

<https://doi.org/10.1038/s43247-024-01541-1>

Modest forest and welfare gains from initiatives for reduced emissions from deforestation and forest degradation

Check for updates

Sven Wunder^{1,2} [✉](mailto:sven.wunder@efi.int), Dario Schulz^{1,3}, Javier G. Montoya-Zumaeta^{4,5}, Jan Börner³, Gabriel Ponzoni Frey³ & Bibiana Betancur-Corredor³

Reduced emissions from deforestation and forest degradation (REDD+) projects and programmes promise to deliver performance-based, cost-effective climate change mitigation. Fifteen years after its conception, we analysed the rigorous counterfactual-based evidence for environmental and welfare effects from such national and subnational initiatives, along with a Theory of Change. Using machine-learning tools for literature review, we compared 32 quantitative studies including 26 primary forest-related and 12 socioeconomic effect sizes. Average environmental impacts were positively significant yet moderately sized, comparable to impacts from other conservation tools, and mostly impermanent over time. Socioeconomic impacts were welfare-neutral to slightly positive. Moderator analysis showed that environmental additionality was likely restricted by project proponents' adverse spatial targeting of low-threat areas. Scarce funding flows from carbon markets and ill-enforced conditionality probably also limited impacts. Hence, important policy and implementation lessons emerge for boosting effectiveness in the current global transition towards larger-scale, jurisdictional action.

In 2003, at the United Nations Framework Convention on Climate Change (UNFCCC) 9th Conference of the Parties, researchers from Brazil and the USA launched the notion of compensated reduction: tropical countries should be ex-post rewarded for reducing their national forest loss below a pre-agreed baseline¹. The European Union's (EU) Joint Research Centre also linked national forest-cover baselines to possible compensations between countries². Simultaneously, a high-level review of the economics of climate change concluded that for US\$5–10 billion, two-thirds of global deforestation could be 'bought out', thus curbing one major source of global greenhouse gas emissions at low costs³. Since much land clearing in the forested hinterlands of the Global South provides only marginal economic returns, conservation opportunity costs there often remain modest, so purportedly they could be compensated rather cheaply. The conceptual scope was later broadened towards an all-inclusive term of political consensus: "reduced emissions from deforestation, forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries" (short: REDD+)⁴.

REDD+ would allegedly work as an international multilevel system of conditional, performance-based payments for environmental services (PES)^{5–7}. In this global architecture, carbon markets would mobilise funding, while recipient national governments would incentivize on-the-ground

landholders and forest-dwelling indigenous populations, invest in economic alternatives, capacity building, and improve protected areas—thus delivering the enabling conditions for achieving emissions reductions on the ground⁸. Hence, REDD+ as a model of intervention is usually associated with global performance-based forest-carbon funding, but implementation is de facto an umbrella term for a broad mix of ground-level initiatives, designed in contextually customised ways.

A swath of local-level REDD+ projects has been implemented since the COP13 Bali Action Plan in 2007^{9,10}. Across governance scales, nested approaches were proposed to resolve issues of attributing carbon credits between projects, subnational programmes, and the national level, including to avoid double-counting¹¹. REDD+ came out strengthened by the UNFCCC Paris Agreement (UNFCCC 2015). Based on CIFOR's International Database on REDD+ Projects (ID-RECCO), 377 REDD+ projects cover 53 million hectares in 56 countries¹². These projects were to avoid some 1% of annual forest-based emissions¹⁰.

Across the Global South, Brazil (48), Colombia (33), and Peru (25) have most initiatives; the density of REDD+ implementation (project area/national forest area) is highest in Kenya, Nepal, Central America, and the Andes region (Fig. 1); conversely, REDD+ implementation density is lower in Central Africa and South Asia. Our map also clearly depicts the nesting

¹European Forest Institute (EFI), Barcelona, Spain. ²Center for International Forestry Research (CIFOR), Lima, Peru. ³Institute for Food and Resource Economics & Center for Development Research, University of Bonn, Bonn, Germany. ⁴Centre for Development and Environment (CDE), University of Bern, Bern, Switzerland.

⁵World Agroforestry Center (ICRAF), Lima, Peru. e-mail: sven.wunder@efi.int

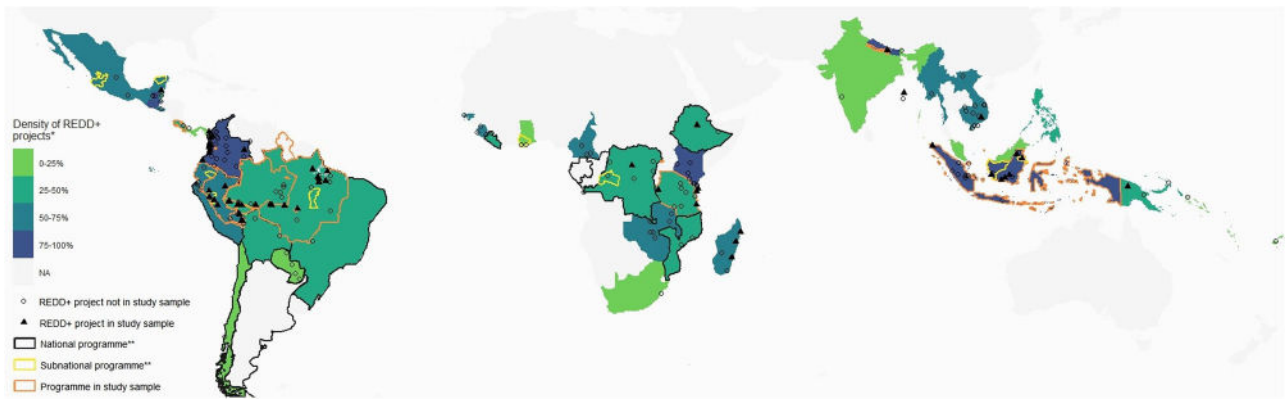


Fig. 1 | Mapping REDD+ projects, programmes, and study sample: implementation and research densities both vary strongly in space. * Density is measured by a country's aggregated REDD+ project area, divided by national forest area in 2020. ** Jurisdictional programmes considered: NICFI (Norway), Central

African Forest Initiative (CAFI), REDD Early Movers (Germany), FCPF (World Bank), Green Climate Fund (GCF), Governor's Climate and Forest Task Force (GCF-TF). For sources and assumptions, see SI.

challenges of REDD+ credits between overlapping national, subnational, and project scales of action (e.g., in Brazil). Finally, mapping the interventions evaluated by at least one of our included studies (triangles; for sample selection, cf. Methods—case sources are listed in the Supplementary Information, SI) also points to some research imbalances: compared to REDD+ implementation, only few rigorous impact evaluations have been done in Asia, Africa, Colombia, and Mexico.

Rather than carbon markets assuming the lead role in financing large-scale jurisdictional-level policy-related implementation, REDD+ has remained 'project-ified', with bilateral or conservation donors financing only incipient actions¹³. Bilateral donors and the UN-REDD programme have implemented so-called REDD+ readiness preparatory processes (increasing forest monitoring capacities, analysing deforestation drivers, etc.); 50+ countries have created national REDD+ programmes¹⁴ (Fig. 1).

The originally envisaged model of REDD+ large-scale national implementation has in practice only in a few countries advanced towards large-scale conditional payments. Those comprise notably Norway's International Climate and Forest Initiative (NICFI), launched as early as 2007⁵, followed by Germany's REDD Early Movers programme¹⁵. More recently, two multilateral organisations started piloting large-scale, results-based payments: the Green Climate Fund (GCF) and the Forest Carbon Partnership Facility (FCPF), as well as efforts to mobilise public-private blended funding under the LEAF coalition.

A heated debate has accompanied REDD+ projects. Strong criticisms of REDD+ processes and impacts have focused *inter alia* on problems related to social inclusion, indigenous rights, and other welfare outcomes^{16–19}. Conversely, more optimistic outlooks stressed the experimental nature of project-scale REDD+ initiatives (with some encouraging outcomes), the time-consuming complexity of governance transitions, and the embryonic stage of the genuine national-level REDD+ that largely remains untested^{14,20,21}.

At this stage, how much do we know about the on-the-ground successes or failures of REDD+? Impact evaluations should answer this question; they are becoming standard tools in many sectors²². Following emphatic calls for solid empirical impact assessments also in environmental and biodiversity conservation²³, the field of environmental impact evaluation has recently expanded, as evidenced by various reviews and meta-studies^{24–26}.

For REDD+ funded initiatives, so far less stylised evidence is available. The literature-synthesising work has been mostly qualitative, ranging from reviews of the REDD+ literature⁴, its governance challenges²⁷ and perspectives for future REDD+ implementation¹⁴. More formal evaluations include lessons from early carbon projects²⁸, and early REDD+ environmental and welfare effects²⁹.

Against this backdrop, our REDD+ meta-study systematically takes stock of the currently expanding evidence, which required careful delimitations (cf. Methods). In our Web-of-Knowledge-based literature search using text-mining algorithms, we target rigorous quantitative evaluations of the environmental and welfare-related impacts of REDD+ interventions. This includes (corporate or NGO) projects, public programmes (e.g., national payments for environmental services (PES) schemes with forest-carbon components), and a few bilateral jurisdictional agreements (e.g. Norway's forest agreements with Guyana and Indonesia). Our focus is on avoided deforestation and degradation, rather than re-, afforestation, or restoration. Included evaluation studies contain impact estimates that can be scaled and ranked, i.e., effect sizes (and their precision) are comparable including across categories defined by relevant contextual and design variables. Finally, as often called for³⁰, we compare REDD+ impact sizes to those of other forest conservation instruments. To our knowledge, no such analysis exists in the REDD+ literature.

Results

Environmental impacts

We used the correlated hierarchical effects model with random effects for our impact calculations. In total, 32 quantitative studies (listed in SI) with 26 forest-related and 12 socioeconomic primary effect sizes fulfilled our data selection criteria (cf. Methods). Figure 2 shows a forest plot of our meta-regression results for comparable forest impact sizes from REDD+ treatments (results fully reproduced in SI). We only have one (not statistically significant) estimate for forest carbon (Fig. 2)—the primary goal and final impact of REDD+ according to its Theory of Change (cf. SI, Fig. S2). Most estimates were for forest-cover proxy outcomes (including absolute and relative forest loss) leading to these impacts, which can be more easily measured. Our mean overall estimated REDD+ effect size of 0.08 (95% Confidence Interval: 0.04–0.11) can comparatively speaking be considered small. The Q-test indicated heterogeneity, meaning one could find true effects outside of this confidence interval. Yet, the positive significant estimate confirmed modest forest conservation gains from REDD+.

Two intervention subgroups can be distinguished in Fig. 2: self-declared REDD+ projects (commercial, NGO-led, or national—cf. upper panel) versus carbon-inclusive multipurpose conservation PES (public, mostly national programmes schemes—middle panel). We found no significant effect difference between the two ($p = 0.57$). The precision of estimates was lower among some of the smaller-sized projects. Even in public PES-for-carbon schemes, the same programmes evaluated in different studies reached considerably divergent estimates—including studies carried out by the same authors (i.e. Arriagada et al. 2011 vs. 2012 both on Costa Rica's PSA; Mohebalian and Aguilar 2016 vs. 2018, both on Ecuador's Socio

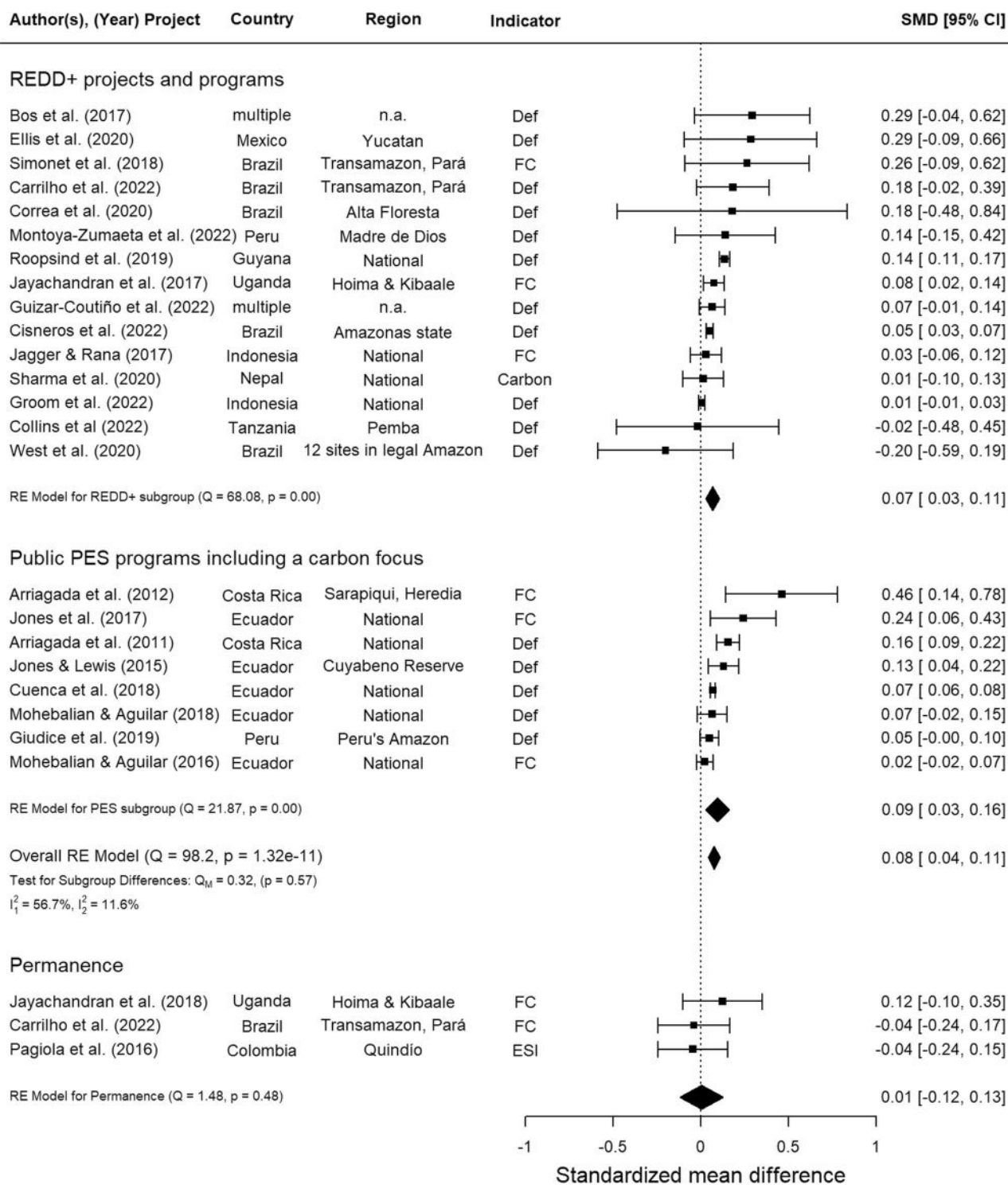


Fig. 2 | Environmental impacts and permanence from REDD+ projects, programmes: small, mostly significant effects. Indicator labels refer to Deforestation (Def) Forest cover (FC), and Ecosystem Services Index (ESI). Random Effects (RE)

models without moderators, standard errors clustered at the country- and study level. We report Cochran's Q test statistic of residual heterogeneity (Q) along with its corresponding p value.

Bosque), seemingly reflecting both variations in output variables and in matching methods (cf. Methods). Of special interest would also be the performance of larger-scale, jurisdictional-level REDD+, given the ongoing implementation shift towards those. These results were moderately encouraging, with conservation impacts in Guyana (Roopsind et al. 2019), Indonesia (Groom et al. 2022) (both NICFI) and Amazonas, Brazil (Cisneros 2022) all being significantly positive, although the latter two very small-sized (all case references in SI).

Looking at secondary impacts, i.e., indicators not directly comparable to forest-cover proxies (SI, Fig. S3a), we observed also here some larger, significant impacts, such as boosting tree-species richness, avoiding wildfire incidence or slowing forest encroachment. Yet, this extended picture remained variable, too. Notably, impacts on forest degradation, the second D in REDD+, were small or statistically insignificant, just like deforestation.

Finally, little is known so far about the permanence of REDD+, i.e. to what extent prospective conservation impacts lasted after the intervention

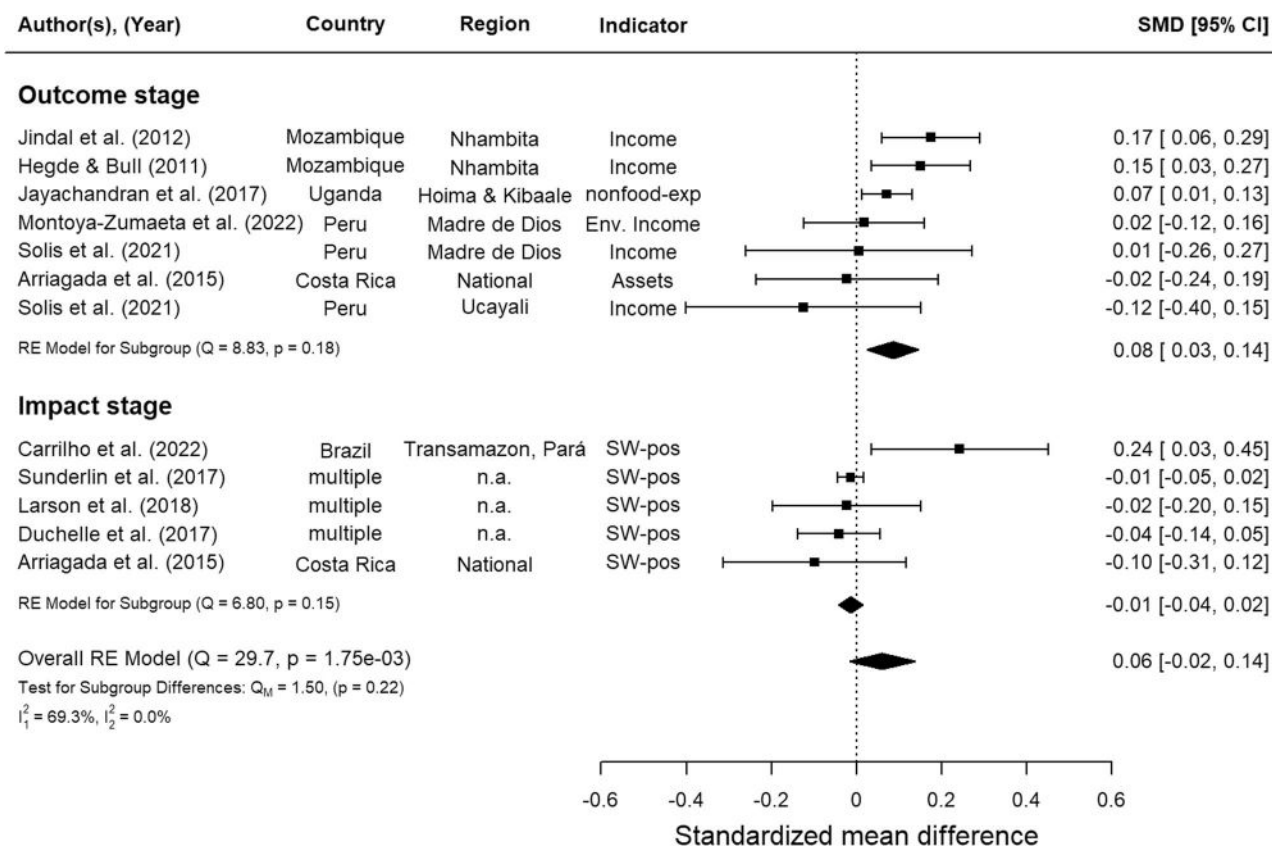


Fig. 3 | REDD+ socioeconomic effects: small positive outcomes (i.e. material welfare proxies) yet insignificant impacts (i.e. subjective wellbeing). Indicator label “SW-pos” refers to self-reported changes in subjective wellbeing. Random

Effects (RE) models without moderators, standard errors clustered at the country- and study level. We report Cochran’s Q test statistic of residual heterogeneity (Q) along with its corresponding p value.

had ended. Estimates across the three permanence studies (excluded from our overall REDD+ effect sizes in Fig. 2 upper panels) originating from Uganda (Jayachandran et al. 2018; see SI literature list), Colombia (Pagiola et al. 2016), and Brazil (Carrilho et al. 2022) differ somewhat, but all coefficients were statistically insignificant (Fig. 3, lower panel). The dominant pattern here was thus that, following confirmed REDD+ deforestation reductions during implementation, post-REDD+ forest loss returned approximately to its pre-intervention rate, but without eliminating the temporary conservation and climate mitigation gains achieved.

Socioeconomic impacts

In our REDD+ Theory of Change (SI, Fig. S2), the most important side-objective of REDD+ was to improve local people’s wellbeing. Figure 3 thus shows comparable socioeconomic effects from rigorous REDD+ impact evaluations. Like for environmental impacts, these were divided between *outcomes* (changes in income, consumption, or asset holdings) and *impacts*: the self-stated subjective wellbeing, and changes herein, on behalf of REDD+ participants and other residents. Also here, our outcome variables were very close impact proxies—and could often be verified more objectively.

Empirically, the two types of indicators performed differently (p = 0.03): at the outcome stage REDD+ had a significant positive, welfare-improving effect size of 0.09 (95% CI: 0.03–0.15), while at the impact stage the mean effect of –0.01 was statistically indifferent from zero. Hence, REDD+ on average tended to make benefit recipients slightly better off materially. Nevertheless, this may not always have boosted self-perceived welfare. The few studies where outcomes and impacts were measured simultaneously (Arriagada et al. 2015; Montoya-Zumaeta et al. 2022) confirmed this trend: the material benefits provided could fall short of community expectations, especially when these were ex-post assessed, after benefit flows had ceased. Thus, self-stated subjective wellbeing may also

become a strategic vote of protest by local participants over unequal REDD+ benefit sharing.

Looking at other socioeconomic outcomes (SI, Fig. S3b), we found more impacts for subgroups to come out negative, such as the subjective wellbeing of female REDD+ participants²¹. This is a reminder that modest average gains in material welfare from REDD+ do not necessarily warrant equity or a do-no-harm principle: distributional and non-material effects may still create (objective or self-perceived) losers.

Contextual and design factors

We conducted a moderator analysis for potential hints about the implicit role of different REDD+ context and design factors in co-determining the above-assessed effect sizes. Hence, we plotted conditional impact sizes against selected variables (Fig. 4A–H). We used all available observations to account for within-study subgroup differences in the design variables. Figure 4 shows the estimated coefficients along with its 95% confidence intervals. Our number of effect-size observations is low for solidly exploring correlations, but we can set hypotheses for future research.

First, this includes baseline deforestation pressure (Fig. 4A, B, E, F), which for other conservation tools correlated positively with impact size^{26,31}: intuitively, the lower ex-ante forest-loss threats were, the harder it would become to counterfactually demonstrate progress. For REDD+, indeed we confirmed a positive relationship between threat and impact, statistically insignificant at national (Fig. 4) but significant (p = 0.057) at the zoomed-in scale of subnational deforestation pressure (Fig. 4) (cf. Methods for classification). Socioeconomically, low-threat REDD+ might have accompanied higher welfare gains (e.g. through lower opportunity costs), as also indicated by the coefficient sign here, yet this correlation is statistically insignificant (Fig. 4E, F).

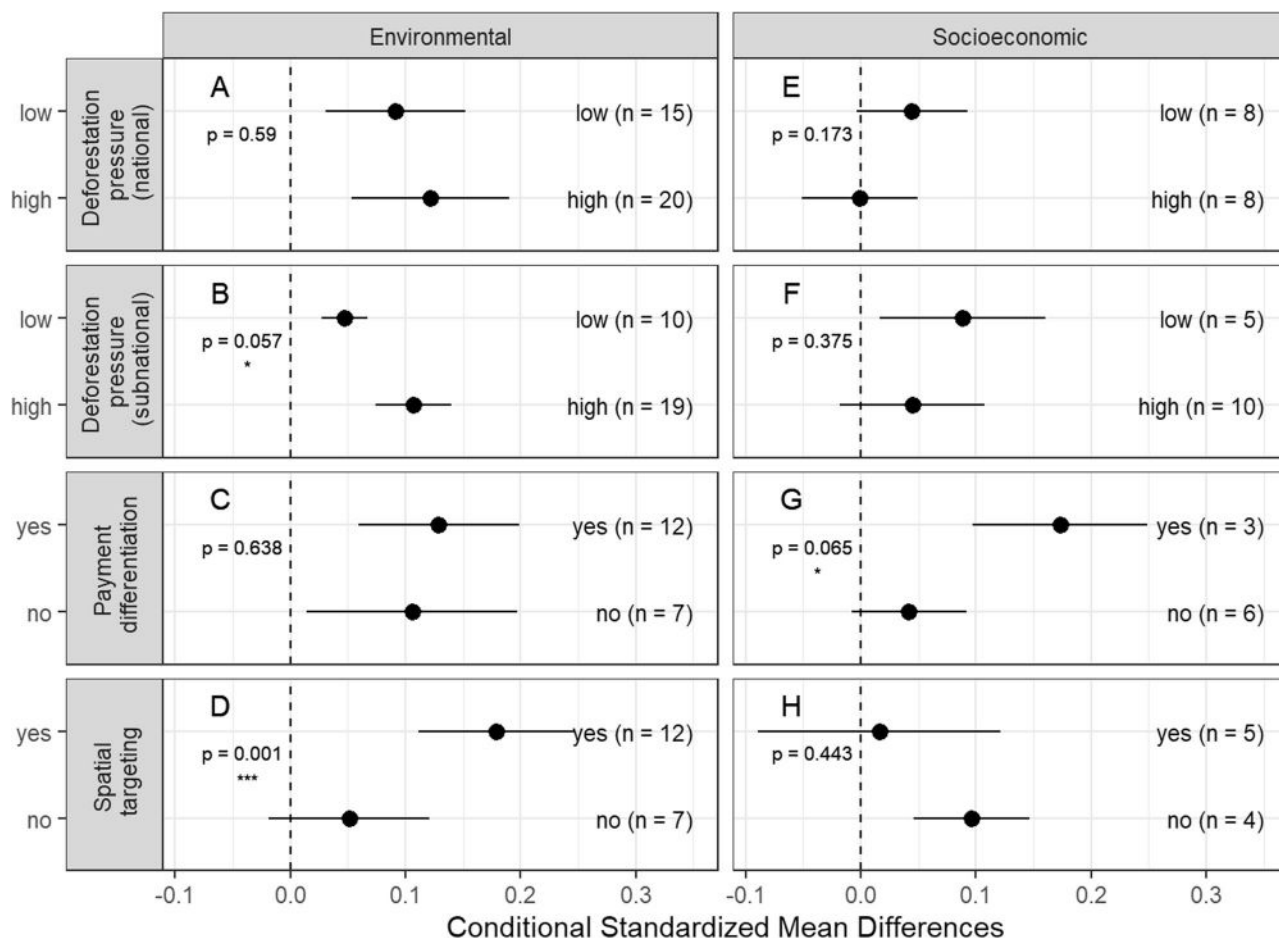


Fig. 4 | Moderator analysis: prior threat levels and spatial targeting matter for environmental, benefit differentiation for socioeconomic effects. A–D (left) show conditional environmental impacts, E–H (right) show conditional socioeconomic impacts. The x-axis in all plots shows the conditional standardised mean differences, y-axis the respective moderator level. From top to bottom, the moderators are contextual (national and subnational deforestation pressure, categorised as high and low) and design (whether there was payment differentiation and spatial targeting).

The models have no intercept, so the coefficients can be interpreted directly as conditional mean effects, along with 95% confidence intervals. We report p values from a Wald test for equality of the moderators, clustered at the study level. The number of observations differs here from Figs. 2 and 3 because: a) we used disaggregated effect sizes, i.e. all available observations per study to exploit within-study variation, and b) we do not show observations with missing moderator values (in particular, all cross-country studies).

Turning to design factors, the PES literature indicates that providing differentiated, beneficiary-customised, rather than uniform benefits can boost environmental effectiveness^{32,33}. Again, we found the expected coefficient sign, but the correlation was insignificant (Fig. 4). Yet, socioeconomic welfare improvements were significantly higher in programmes with differentiated, rather than uniform benefits ($p = 0.045$). This indicates that social customisation may have been important in REDD+ benefit-sharing strategies.

As for the second REDD+ design variable, spatially targeting lands with high density of/high threat towards environmental benefits is in the PES literature also featured as key for additionality^{32,33}. Indeed, spatial targeting for forest-carbon density and/or expected deforestation strongly correlated ($p = 0.001$) with environmental impacts (Fig. 4D), but we only observed initiatives with spatial targeting in Latin America (cf. SI, Table S2), which restricts our ability to clearly distinguish between design-induced and any prospective regionally determined differences. Yet, spatial targeting likely contributed to the larger effects observed in Latin America. Meanwhile, it did not moderate the socioeconomic outcomes (Fig. 4H).

Comparing with other conservation instruments

Finally, we know generally too little about the comparative performance across conservation instruments³⁰. In Fig. 5, we compared the normalised Hedges' G effect sizes recorded for the two types of REDD+ with those of

pre-existing conservation instruments, such as other incentives, disincentives, and enabling actions—drawing on previous studies^{26,31}. As in Fig. 2, we accounted for dependent effect sizes from the same study by assuming a correlation of 0.8 and used robust variance estimation.

As for performance, the two REDD+ subgroups (defined as in Fig. 2) compared fairly well to the other three instrument categories, in terms of mean effect sizes to protect forests (2nd and 3rd numerical ranks, among five). However, there was also a large variability of REDD+ outcomes underlying the rather small intervention samples. Consequently, no statistically significant differences between REDD+ and any of the alternative conservation instruments could be found. Comparatively speaking, REDD+ exhibited a middling, yet also changeable conservation performance.

Discussion

Since 2007, the world has incipiently gathered experiences with REDD+, a tool designed to conserve and enhance forest carbon in non-Annex I countries (i.e., largely developing/ emerging economies, plus China), in exchange for economic compensations from the industrialised Global North. REDD+ is an objectively desirable end (the goal of reducing forest-based emissions) but has equally become a controversial means of using 'market-based' international offsets to help accelerate climate change mitigation.

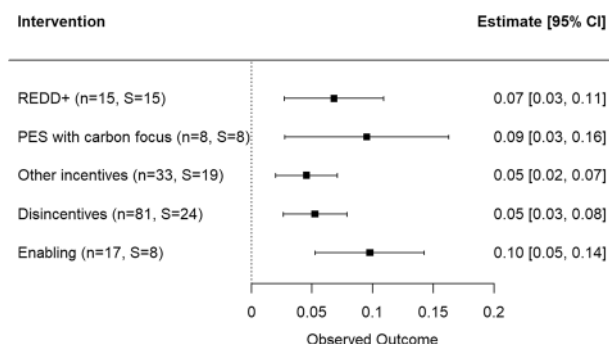


Fig. 5 | Environmental impacts from REDD+ vs. other conservation tools: comparing Cohen's D effect sizes. REDD+ data from own calculations; other tool effect sizes derived from Börner et al. (2020). REDD+ mean impacts rank 2nd and 3rd among the five tools, but differences are statistically insignificant.

A broad range of REDD+ pilot projects has thus emerged. Jointly, they annually planned to avoid 84 million tCO₂ of emissions (over 33 years of mean lifespan) or 'around 1% of annual emissions from deforestation, forest degradation, harvesting and peat fires in the tropics'¹⁰. While potentially important, leakage and credit performance apart only 5% of the correspondingly needed carbon credits had so far been sold on the voluntary market. In terms of de facto avoiding existing deforestation at scale, REDD+ projects have thus been but a drop in the sea¹⁰.

Effectively, REDD+ projects have been starved out by a grossly insufficient global willingness to pay for mitigating climate change. Uncertain funding prospects had also many projects quasi-placeholders waiting whether funding flows would materialise, meanwhile keeping on-the-ground treatment intensities low^{9,14}. In particular, implementers have been hesitant to introduce PES-type of continuous compensations to landholders, since implementers could not promise continuity^{34,35}. Obviously, this deteriorated the framework conditions under which REDD+ projects were hoped to deliver efficient results.

Yet, this does not per se question the potential usefulness of REDD+ projects in providing valuable pilot lessons for any upscaling to jurisdictional REDD+. Above, we have taken stock of the experiences so far. We carefully delimited which initiatives were to be labelled as REDD+—either by proponents, or by analysts. We also screened which impact evaluations were sufficiently rigorous to ensure internal validity and deliver trustworthy results, based on realistic counterfactuals enabling credible causal attribution. With many new empirical studies emerging recently, our larger and geographically more balanced sample than in previous meta-studies should also increase confidence in our results.

We can thus shed some light on the effectiveness and welfare implications of REDD+ initiatives. As for forest-cover and carbon effects—the ultimate *raison d'être*—REDD+ initiatives have had small-sized effects, similar to what other conservation instruments have (not) achieved^{26,31}. This holds for both specialised REDD+ projects/ programmes and cases where REDD+ has been integrated into national PES programmes. When interventions stopped, prior pressures tended to resume, yet typically without 'catching up' with the REDD+ induced halt in deforestation, thus not fully undoing REDD+ conservation gains (partial permanence)³⁶. Overall, given limited carbon-market financial flows, and the harsh critique against REDD+, environmental impacts in our mixed sample of private-public REDD+ initiatives indeed remained small in size, but were still larger than those found for another sample featuring only private-sector REDD+ projects³⁷.

For lack of uniform cost data, we could above not systematically compare cost-efficiency parameters. The few available case studies with REDD+ costs data point to highly variable, in some cases elevated transaction costs, but also declining with scale^{38–40}. A move towards larger-scale jurisdictional REDD+ programmes may thus also push towards more 'bang for the buck' in climate-change mitigation.

On the socioeconomic side of local benefits provided—REDD+'s primary side-objective—our results on average portrayed small positive contributions to local livelihood outcomes (e.g. incomes, assets), yet variable and mostly insignificant impacts (e.g. subjective wellbeing). Notably, new incentive-based projects also tended to locally build easy-to-disappoint expectations regarding future benefit flows⁴¹. Customised, rather than equal benefit transfers seemed to improve socioeconomic outcomes. While not everybody locally may have gained from REDD+, a narrow outcome range from welfare neutrality to modest livelihood improvements was most common. Our moderate quantitative findings thus complement the more negative picture typically portrayed by the qualitative REDD+ literature stressing deficiencies in participation, equity, land-tenure and governance issues^{17–19,42}.

But why have the environmental impact results of this innovative tool overall not been much better? Above we have incipiently pointed to several design flaws, such as adverse selection bias and inadequate spatial targeting. Insufficient on-the-ground enforcement of contractual conditionality is, however, another commonly noted deficiency in REDD+ implementation: often implementers preferred to tolerate land-use violations, safeguarding instead the social capital built with local communities (e.g. Brazil¹³; Cisneros et al. 2020; Peru: Montoya-Zumaeta et al. 2022; Giudice et al. 2019). Moreover, by tolerating exaggerated local baselines of future deforestation, the bar for REDD+ credits was set far too low; unsurprisingly, the majority of credits were non-performing^{37,44}, accelerating a public fatigue with environmental offsets.

Another critical qualitative issue surrounds complexity. REDD+ is to forest carbon what integrated conservation and development projects (ICDP) projects have been to biodiversity: an umbrella term under which a ratatouille of composite, heterogeneous interventions has gathered. Many REDD+ initiatives are 'ICDP-like', in terms of using the same integrated, multifaceted approach: trying a bit of everything to satisfy multiple stakeholders and minimise risks of total failure. Unfortunately, the ICDP approach has had a dismal impact-producing record²⁶.

After 2007, many pre-existing ICDP projects looking for fresh funding were remodelled as REDD+ initiatives, producing an opportunistic self-selection bias. In Indonesia, for instance, many REDD+ projects were implemented by biodiversity-focused organisations; the targeted forest areas were more biodiversity-rich than carbon-dense, and only about one quarter of the project areas was truly threatened by deforestation⁴⁵. In turn, many genuinely new private sector initiatives adversely targeted de facto low-threat areas: avoiding deforestation would here become a low-hanging fruit⁴⁴. Hence, many REDD+ projects may have served more as a proof of concept than as a real test of whether the avoided-deforestation approach is holding water.

Arguably, it is no shame for pilot projects to underperform or fail, if useful lessons are indeed learned for current or future initiatives⁴⁶. Did early REDD+ interventions thus maximise this learning and upscaling potential? Hardly so, mainly since too many projects were carried out in 'high-and-far', i.e. market-remote low-pressure settings, thus not taking the bull by the horns. Particularly the ICDP-type model was also too complex in design and transaction cost-heavy in implementation to replicate at scale. Furthermore, REDD+ implementers almost never facilitated impact evaluation through (quasi-)experimental rollout of multiple design options of action. Hence, we stand back with many highly customised ICDP-like 'boutique projects', including multiple components of action; yet we know very little about how well these components worked, and why.

What about full national REDD+ programmes, as an alleged future of upscaled REDD+? For now, only Norway's NICFI programmes in Guyana (Roopsind et al. 2019) and Indonesia (Groom et al. 2022) have been evaluated, finding for both small yet statistically significant forest-protecting impacts. The Guyana case is not without controversy though: deforestation especially from gold mining actually increased during NICFI support, but less so than was predicted by a synthetic matching model for a no-REDD+ counterfactual, based on other comparable high-forest low-deforestation countries with strong mining sectors.

For future research, doing further analyses of larger-scale REDD+ programmes, be it NICFI or more recently GCF and FCPF interventions, looks promising, but impact evaluation needs to be integrated early into programme design. These impact evaluation analyses should in turn not just provide average effect estimates, but equally be challenged to investigate heterogeneous impacts, enabling us to tell causally plausible stories about where, how, and why REDD+ might work or fail.

Methods

Delineating the concept of REDD+

As we saw above, REDD+ has typically been seen as a prototype type of action (i.e., a means) that generically remains described exclusively by its outcomes of reduced emissions (i.e., an end). This is fundamentally different from other interventions; for instance, protected areas or forest certification, describing means not ends. Observers can thus conceptually come to confuse a model for action (the alleged market-based offsetting strategy of REDD+) with an expected final goal (of having forests store more carbon)³⁵.

Here we thus explicitly walked through the typical stages and assumptions underlying a REDD+ intervention, using a Theory of Change (ToC) approach, designed for causally linking the stages of inputs, treatments, outputs, outcomes and impacts⁴⁷. Figure S2 (SI) outlines these stages going from left to right, with key critical assumptions flagged in bubble shapes.

As for inputs, REDD+ is directly triggered by, and thus essentially dependent on the presence of external finance flows, be it from global markets for carbon credits (as originally envisaged), or from bilateral development/environmental donors (such as Norway's NICFI), multilaterals with a climate mandate (e.g. the Global Environment Facility or the Green Climate Fund), and private-sector non-market flows for direct emission offsets, based on notions of corporate social responsibility.

Generally, serious claims for REDD+ achievements can eventually only be made if knowledge about pre-existing carbon stocks, land-use trends, key drivers and stakeholders triggering forest loss (and protection) jointly can be merged into a credible baseline: what would have happened under the laissez-faire baseline assumption of 'no-REDD+ intervention'? Notably, a proper assessment of levels/ changes in threat is quintessential: if threats from deforestation drivers are rising, treatment may have to be intensified. Conversely, if a projected threat was not to materialise at all, then neither the dynamic counterfactual nor the project will exhibit any deforestation.

REDD+ treatments are highly heterogeneous in their composition. We thus distinguish between the subcategories of incentives, disincentives, and enabling measures⁴⁸. First, invariably some incentives are present in REDD+ as a general local benefit-sharing mechanism, or compensation for the opportunity costs of newly introduced/ enforced restrictions in forest use or conversion to alternative land uses. Incentives can either be conditioned upon compliance with certain land-use rules (e.g. PES-type of contracts), or unconditional investments into alternative, environmentally more benign livelihoods, social sectors (health, education), etc. Often, REDD+ interventions also entail disincentives, through newly introduced restrictions or a more thorough monitoring and sanctioning of non-compliance with already existing ones. Typically, REDD+ has thus included both carrots and sticks. Third, enabling measures as a residual category include tools such as the free prior informed consent (FPIC) of local people's participation in REDD+, a clarification of land tenure and access rules, etc.

Many real-world REDD+ projects and programmes, such as the Bolsa Floresta Programme in Brazil's Amazonas State (Cisneros et al. 2019), or the Sustainable Settlements in the Amazon project in the Transamazon region of Pará State (Simonet et al. 2019; Carrilho et al. 2022) have been using the full spectrum of conditional and unconditional incentives, disincentives, and enabling measures. Pilot interventions experimented with different components, but an underlying belief prevailed that holistic, locally customised approaches carried higher probabilities of success, especially in market-remote, cash-strapped frontier regions. Unsurprisingly, many REDD+ projects are in their holistic range of actions 'ICDP-like', with a

predominant focus on non-conditional livelihood enhancements^{9,34}. For the same reason, REDD+ projects have also had much to learn from ICDPs⁴⁹.

Public PES programmes with a partial focus on forest carbon goals constitute a second type of intervention. Often, carbon financing has helped to boost the funding of these national-level, or at least regional-scale programmes. Costa Rica's PSA, Peru's National Forest Conservation Programme, and Ecuador's Socio Bosque all constitute such examples, although the latter two combined PES with ICDP components (Giudice et al. 2019; Jones and Lewis 2015). Hence, with forest carbon enhancement for climate change mitigation being flagged as an explicit goal, these PES-based interventions need to be included as another pathway of implementing REDD+.

Outputs are to be understood as the immediate, often short-term results of the 'treatment': the treated recipients need to understand the goals and modalities of the intervention, the rules of the game (incl. land and resource tenure) are clarified, and (dis)incentives well-applied. Delivered outputs imply that stakeholder motivations have been successfully aligned with the goals of the intervention. For this to occur, treatments need to have been well-designed and carefully implemented. From the PES literature, we know that spatial targeting in the selection of participants and their to-be-treated land areas constitutes an Achilles Heel, vis-à-vis two complementary dimensions: a) the site-specific environmental service density (here: forest carbon stocks per hectare), and b) the on-site projected threat (here: of deforestation/degradation) of that stock to become endangered over time. Also, customisation of the benefits (e.g. multiple payment levels) can help making the intervention more cost-effective and equitable^{32,33}.

The outcome level is where the REDD+ rubber hits the road: do critical stakeholders undertake the required behavioural on-the-ground changes? That is, do they reduce forest clearing, charcoal making, or timber harvesting in the REDD+ required manner (environmental outcomes)? Similarly, do income, consumption, and assets increase among those targeted stakeholders (socioeconomic outcomes)? These are all measurable indicators that can potentially be impact-evaluated.

The final transition towards impacts—the overarching primary carbon-related goal of reduced forest-based emissions, as well as ethically and politically important side-objectives related to biodiversity, self-perceived human wellbeing, equity, and tenure security—entail further subtleties.

First, a reality check is to what extent intervention-targeted stakeholders and deforestation drivers have been adequately aligned. For instance, many REDD+ projects are focused on addressing smallholders to reduce their deforestation, but a local surge in land grabbing from more powerful external agents might render these efforts less fruitful in terms of mitigating deforestation.

Second, income and consumption outcomes trigger development feedback loops on the final impacts. Rebound effects refer to treatment-induced changes in household incomes potentially affecting consumption patterns (e.g., higher incomes stimulating meat and dairy consumption) that per se change ecological footprints. Magnet effects refer to the potential of these income changes to attract outside migrants, e.g. through successful employment creation in REDD+ projects. Pull migration could have a bearing on land use, as migrants open up new land plots for subsistence agriculture. Both effects are well-established in the PES literature³¹.

Third, the goal of mitigating climate change is both universal and perpetual. Classical concerns vis-à-vis REDD+ projects are thus to what extent these time- and place-bound interventions contribute to the hoped-for universal and perpetual impacts. As for permanence, the impact of a time-limited treatment on carbon stocks may also only be transitory—though as such still important for mitigating climate change in the short run. Conversely, to the extent the treatment triggers desirable structural changes at the output and outcomes level, permanence might be increased.

Likewise, a REDD+ treatment may not only reduce on-site deforestation, but also push some pressures outside the intervention area—a phenomenon known as leakage. This spillover effect will typically diminish, though not fully erase REDD+ mitigation impacts^{50,51}. The larger the scale and policy embeddedness of the REDD+ intervention, the less leakage we

should expect—a key argument for favouring national programmes over REDD+ projects. The size of leakage in conservation incentive programmes is seldom quantitatively estimated³¹. For high-value products sold on international markets, such as harvesting precious timbers, leakage may be exceptionally high⁵². In general, the higher the price elasticity and the geographical mobility on output and input markets (incl. access to land), the larger leakage we should expect³¹.

Sample delimitation

As mentioned, we aim to take stock of the currently available evidence from rigorous quantitative impact evaluations for REDD+ interventions. This means that we needed to apply various a priori filters of inclusion (cf. Table S1), related both to the underlying REDD+ intervention (Factors 1–4), and subsequently to the case study evaluating its impacts (Factors A–F).

As REDD+ interventions (1), we understood here, firstly, actions that implementers self-denominated using the RED(D)+ label, and secondly, other actions that fully or partially featured forest-based climate mitigation/ carbon outcomes in an explicit way. This included also national-level PES programmes pretending to further forest-carbon objectives; in turn, some large watershed-focused PES programmes (e.g. in China, Mexico, and Vietnam) remained excluded. As for actions (2), many forest carbon programmes included both conservation/ regeneration of standing forests and afforestation/reforestation (A/R) activities; those focused entirely on A/R did not functionally fit the REDD+ definition, and we thus excluded them. In terms of scale (3), we chose to be inclusive of both subnational REDD+ (incl. projects) and emerging national programmes, keeping in mind they likely had different characteristics—cf. also (1). Finally, as a temporal cut-off point for the start of REDD implementation (4), we used year 2007, coinciding with the Bali UNFCCC COP13: pre-2007 forest-carbon initiatives (Joint Implementation, Clean Development Mechanism, etc.) were of comparative interest²⁸, but were inevitably bound to differ from REDD+.

A second layer of filters refers to the analytical level. First, we screened both peer-reviewed and grey-literature studies (A)—considering in a quickly-moving field also recent working paper-staged contributions (assessed by us as ‘high-quality’). As for analysed impacts (B), we looked at both forest carbon (main goal) and welfare effects (primary side-objective). As bottom-line, we understood effects to be observed at the right-hand side of the REDD+ ToC, i.e. both outcomes and impacts (see above). Impact evaluations are often stated in terms of outcomes (C), such as forest cover (deforestation areas, rates) and land-use proxies (e.g. fire incidences), which are more precisely observable than forest carbon in the short-to-medium term. More process-oriented, intermediary outputs (middle part of ToC) were not of our interest (D): they were often more qualitative than quantitative, and less clearly (sometimes, ambiguously) linked to REDD+ bottom-line outcomes. Notably, we included subjectively stated wellbeing (do you now feel better/worse-off/ unchanged than prior to the REDD+ project start?), as a popular socioeconomic bottom line of evaluation (E). Admittedly, these indicators feature potential response biases, and are thus best triangulated with more objectively measurable socioeconomic outcomes.

The final, yet ponderous filtering criterion refers to the quality of impact evaluation (F). To rigorously attribute impacts to interventions, counterfactuals are needed: what would have happened without the REDD+ intervention? We only included impact studies using counterfactuals, i.e. experimental and quasi-experimental methods. This included the alleged ‘gold standard’ of randomised controlled trials (RCT), and Before-After-Control-Intervention (BACI) designs. Various econometric techniques attempt to ex-post model counterfactuals, including using matching to identify adequate control observations, or selecting non-treated units to synthesise control units. Yet, recall techniques can also be used to ex post gather baseline data in the field. To make impact estimates quantitatively comparable, we also needed standard deviation estimates. Many case-study authors did not publish these; we contacted several to obtain this supplement.

Study identification strategy

Our literature search strategy, data extraction procedures, and meta-analysis protocols were registered on the Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/MYDBK>.

We started by screening our pool of studies from prior REDD+ reviews^{26,29,53–55}. Initially, 15 eligible studies with quantitative estimates of REDD+ and carbon-focused PES projects using counterfactual impact evaluation methods was identified. A Boolean search string based on title and abstract of this initial sample was semi-automatically generated, following the method described by Grames et al.¹⁹ (Supplementary Note S1, Fig. S1).

We extracted study characteristics such as location, intervention details, sample characteristics along with Hedge’s G effect sizes (SI, Table S4 and Note S2 for calculation details). Our final sample comprises a total of 30 REDD+ interventions, analysed in 32 studies (a list of all included studies is given in the SI), with 52 effect sizes being included (35 forest-related, 17 socioeconomic outcomes). This includes disaggregated effects being used in the moderation analysis. For the main analyses, we aggregated effects resulting in 23 and 12 estimates for environmental and socioeconomic indicators, respectively. For a meta-study, this remains a fairly small sample, restricting also our analytical options: although the number of rigorous impact studies has expanded rapidly in recent years (more than half of the articles included have been published since 2017), more is needed to reach a critical mass for detailed statistical analysis. Our studies are just about equally divided between specialised REDD+ projects/programmes and PES schemes; yet the latter concentrate on fewer cases. In the former category, some studies are multi-case comparisons, e.g. a pool of Amazon Fund-financed and VCS-certified private REDD+ projects⁴⁴ and cases from CIFOR’s Global Comparative Study on REDD+ (GCS-REDD) (Bos et al. 2017, Duchelle et al. 2017, Larson et al. 2018; Sunderlin et al. 2017).

How well does our final sample represent the REDD+ universe? For recall, it is shaped by the filters we have applied (cf. Table S1), overlaying geographically an initial implementation bias (where have REDD+ investors gone?) with a *research bias* (where have scientists preferred to work, and found access to data?), and *publication bias* (is it more likely that positive results are published than negative or null results?). Our small sample mirrors an ‘absolute’ implementation bias towards Latin America (Brazil, Andes, Mesoamerica); see SI Table S3. It covers less well some ‘high-density’ REDD+ countries (Kenya, Colombia, Guatemala). We did find evidence for a moderate publication bias based on Egger’s regression test on funnel plot asymmetry (Supplementary Note S4, Figs. S4a–d): environmentally positive, significant results have a slightly higher likelihood of getting published. On aggregate, the external validity of our sample is deficient, yet still vastly exceeds that of earlier meta-studies of forest conservation incentives, having been based on smaller and geographically much more biased samples^{24,25}.

Meta-analysis

The meta-analysis was carried out using the standardised mean difference (Hedges’g) as the outcome measure. We use the *metafor* along with *club-Sandwich* packages in R (version 4.3.1). A multi-level random-effects model was fitted to the data, including random effects at the study and country level. For the main estimates, we assumed a correlation of 0.8 within studies and countries, and report robust variance estimates based on the correlated hierarchical effects procedure⁵⁶. We conducted subgroup analyses, testing for differences between self-declared REDD+ and PES-cum-carbon programmes. Similarly, we tested for differences between the outcome and impact levels of the socioeconomic variables. For the moderation analysis we also included binary moderators indicating a) deforestation pressure (1 for high pressure; 0 otherwise); b) spatial targeting (1 if study explicitly mentions ecosystem service density and/or deforestation threat as determining factors for the location and/or intensity of the intervention; 0 otherwise); and c) benefit differentiation (1 if study explicitly mentions differently sized benefit levels within the same scheme; 0 otherwise).

Our binary division between high and low deforestation threat was based on the position vis-à-vis the mean annual deforestation rate over the

period 2001–21 across all countries ($0.28\% \text{ y}^{-1}$) from Global Forest Watch (GFW). We compare this threshold with the average case-level deforestation rate during the last 5 years prior to REDD+ start.

Environmental effects

Among the 23 observations in our main analysis, the observed standardised mean differences ranged from -0.1999 to 0.4623 ; most estimates were positive (91%). The Q -test indicated heterogeneity among the true outcomes ($Q(22) = 98.2097$, $p < 0.0001$, $\tau^2 = 0.0018$, $I^2 = 78.5150\%$); while the average outcome was estimated to be positive, in some studies the true outcome might be negative. Inspection of the studentized residuals did not reveal any values larger than ± 3.0654 , indicating the absence of outliers in this model. Additionally, based on Cook's Distance, one study (Roopsind et al. 2019 on Guyana; cf. SI literature) appeared to exert a notable influence. Figure S4a, b presents funnel plot of the estimates. The regression test showed funnel plot asymmetry ($p = 0.032$), although the rank correlation test did not ($p = 0.8346$).

Socioeconomic effects

For the 12 included observations, the observed standardised mean differences ranged from -0.125 to 0.242 ; half of the estimates were negative (51%). The Q -test indicated heterogeneity among the true outcomes ($Q(11) = 27.74$, $p = 0.0035$, $\tau^2 = 0.0045$, $I^2 = 62.8\%$); although the average outcome was estimated to be positive, in some studies the true outcome may be negative. The studentized residuals showed no values exceeding ± 2.87 , i.e., no indication of outliers. Based on Cook's Distance, none of the studies could be considered overly influential. Figure S4a, b presents funnel plots of the estimates. Neither the rank correlation nor the regression test indicated any funnel plot asymmetry ($p = 0.8406$ and $p = 0.7048$, respectively).

Robustness checks

Lacking significant differences between the two forest-size outcomes (forest cover, deforestation rate), we included them both in the same primary-effect analysis. In addition, we found no evidence that impact estimates would vary systematically with programme duration. We tested to what extent our results were driven by a few influential studies⁵⁷ by a) consecutively excluding studies with high weights, namely Groom et al. (2022) and Guizar et al. (2022); b) excluding studies using the synthetic control method, and c) excluding studies with a Cook's Distance larger than two standard deviations. The coefficient sizes slightly changed, but our conclusions remained robust.

Several studies employed matching techniques, and to calculate effect sizes one requires the correlation between pairs of observations (Borenstein 2009:29, Formula 4.27)⁵⁸. Due to missing data, we assumed a correlation coefficient of 0.5 for our main specification but tested also more extreme values (0.3 and 0.7) as a robustness check (Supplementary Note S3, Fig. S6). Indeed, we found that REDD+ effect estimates are sensitive to the assumed parameter in the calculation method, but not enough to alter our findings in Figs. 2 and 3.

Risk of bias assessment

The risk of bias assessment (Supplementary Note S5, Figs. S5a–d) revealed variable methodological quality, both for studies reporting environmental and socioeconomic outcomes: some revealed low risk of bias, others some concerns, or high risk. Caution is needed in interpreting findings, particularly for studies with high risk of bias, as they may impact the overall strength of evidence. For environmental outcomes, bias sources included missing data and deviations from intended interventions; for socioeconomic outcomes, randomisation and deviations from interventions were significant sources of bias.

Data availability

All research data and procedures have been made publicly and enduringly available through the Open Science Framework (OSF) data repository: <https://doi.org/10.17605/OSF.IO/MYDBK>.

Received: 13 October 2023; Accepted: 5 July 2024;

Published online: 22 July 2024

References

- Santilli, M. P. et al. Tropical deforestation and the Kyoto Protocol: an editorial essay. *Clim. Change* **71**, 267–276 (2005).
- Mollicone, D. et al. An incentive mechanism for reducing emissions from conversion of intact and non-intact forests. *Clim. Change* **83**, 477–493 (2007).
- Stern, N. *Stern Review: The Economics of Climate Change*. Cambridge University Press (2006).
- Turnhout, E. et al. Envisioning REDD+ in a post-Paris era: between evolving expectations and current practice. *Wiley Interdiscip. Rev. Clim. Chang.* **8**, 1–13 (2017).
- Angelsen, A. REDD+ as result-based aid: general lessons and bilateral agreements of Norway. *Rev. Dev. Econ.* **21**, 237–264 (2017).
- Pagiola, S. *Using PES to Implement REDD*. World Bank, Latin America and Caribbean, Sustainable Development Department (pp. 22) (2011).
- Wunder, S. Can payments for environmental services reduce deforestation and forest degradation? In Angelsen, A. & Brockhaus, M. (eds) *Realising REDD+: National Strategy and Policy Options* 213–221 (CIFOR, 2009).
- Wertz-Kanounnikoff, S., and Angelsen, A. Global and national REDD+ architecture. In Angelsen, A. (ed) *Realising REDD* 13–24 (CIFOR, 2009).
- Sills, E. O. et al. (Eds). *REDD+ on the Ground: A Case Book of Subnational Initiatives Across the Globe*. (CIFOR, 2014).
- Simonet G. et al. *ID-RECCO, International Database on REDD+ projects and programs, linking Economic, Carbon and Communities data. version 3.0.* <http://www.reddprojectsdatabase.org> (2018a).
- Pedroni, L., Dutschke, M., Streck, C. & Porrúa, M. Creating incentives for avoiding further deforestation: the nested approach. *Clim. Policy* **9**, 207–220 (2009).
- Atmadja, S. S. et al. How do REDD+ projects contribute to the goals of the Paris Agreement? *Environ. Res. Lett.* **17**, 044038 (2022).
- Nepstad, D. et al. More food, more forest, few emissions, better livelihoods: linking REDD+, sustainable supply chains and domestic policy in Brazil, Indonesia and Colombia. *Carbon Manag.* **4**, 639–658 (2013).
- Duchelle, A. E. et al. Forest-based climate mitigation: lessons from REDD + implementation. World Resources Institute Issues Paper (2019).
- Pistorius, T. and Kiff, L. The Politics of German Finance for REDD. Center for Global Development. Working Paper No. 390. <https://doi.org/10.2139/ssrn.2622776> (2015).
- Fletcher, R., Dressler, W., Buescher, B. & Anderson, Z. R. Questioning REDD+ and the future of market-based conservation. *Conserv. Biol.* **30**, 673–675 (2016).
- Corbera, E. Problematizing REDD+ as an experiment in payments for ecosystem services. *Curr. Opin. Environ. Sustain.* **4**, 612–619 (2012).
- Griffiths, T. *Seeing “RED”? “Avoided deforestation” and the Rights of Indigenous Peoples and Local Communities., Forest Peoples Programme, Moreton-in-Marsh, UK* (2007)
- Chomba, S., Kariuki, J., Lund, J. F. & Sinclair, F. Roots of inequity: how the implementation of REDD+ reinforces past injustices. *Land Use Policy* **50**, 202–213 (2016).
- Angelsen, A. et al. Learning from REDD+: a response to Fletcher et al. *Conserv. Biol.* **31**, 718–720 (2017).
- Pham T. T. et al. Strategic alignment: integrating REDD+ in NDCs and national climate policies. In Angelsen, A., Martius, C., De Sy, V., Duchelle, A. E., Larson, A. M. & Pham, T. T. (eds) *Transforming REDD +: Lessons and New Directions*. (CIFOR, 2018).
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. and Vermeersch, C. M. J. *Impact Evaluation in Practice*, 2nd edition. Washington, DC: Inter-American Development Bank and World Bank. <https://doi.org/10.1596/978-1-4648-0779-4> (2016).

23. Ferraro, P. & Pattanayak, S. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biol.* **4**, e105 (2006).
24. Samii, C., Lisiecki, M., Kulkarni, P., Paler, L. and Chavis, L. Effects of payment for environmental services and decentralized forest management on deforestation and poverty in low and middle-income countries: a systematic review. *Campbell Syst. Rev.* **91** (2014).
25. Snilsveit, B. et al. Incentives for climate mitigation in the land use sector—the effects of payment for environmental services (PES) on environmental and socio-economic outcomes in low-and middle-income countries: a mixed-methods systematic review. *Campbell Syst. Rev.* **15**, e1045(2019)
26. Börner, J., Schulz, D., Wunder, S. & Pfaff, A. The effectiveness of forest conservation policies and programs. *Annu. Rev. Resour. Econ.* **12**, 45–64 (2020).
27. Ravikumar, A., Larson, A. M., Duchelle, A. E., Myers, R. & Gonzales Tovar, J. Multilevel governance challenges in transitioning towards a national approach for REDD+: evidence from 23 subnational REDD+ initiatives. *Int. J. Commons* **9**, 909–931 (2015).
28. Caplow, S., Jagger, P., Lawlor, K. & Sills, E. O. Evaluating land use and livelihood impacts of early forest carbon projects: lessons for learning about REDD+. *Environ. Sci. Policy* **14**, 152–167 (2011).
29. Duchelle, A. E., Simonet, G., Sunderlin, W. D. & Wunder, S. “What is REDD+ achieving on the ground?”. *Curr. Issues Sustain.* **32**, 134–140 (2018).
30. Editorial: Weighing up policy tools. *Nat. Sustain.* **4**, 561–561, <https://doi.org/10.1038/s41893-021-00754-y> (2021).
31. Wunder, S., Börner, J., Ezzine-de-Blas, D., Feder, S. & Pagiola, S. Payments for environmental services: past performance and pending potentials. *Annu. Rev. Resour. Econ.* **12**, 209–234 (2020).
32. Ezzine-de-Blas, D., Wunder, S., Ruiz-Pérez, M. & Moreno-Sanchez, R. D. P. Global patterns in the implementation of payments for environmental services. *PLoS ONE* **11**, e0149847 (2016).
33. Wunder, S. et al. From principles to practice in paying for nature’s services. *Nat. Sustain.* **1**, 145–150 (2018).
34. Sunderlin W. D. et al. The challenge of establishing REDD+ on the ground: insights from 23 subnational initiatives in six countries. Occasional Paper 104. Bogor, Indonesia: (CIFOR, 2014).
35. Wunder, S. et al. REDD+ in theory and practice: how lessons from local projects can inform jurisdictional approaches. *Front. For. Glob. Chang.* **3** <https://doi.org/10.3389/ffgc.2020.00011> (2020).
36. Carrilho, C. D. & Wunder, S. What happens after conservation payments stop? *CIFOR Infobrief #379*. Bogor: CIFOR. <https://doi.org/10.17528/cifor/008803> (2023).
37. West, T. A. P. et al. Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science* **381**, 873–877 (2023).
38. Nantongo, M. & Vatn, A. Estimating transaction costs of REDD+. *Ecol. Econ.* **156**, 1–11 (2019).
39. Thompson, O. R. R. et al. Reducing emissions from deforestation and forest degradation (REDD+): transaction costs of six Peruvian projects. *Ecol. Soc.* **18**, (2013).
40. Rakatama, A., Pandit, R., Ma, C. & Iftekhhar, S. The costs and benefits of REDD+: a review of the literature. *For. Policy Econ.* **75**, 103–111 (2017).
41. Montoya-Zumaeta, J. G., Rojas, E. & Wunder, S. Adding rewards to regulation: the impacts of watershed conservation on land cover and household wellbeing in Moyobamba, Peru. *PLoS ONE* **14**, e0225367 (2019).
42. Milne, S. et al. Learning from ‘actually existing’ REDD+: a synthesis of ethnographic findings. *Conserv. Soc.* **17**, 84–95 (2019).
43. Rosa da Conceição, H., Börner, J. & Wunder, S. REDD+ as a public policy dilemma: understanding conflict and cooperation in the design of conservation incentives. *Forests* **9**, 725 (2018).
44. West, T. A. P., Börner, J., Sills, E. O. & Kontoleon, A. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proc. Natl Acad. Sci. USA* **117**, 24188–24194 (2020).
45. Murray, J. P., Grenyer, R., Wunder, S., Raes, N. & Jones, J. P. Spatial patterns of carbon, biodiversity, deforestation threat, and REDD+ projects in Indonesia. *Conserv. Biol.* **29**, 1434–1445 (2015).
46. Chambers, J. et al. The right to fail? Problematizing failure discourse in international conservation. *World Dev.* **150**, 105723 (2022).
47. Weiss, C. H. How can theory-based evaluation make greater headway? *Eval. Rev.* **21**, 501–524 (1997).
48. Börner, J. and Vosti, S. Managing tropical forest ecosystem services: an overview of options. *Governing the provision of ecosystem services*. R. Muradian and L. Rival. Dordrecht, Springer: 21–46 (2013).
49. Blom, B., Sunderland, T. & Murdiyarto, D. Getting REDD to work locally: lessons learned from integrated conservation and development projects. *Environ. Sci. Policy* **13**, 164–172 (2010).
50. Pfaff, A. & Robalino, J. Spillovers from conservation programs. *Annu. Rev. Resour. Econ.* **9**, 299–315 (2017).
51. Meyfroidt, P. et al. Focus on leakage and spillovers: informing land-use governance in a tele-coupled world. *Environ. Res. Lett.* **15**:9 <https://iopscience.iop.org/article/10.1088/1748-9326/ab7397> (2020).
52. Sohngen, B. & Brown, S. Measuring leakage from carbon projects in open economies: a stop timber harvesting project in Bolivia as a case study. *Can. J. For. Res.* **34**, 829–839 (2004).
53. Burivalova, Z. et al. What works in tropical forest conservation, and what does not: effectiveness of four strategies in terms of environmental, social, and economic outcomes. *Conserv. Sci. Pract.* **1**, e28 (2019).
54. Montoya-Zumaeta, J. G. Reconciling conservation and development? Impacts of incentive-based forest conservation in Peru (Ph.D. Thesis), (Australian National University (ANU), 2021).
55. Games, E. M., Stillman, A. N., Tingley, M. W. & Elphick, C. S. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods Ecol. Evol.* **10**, 1645–1654 (2019).
56. Pustejovsky, J. clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. *R package version 0.5.2*. <https://CRAN.R-project.org/package=clubSandwich> (2020).
57. Aguinis, H., Gottfredson, R. K. & Joo, H. Best-practice recommendations for defining, Identifying, and handling outliers. *Organ. Res. Methods* **16**, 270–301 (2013).
58. Borenstein, M. Effect sizes for continuous data. In Cooper, H. Hedges, L. V. & Valentine, J. C. (eds) *The Handbook of Research Synthesis and Meta-analysis* (2nd ed.) 221–235 (Russell Sage Foundation, 2009).

Acknowledgements

We are grateful for financial support from Norad through CIFOR’s Global-Comparative Study on REDD+, Germany’s Federal Ministry for Food and Agriculture, (BMEL) through EFI’s NewGo project, and the German Research Foundation under its Excellence Strategy [EXC-2070-390732324-PhenoRob], and the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101081744. We thank multiple authors of primary studies for sharing with us unpublished quantitative material from their respective impact evaluations, and commentators on previous versions of this document: Amy Duchelle, Steven Dörr, Leif-John Fosse, Yitagesu Tekle, Thales West, and Georg Winkel.

Author contributions

S.W., D.S. and J.B. conceptualised the project and designed the methodology, D.S. and J.G.M.Z. curated the data, D.S., S.W., and J.G.M.Z. led the formal analyses; B.B.C. led the risk of bias assessment; S.W. and J.B. acquired funding and supervised the work, G.P.F. undertook the spatial analysis visualising the global map, S.W. wrote the first article draft, D.S. wrote the SI and visualised the data, and all authors contributed to the review and editing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s43247-024-01541-1>.

Correspondence and requests for materials should be addressed to Sven Wunder.

Peer review information *Communications Earth & Environment* thanks Adeniyi P. Asiyebi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Martina Grecequet. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024