakage poses challenges for conservation evaluation (Pfaff & Robalino, 2017). As noted, we controlled for within-community leakage by analyzing deforestation at the community scale (as deforestation driven to areas near enrolled parcels would simply reduce our impact estimate). We could not control for between-community leakage, which, if preferentially occurring from treatment to control, would bias our estimated impact upward. However, we argue that such a bias is unlikely because treated communities' neighbors are randomized (thus the effect should cancel out). Also, local deforestation is mostly due to small-scale conversion to agriculture for local markets, so households are unlikely to clear land far from their home.

2.3.4 | Modeling details

Our primary, ITT analysis compared deforestation between treated and control communities regardless of the extent to which *Watershared* agreements were signed. This estimates the effectiveness of the *Watershared* intervention as rolled out in the region.

To explore the potential efficacy of the intervention, we further developed a "continuous treatment" (CT) model, which has some analogy to "as-treated" models commonly used in the medical trials literature. However, in our situation, treatment is continuous (% land area enrolled).

We followed published guidelines for analysis of intervention effects in randomized trials (European Medicines Agency, 2015). In addition to uptake propensity, we included as control variables those used for initial stratification of the control and treatment group (population and cattle density), the baseline value of our continuous outcome measure (deforestation 2000–2010), and other geographical variables expected a priori to be strongly associated with the outcome (limited using a screening model; SI 5).

All our models were fitted using GAM (Wood, 2011) to account for nonlinear relationships and nonnormal errors, leading to our use of the Tweedie distribution family for deforestation (percentage) and beta for uptake propensity (proportion), all selected based on a priori expectation combined with model comparisons for fit. The ITT and CT predictor set was identical apart from whether the intervention was coded as a binary control/treatment variable or % uptake across communities (see Table 1 for all predictors). In both cases, the intervention variable was interacted with uptake propensity, which would indicate whether treatment has an effect above and beyond the effect of endogenous factors. In other words, if there is no deforestation difference between control and treatment communities with high predicted uptake, it implies that the scheme has had no effect above the "null" behavior under predisposing conditions. Other plausible interactions between predictors were tested for significance and included where necessary. The effect on the impact evaluation exerted by data points with high leverage (Cook's distance) was evaluated by repeating analysis without them, which provided us with a more conservative estimate of the effect size of the PES scheme.

Significance of predictors as well as variable selection was determined using GAM internal Wald tests and by allowing for shrinkage (Wood, 2017). Model performance was examined by inspection of residuals (Faraway, 2006). The effect size of the intervention as per the CT model was approximated by predicting % deforestation in five scenarios where % uptake was set at 0, 20, 40, 60, and 80%. For each scenario, we made 30 plausible predictions of the effect of the intervention based on the model confidence interval. The percentages for each community were multiplied by its forest cover to attain deforestation in hectares, as well as an overall % change in deforested hectares out of available forest in 2010.

All statistical analysis was undertaken with R (R Core Team, 2017), using package mgcv (Wood, 2011, 2017). R

TABLE 1 Variables used to investigate *Watershared* intervention effect on deforestation rates in the intention-to-treat (ITT) and continuous-treatment (CT) models

Variable	Unit	Model	Inclusion reason	Data source
Deforestation (2011–2016)	%	Both	Response variable	Global Forest Change v1.4[a]
Baseline deforestation (2000–2010)	%	Both	Baseline data for response	Global Forest Change v1.4[a]
Population density	*n*/ha	Both	Stratification	Baseline data collected by *Natura* in 2010
Forest cover	%	Both	Hypothesis-based	Global Forest Change v1.4[a]
Community area	ha	Both	Control for large variability in community size	Community boundary shapefile, generated by our team from National Institute of Agrarian Reform data with field validation
Chaco (ecoregion)	%	Both	Screened	Ecoregion shapefile (see SI 1)
Slope	°	Both	Screened	USGS (2004)
Uptake propensity	%	Both	Endogeneity control	Modeled based on data in Bottazzi et al. (2017)
Control/treatment	0/1	ITT	Intervention	*Natura*
Uptake	%	CT	Intervention	Shapefile showing *Watershared* agreements provided by *Natura*

[a] Global Forest Change data was accessed at http://earthenginepartners.appspot.com/science-2013-global-forest

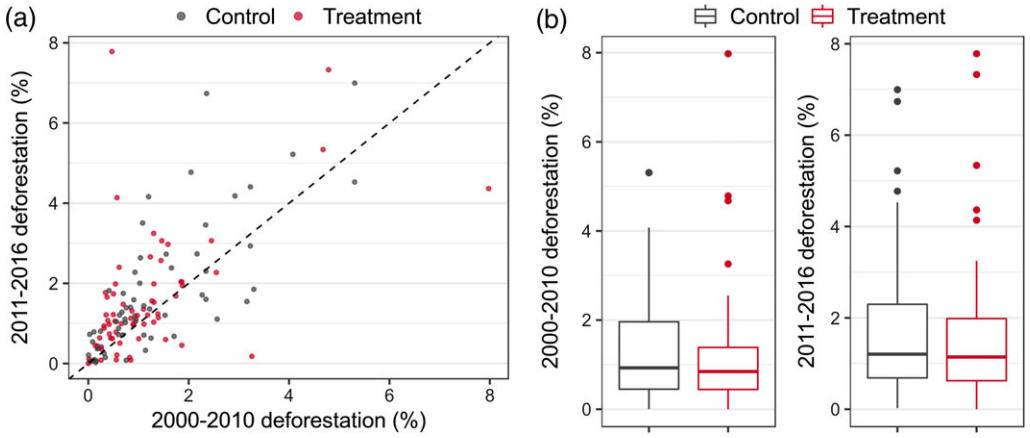

FIGURE 2 (a) Deforestation data for the baseline (2000–2010) and intervention (2011–2016) periods (dashed line = 1:1) and (b) their differences between control and treatment communities. Horizontal line = median, hinges = 25th and 75th percentiles, and whiskers extend to 1.5 × interquartile ranges from each hinge. Points represent data lying beyond this range

code with all data are available at [https://doi.org/10.6084/m9.figshare.7418264].

3 | RESULTS

3.1 | Distribution of, and trends in, deforestation

Total deforestation in the baseline period (2000–2010) was 4,147 ha (±742 ha) but was variable across communities (mean 1.2%, median 0.9%; Figure 2a). With the caveat that any systematic difference between randomized cohorts is necessarily due to chance and therefore invalidates the premise for frequentist significance tests, we note that there was no significant difference (Wilcoxon rank sum test) in either measure between control and treatment communities (Figure 2b), supporting the visual inspection of balance between control and treatment. The control and treatment communities were also largely balanced in the potential drivers of deforestation we identified (SI 6).

Communities with high baseline deforestation tended to also have high deforestation during the intervention period, however there was considerable scatter around this relationship (Figure 2a). The total area of deforestation during the intervention (2011–2016) was 6,042 ha (±3,933 ha); again, this was variable across communities (mean 1.7%, median of 1.2%). Considering the intervention period is shorter, this implies increased overall deforestation in the intervention period.

3.2 | Modeling propensity of uptake in *Watershared* agreements

Our model of uptake propensity explained 50% of uptake (Figure S4.1a), with considerable and slightly biased scatter around the 1:1 line. Testing different model family specifications and predictor interactions did not improve fit, suggesting omitted variable bias; however control and treatment communities were largely balanced in uptake propensity (Figure S4.1b in SI4).

Responses to the question "Why did you not join the scheme," provide some evidence that the unexplained variation in uptake can be explained, at least in part, by nonconfounding factors. Not having attended a sign-up meeting was the most common reason (50%) given by nonparticipants instead of for example, lack of interest (SI 7). While not attending a meeting may be correlated with some confounders, it could also reflect variation in the way in which the program was offered across the study area.

3.3 | ITT model (79.7% deviance explained—see SI 8)

ITT analysis revealed no significant difference overall (i.e., intercept) between deforestation in control and treatment communities after accounting for control variables including uptake propensity (Figure 3). The slope of uptake propensity varied between control and treatment: uptake propensity was significant for the treatment communities but insignificant for control communities ($p = 0.016$ vs. $p = 0.11$). For treatment communities, the relationship suggested decreasing deforestation with increasing uptake propensity (Figure 4a). However, following removal of data points with high leverage on model outcomes ($n = 3$), there was no significant control/treatment difference in the relationship between deforestation and uptake propensity and therefore the effect is volatile (Figure 4b).

3.4 | Continuous-treatment model (80.3% deviance explained—see SI 8)

The continuous-treatment model indicated a significant negative relationship between both increasing % uptake and % uptake propensity (as interaction) and deforestation ($p = 0.008$; Figure 5a; SI 8). For this model, removing the data points with high leverage on model outputs did not remove the treatment effect (Figure 5b). The treatment effect is small. If an 80% uptake were achieved our models suggest

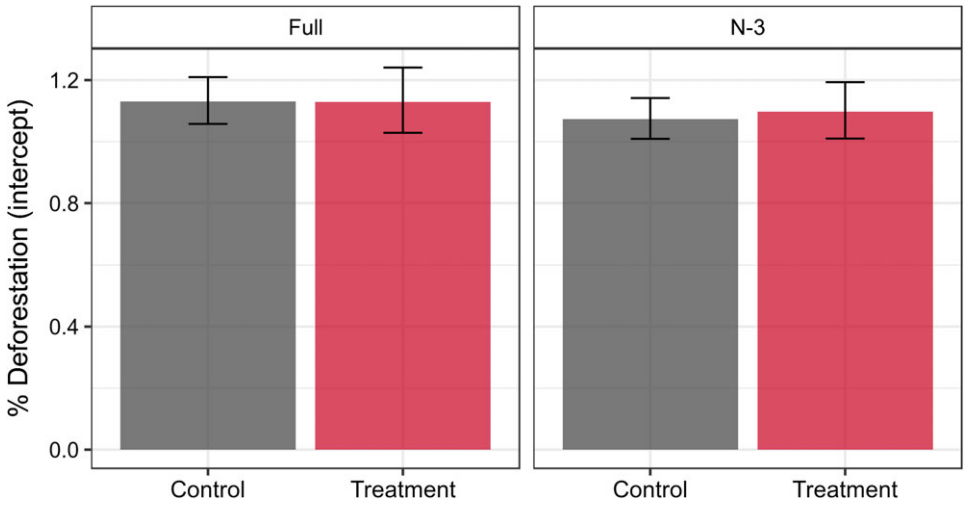

FIGURE 3 Difference in deforestation between control and treatment communities based on the intention-to-treat model with (full) and without ($N - 3$) communities with high leverage on model results. Bars are standard errors

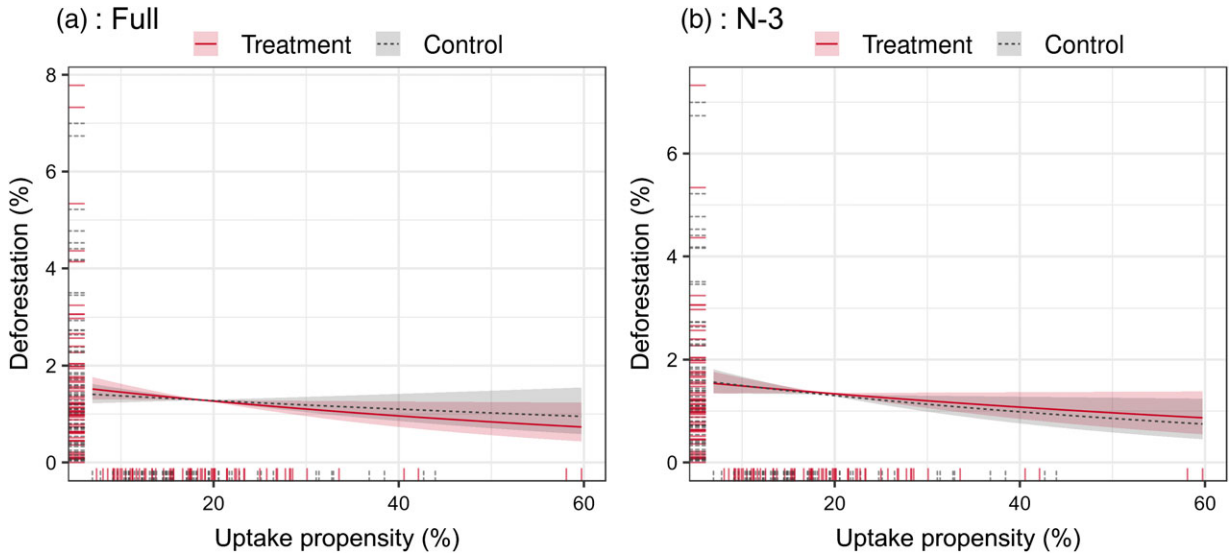

FIGURE 4 The intention-to-treat model suggests decreasing deforestation with increasing uptake propensity for treatment but not control communities when all communities are included (a). However when three communities with high leverage on model results are discarded, there is no difference between control and treatment cohorts (b). The rug shows the distribution of real data points

this would result in a reduction of deforestation of just 670 ha compared with 0% uptake. This represents a reduction in deforestation rate from 1.58 ± 0.015% (1 SD) (with 0% uptake scenario) to 1.41 ± 0.042 (1 SD).

4 | DISCUSSION

4.1 | Did the *Watershared* intervention effectively slow deforestation?

We did not detect a measurable impact of *Watershared* on deforestation using the ITT model. This suggests that, as implemented in the landscape, *Watershared* was not effective at slowing deforestation. While there is some evidence that deforestation was reduced for communities with a higher propensity to take up the scheme in treatment communities (but not in control communities; Figure 4), this effect was driven by three communities with high leverage. Our exploration of efficacy (CT model) showed that deforestation decreased slightly with increasing uptake regardless of uptake propensity, which suggests that improvement of uptake rates could potentially lead to effective intervention.

Interpretation of our CT model rests on the assumption that deviation from intended treatment (both uptake in control communities, and some element of variation of uptake in treated communities) was independent of confounding factors (McNamee, 2009). We are confident that confounding factors did not drive the cases of uptake in control communities; they were the result of an accident of geography (people living in treatment communities who owned land in control communities) and limited monitoring of the RCT (they should not have been allowed to enroll that land). The 50% of uptake variation

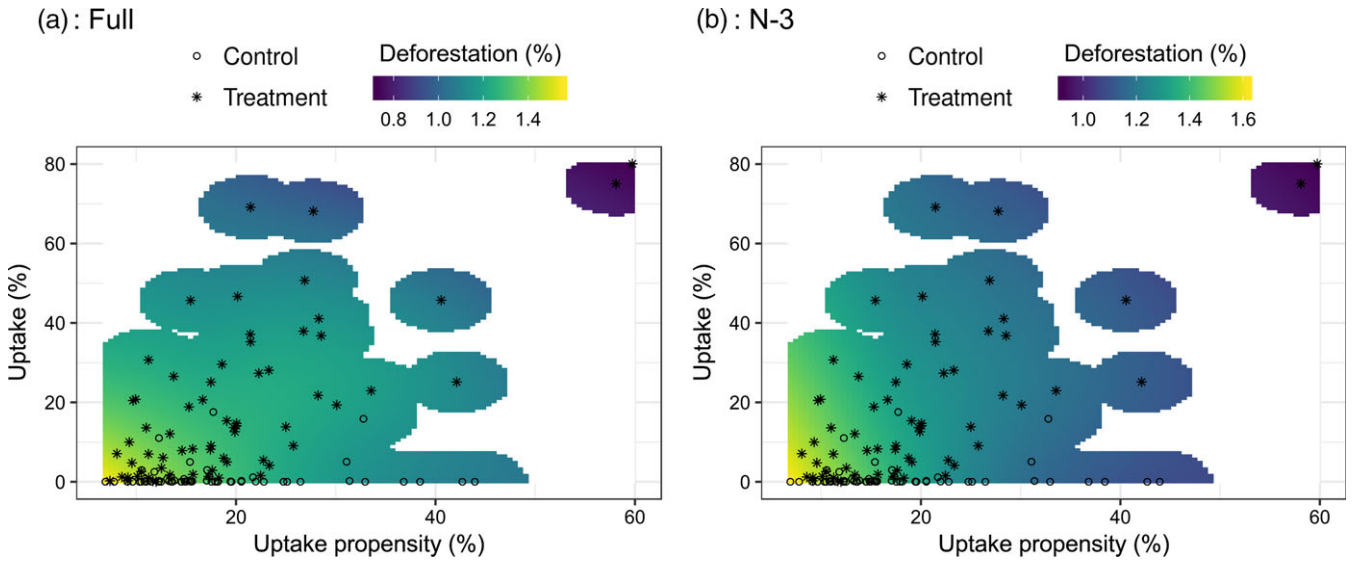

FIGURE 5 The results of the continuous-treatment model with (a) and without (b) three communities with high leverage on model results. The estimated effect of uptake can be seen by comparing deforestation (color scale) between communities with similar uptake propensity (*x*-axis) but different actual uptake (*y*-axis)

we could not explain with our propensity score, while potentially influenced by unobserved confounding factors, is also plausibly due to differences in the way the scheme was offered between communities. Opportunities to enroll land may have varied because of timing of visits by *Natura*, or links between technicians and community members affecting how effectively news of the meetings spread in some communities, or chance (people being sick, or away). These possibilities, although not directly monitored as part of this RCT, are supported by our interviews with nonparticipants. Fifty percent of respondents gave "did not attend meeting" as the reason for not taking up the agreements.

The randomization increases confidence in our analysis. Given that uptake propensity scores and preintervention deforestation rates, inter alia, were balanced between control and treatment communities, we can reasonably expect balance also in unobservable confounders. For example, treatment communities with both high uptake propensity and high uptake, can be expected to be balanced in the analysis with similar control communities who would have taken up the scheme if offered it. In the absence of being able to perfectly model propensity to take up the scheme, randomization was therefore very useful for supporting causal inference.

It is important to note that our estimated effect is very small, and potentially trivial. If 80% uptake was achieved across the landscape (unlikely to be achievable), our scenario modeling suggests that deforestation would reduce from 1.58% (with 0% uptake scenario) to 1.41%. The only other published RCT evaluation of a PES program looked at the impact of payments to households in Uganda over a 2-year period (Jayachandran et al., 2017) and found a much larger reduction in deforestation rate (from 9.1% to just 4.2%). However, this project operated in a small area (<99,300 ha vs. 489,400 ha here) with higher deforestation rates. Low baseline deforestation inevitably reduces the scope of impacts of a program seeking to reduce deforestation rates (Alix-Garcia et al., 2012).

4.2 | Might the *Watershared* program have had other environmental impacts?

The *Watershared* program was introduced with the aim not only of conserving forest cover, but also conserving biodiversity (potentially damaged by forest degradation) and ensuring the supply of locally valued ecosystem services (particularly the quality and quantity of downstream water; Asquith, 2016). A recent analysis of the impact of the *Watershared* scheme on water quality using the same RCT design as in this study showed that while excluding cattle from water sources reduced *Escherichia coli* contamination at that location, there was no difference between control and treatment communities in the quality of their water (Pynegar et al., 2018). Pynegar et al. (2018) suggest that the lack of impact on water quality is because so little land was enrolled in level 1 contracts, and the scheme involved no targeting meaning that not all the land enrolled had the potential to impact water quality. It is possible that the scheme may have had a positive impact on local biodiversity through cattle exclusion. However, although detailed data on amphibians, reptiles and dung beetles were collected at endline, this data has not yet been examined.

4.3 | How could the impact of the *Watershared* program be increased?

Watershared already fulfils some criteria recently identified as correlating with PES success (Börner et al., 2017) such as compliance monitoring and in-kind payments. However, it

did not deliver reductions in deforestation with the levels of uptake which were achieved in the study area. We mention above that low uptake in some communities may have been driven at least partly by differences in how the intervention was presented between communities, which could be beneficial to explore in the future.

The value of the in-kind incentives is also likely to have a role in uptake. While it is difficult to draw comparisons across countries with different economies, the value of the incentives in *Watershared* are low compared with other program (the value of the incentives for the most restricting agreements is $10 a hectare plus the equivalent of a $100 value joining bonus, but just $1 a hectare for the least restricting agreements; SI 2). For comparison, Mexico's program pays 27–36 USD ha^{-1} $year^{-1}$ depending on forest type (Muñoz-Piña, Guevara, Torres, & Braña, 2008), Costa Rica's national program pays 45–163 USD ha^{-1} $year^{-1}$ (Wunder, Engel, & Pagiola, 2008), and the Ugandan PES program paid 28 USD ha^{-1} $year^{-1}$ (Jayachandran et al., 2017). Those promoting *Watershared* argue that it works through nudging, by emphasizing environmental norms and reciprocity rather than paying the opportunity cost, so the level of incentives is relatively unimportant (Asquith, 2016). There is evidence that farmers enroll due to the perception that they or their community will benefit from improved water quality (Bottazzi et al., 2018). However both theory (Persson & Alpízar, 2013) and empirical data (Arriagada, Sills, Pattanayak, & Ferraro, 2009) do predict low incentives lead to low participation. We suggest that higher valued incentives could increase uptake of *Watershared.*

Our evidence suggests that even if uptake could be greatly increased, the reduction in deforestation would be modest. A common problem in all PES schemes is adverse selection; participants enroll land which is unlikely to be cleared anyway, resulting in low additionality (Börner et al., 2017). A recent analysis suggests that only 13% of the land area enrolled in *Watershared* agreements has resulted in additional conservation (Bottazzi et al., 2018). If higher payments could increase additionality as well as uptake this may therefore increase the efficacy of the intervention.

We finally note that the impact of *Watershared* may also increase and/or materialize with time, as found for a number of PES schemes (Grima, Singh, Smetschka, & Ringhofer, 2016), especially where livelihood changes are incentivized (Börner et al., 2017). For example, many of the *Watershared* incentives involve either waiting (fruit tree saplings reaching maturation) or mastery (bee keeping, effective irrigation) before becoming a financially viable alternative to the status quo.

4.4 | What can RCT contribute to conservation impact evaluation?

Establishing causality in environmental policies by properly identifying counterfactual outcomes is essential if environmental policy decisions are to be based on evidence (Ferraro & Hanauer, 2014). Quasi-experimental approaches represent a huge advance over what passed for conservation evaluations in the past, and their increasing use is very positive. However, post-hoc analysis is only as reliable as the counterfactual scenario which can be created statistically and recent evidence demonstrates how even supposedly robust methods such as difference-in-differences can result in biases in impact estimates (Daw & Hatfield, 2018). As much as possible, therefore, conservation interventions should be explicitly designed to allow robust evaluation (Ferraro & Hanauer, 2014). Randomizing a conservation intervention can help to facilitate an evaluation by reducing the role of confounding factors, as well as providing a satisfactory pool of counterfactuals in cases of nonrandom uptake.

The *Watershared* RCT suffered from some contamination of the control and considerable variability in uptake. Despite this "noise", the randomized design was an improvement from a nonrandomized alternative. This is because unobserved confounders driving uptake are likely to exist, which quasi-experimental methods such as matching cannot account for. The existence of a control balanced in all factors for which we have data gives us confidence that the observed effect (or lack of) is not due to these unobserved confounders. For example, there were low uptake rates in the northern sector which would not have been expected a priori, however randomization ensured that comparable controls existed.

Despite calls for more randomized experiments in conservation impact evaluation, their use remains rare. *Watershared* is one of only three randomized impact evaluations of landscape-scale conservation interventions we are aware of (the others are: Jayachandran et al., 2017; Wilebore, Voors, Bulte, Coomes, & Kontoleon, in press). There are ethical and practical challenges meaning that full RCTs are not always appropriate (Baylis et al., 2016; Deaton & Cartwright, 2018; Ferraro, 2011; Pynegar et al., 2018). However, where possible, randomization certainly offers valuable opportunities for improving causal inference (Ferraro, 2011). The *Watershared* RCT is the result of a collaboration between practitioners, who had the foresight to implement their intervention in a randomized design, and researchers. More such collaborations would facilitate a growth in the robust evaluations that conservation so desperately needs. We hope that conservation can avoid the polarized debate surrounding the value of knowledge generated from RCTs in other fields (Ravallion, 2009), and that randomization can be added to the conservationist toolkit where appropriate.

ACKNOWLEDGMENTS

We thank colleagues at Fundación Natura Bolivia especially Maria Teresa Vargas, Tito Vidaurre, Johnny Severiche, Roger Coronado and Victoria Aguilera, and Patrick Bottazzi.

James Gibbons, Gavin Simpson, Alex Pfaff, and Paul Ferraro provided valued input on analytical approaches, giving very generously of their time. We are also grateful to Kelsey Jack who designed the initial randomization in 2010. This research was funded by grant RPG-2014-056 from the Leverhulme Trust and grant NE/L001470/1 from the UK's Ecosystem Services and Poverty Alleviation program.

CONFLICT OF INTEREST

There is the potential for conflict of interest as *Natura* were involved in the research but are also the implementers of the *Watershared* program which is the focus of our research. However, while N.A. (who was involved in founding *Natura*) is a co-author, he was not directly involved in the analysis (his role was providing context for helping us design the analysis and interpret results). E.P. started working for *Natura* after this analysis was complete.

Author contributions

E.W. and J.P.G.J. conceived the analysis with input from the other authors. R.D. developed the forest change product and conducted the validation (with help from Crespo). E.W. and J.P.G.J. wrote the paper. N.A. developed the randomization which made the analysis possible.

ORCID

Julia P. G. Jones https://orcid.org/0000-0002-5199-3335

REFERENCES

Alix-Garcia, J. M., Mcintosh, C., Sims, K. R. E., & Welch, J. R. (2013). The ecological footprint of poverty alleviation: Evidence from Mexico's Oportunidades program. *The Review of Economics and Statistics*, *95*(2), 417–435.

Alix-Garcia, J. M., Shapiro, E. N., & Sims, K. R. E. (2012). Forest conservation and slippage: Evidence from Mexico's National Payments for Ecosystem Services Program. *Land Economics*, *88*(4), 613–638. https://doi.org/10.1353/lde.2012.0052

Alix-Garcia, J. M., Sims, K. R. E., Orozco-Olvera, V. H., Costica, L. E., Fernández Medina, J. D., & Romo Monroy, S. (2018). Payments for environmental services supported social capital while increasing land management. *Proceedings of the National Academy of Sciences of the United States*, *115*, 7016–7021.

Arriagada, R. A., Sills, E. O., Pattanayak, S. K., & Ferraro, P. J. (2009). Combining qualitative and quantitative methods to evaluate participation in Costa Rica's program of payments for environmental services. *Journal of Sustainable Forestry*, *28*(3–5), 343–367.

Asquith, N. M. (2016). Watershared: Adaptation, mitigation, watershed protection and economic development in Latin America. *Climate & Development Knowledge Network*, *1*, 1–7.

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., … Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, *9*(1), 58–64. https://doi.org/10.1111/conl.12180

Börner, J., Baylis, K., Corbera, E., Ezzine-de-Blas, D., Honey-Rosés, J., Persson, U. M., & Wunder, S. (2017). The effectiveness of payments for environmental services. *World Development*, *96*, 359–374. https://doi.org/10.1016/j.worlddev.2017.03.020

Bottazzi, P., Vargas Rios, M. T., Asquith, N., Crespo, D., Vidaurre, T., García, M., … Jones, J. P. G. (2017). Baseline and endline socio-economic data from a randomised control trial of the Watershared intervention in the Bolivian Andes. *ReShare UKData Archive*. https://doi.org/10.5255/UKDA-SN-852623

Bottazzi, P., Wiik, E., Crespo, D., & Jones, J. P. G. (2018). Payment for environmental "self-service": Exploring the links between Farmers' motivation and additionality in a conservation incentive programme in the Bolivian Andes. *Ecological Economics*, *150*, 11–23. https://doi.org/10.1016/j.ecolecon.2018.03.032

Clements, T., & Milner-Gulland, E. J. (2015). Impact of payments for environmental services and protected areas on local livelihoods and forest conservation in northern Cambodia. *Conservation Biology*, *29*(1), 78–87. https://doi.org/10.1111/cobi.12423

Daw, J. R., & Hatfield, L. A. (2018). Matching and regression to the mean in difference-in-differences analysis. *Health Services Research*. https://doi.org/10.1111/1475-6773.12993

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Eklund, J., Guillaume Blanchet, F., Nyman, J., Rocha, R., Virtanen, T., & Cabeza, M. (2016). Contrasting spatial and temporal trends of protected area effectiveness in mitigating deforestation in Madagascar. *Biological Conservation*, *203*, 290–297. https://doi.org/10.1016/j.biocon.2016.09.033

European Medicines Agency. (2015). *Guideline on adjustment for baseline covariates in clinical trials*. London, England: Author.

Faraway, J. J. (2006). Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models. In *CRC texts in statistical science series* (Vol. 62, 2nd ed.). Boca Raton, FL: Chapman and Hall. https://doi.org/10.1111/j.1541-0420.2006.00596_12.x

Ferraro, P. J. (2011). *Experimental project designs in the global environment facility designing projects to create evidence and catalyze investments to secure global environmental benefits.* Retrieved from http://www.unep.org/stap/

Ferraro, P. J., & Hanauer, M. M. (2014). Advances in measuring the environmental and social impacts of environmental programs. *Annual Review of Environment and Resources*, *39*(1), 495–517. https://doi.org/10.1146/annurev-environ-101813-013230

Ferraro, P. J., & Pattanayak, S. K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, *4*(4), e105. https://doi.org/10.1371/journal.pbio.0040105

Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton, NJ: Princeton University Press.

Grillos, T. (2017). Economic vs non-material incentives for participation in an in-kind payments for ecosystem services program in Bolivia. *Ecological Economics*, *131*, 178–190. https://doi.org/10.1016/j.ecolecon.2016.08.010

Grima, N., Singh, S. J., Smetschka, B., & Ringhofer, L. (2016). Payment for ecosystem services (PES) in Latin America: Analysing the performance of 40 case studies. *Ecosystem Services*, *17*, 24–32. https://doi.org/10.1016/j.ecoser.2015.11.010

Gupta, S. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109–112. https://doi.org/10.4103/2229-3485.83221

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Jayachandran, S., de Laat, J., Lambin, E. F., Stanton, C. Y., Audy, R., & Thomas, N. E. (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, *357*(6348), 267–273. https://doi.org/10.1126/science.aan0568

Landell-Mills, N., & Porras, I. T. (2002). *Silver bullet or fools' gold? A global review of markets for forest environmental services and their impact on the poor*. London, England: NEPAD.

McNamee, R. (2009). Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Statistics in Medecine*, *28*, 2639–2652.

Miteva, D. A., Pattanayak, S. K., & Ferraro, P. J. (2012). Evaluation of biodiversity policy instruments: What works and what doesn't? *Oxford Review of Economic Policy*, *28*(1), 69–92. https://doi.org/10.1093/oxrep/grs009

Muñoz-Piña, C., Guevara, A., Torres, J. M., & Braña, J. (2008). Paying for the hydrological services of Mexico's forests: Analysis, negotiations and results. *Ecological Economics*, *65*(4), 725–736. https://doi.org/10.1016/j.ecolecon.2007.07.031

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

Pattanayak, S. K., Wunder, S., & Ferraro, P. J. (2010). Show me the money: Do payments supply environmental services in developing countries? *Review of Environmental Economics and Policy*, *4*(2), 254–274. https://doi.org/10.1093/reep/req006

Persson, U. M., & Alpízar, F. (2013). Conditional cash transfers and payments for environmental services—A conceptual framework for explaining and judging differences in outcomes. *World Development*, *43*, 124–137. https://doi.org/10.1016/j.worlddev.2012.10.006

Pfaff, A., & Robalino, J. (2017). Spillovers from conservation programs. *Annual Review of Resource Economics*, *9*(1), 299–315. https://doi.org/10.1146/annurev-resource-100516-053543

Pullin, A. S., & Knight, T. M. (2001). Effectiveness in conservation practice: Pointers from medicine and public health. *Conservation Biology*, *15*, 50–54. https://doi.org/10.1046/j.1523-1739.2001.99499.x

Pynegar, E. L., Jones, J. P. G., Gibbons, J. M., & Asquith, N. M. (2018). The effectiveness of payments for ecosystem services at delivering improvements in water quality: Lessons for experiments at the landscape scale. *PeerJ*, *6*, e5753. https://doi.org/10.7717/peerj.5753

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rasolofoson, R. A., Ferraro, P. J., Jenkins, C. N., & Jones, J. P. G. (2015). Effectiveness of community forest management at reducing deforestation in Madagascar. *Biological Conservation*, *184*, 271–277. https://doi.org/10.1016/j.biocon.2015.01.027

Ravallion, M. (2009). Should the randomistas rule? *The Economists' Voice*, *6*(2), 8–12. https://doi.org/10.2202/1553-3832.1368

Robalino, J., & Pfaff, A. (2013). Ecopayments and deforestation in Costa Rica: A nationwide analysis of PSA's initial years. *Land Economics*, *89*(3), 432–448. https://doi.org/10.3368/le.89.3.432

Salzman, J., Bennett, G., Carroll, N., Goldstein, A., & Jenkins, M. (2018). The global status and trends of payments for ecosystem services. *Nature Sustainability*, *1*(3), 136–144. https://doi.org/10.1038/s41893-018-0033-0

Samii, C., Lisiecki, M., Kulkarni, P., Paler, L., & Chavis, L. (2014). Effects of payment for environmental services (PES) on deforestation and poverty in low and middle income countries: A systematic review. *Campbell Systematic Reviews*, (11), 10. https://doi.org/10.4073/csr.2014.11

Sánchez-Azofeifa, G. A., Pfaff, A., Robalino, J. A., & Boomhower, J. P. (2007). Costa Rica's payment for environmental services program: Intention, implementation, and impact. *Conservation Biology*, *21*(5), 1165–1173. https://doi.org/10.1111/j.1523-1739.2007.00751.x

Sills, E. O., de Sassi, C., Jagger, P., Lawlor, K., Miteva, D. A., Pattanayak, S. K., & Sunderlin, W. D. (2017). Building the evidence base for REDD+: Study design and methods for evaluating the impacts of conservation interventions on local well-being. *Global Environmental Change*, *43*, 148–160. https://doi.org/10.1016/j.gloenvcha.2017.02.002

Sims, K. R. E. (2010). Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, *60*, 94–114. https://doi.org/10.1016/j.jeem.2010.05.003

Ten Have, T. R., Normand, S.-L. T., Marcus, S. M., Brown, C. H., Lavori, P., & Duan, N. (2008). Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatric Annals*, *38*(12), 772–783. https://doi.org/10.3928/00485713-20081201-10

USGS. (2004). *Shuttle radar topography mission, 1 arc second scene SRTM_u03_n008e004, unfilled unfinished 2.0, February 2000*. College Park, MD: Global Land Cover Facility, University of Maryland.

Wilebore, B., Voors, M., Bulte, E., Coomes, D., & Kontoleon, A. (in press). Unconditional transfers and tropical forest conservation. Evidence from a randomized control trial in Sierra Leone. *American Journal of Agricultural Economics*.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton, FL: CRC Press.

Wunder, S. (2007). The efficiency of payments for environmental services in tropical conservation. *Conservation Biology*, *21*(1), 48–58. https://doi.org/10.1111/j.1523-1739.2006.00559.x

Wunder, S. (2015). Revisiting the concept of payments for environmental services. *Ecological Economics*, *117*, 234–243.

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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wiik E, d'Annunzio R, Pynegar E, Crespo D, Asquith N, Jones JPG. Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conservation Science and Practice*. 2019;e8. https://doi.org/10.1002/csp2.8