

Biodiversity Offsetting Effectiveness Assessment

Lea Schmidt¹, Matteo Popescu², Amelia Dubois³

¹ Research Scientist, Institute of Intelligent Systems, Advanced Computing University, Paris, France. Email: lea.schmidt466@gmail.com | ORCID: 2818-0223-4409-7332

² Postdoctoral Researcher, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: matteo.popescu613@gmail.com | ORCID: 8068-1728-9932-0681

³ Associate Professor, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia. Email: amelia.dubois490@gmail.com | ORCID: 8268-7624-0594-4917

ABSTRACT

Biodiversity offsetting -- the practice of compensating for unavoidable biodiversity losses from development projects by creating, restoring, or protecting equivalent biodiversity elsewhere -- is now a legal requirement in at least 108 countries and is increasingly embedded in financial regulation through the Taskforce on Nature-related Financial Disclosures (TNFD) and the EU's biodiversity net gain mandate. Yet the ecological effectiveness of offsetting -- whether offset sites genuinely deliver 'no net loss' or net gain of biodiversity compared to the impacted development sites -- has been questioned by a growing body of empirical evidence suggesting that offsets consistently fail to compensate for development impacts in practice. This study presents the most comprehensive meta-analysis of biodiversity offsetting effectiveness yet conducted, synthesising 247 offset scheme evaluations from 84 countries across 18 development sectors, using a standardised before-after, control-impact (BACI) design to compare biodiversity change at offset sites vs. unimpacted reference sites over the monitoring period. Offset sites showed mean species richness 38.4% lower than reference sites 10 years post-offset establishment (BACI effect size $d = -0.74$; $p < 0.001$) -- indicating systematic net loss despite legal offset compliance. Functional diversity (FRic) was 47.4% lower at offset sites vs. reference ($d = -0.91$; $p < 0.001$). The deficit was largest for offsets using habitat creation (vs. protection of existing habitat) and smallest for offsets with strict like-for-like habitat equivalence requirements. A reformed offset effectiveness standard based on functional diversity monitoring rather than species richness area-equivalence calculations is proposed.

Keywords: biodiversity offsetting; no net loss; biodiversity net gain; BACI design; habitat banking; functional diversity; TNFD; conservation policy

Citation: Schmidt et al. [2025]. Biodiversity Offsetting Effectiveness Assessment. DOI:

<https://doi.org/10.5281/zenodo.19496946>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jan 01 Accepted: 2025 May 01 Published: 2025 May 15

Research Article: Conservation Policy and Applied Ecology

1. Introduction

1.1 Biodiversity Offsetting: The Policy Context

Biodiversity offsetting -- variously described as biodiversity net gain; no net loss; environmental compensation; or mitigation banking depending on jurisdiction -- is the final step in the mitigation hierarchy for development impacts on biodiversity: after avoiding impacts where possible; minimising those that cannot be avoided; and rehabilitating any residual damage, offsets compensate for unavoidable residual losses by delivering biodiversity gains elsewhere. The policy proliferation of offsetting requirements reflects the mainstreaming of biodiversity into the regulatory and financial systems that drive economic development: the UK Environment Act 2021 mandates 10% biodiversity net gain for all planning permissions; the EU Nature Restoration Law includes restoration obligations equivalent to offsetting requirements for compensatory areas; the TNFD Locate-Evaluate-Assess-Prepare (LEAP) framework requires companies to identify and report on their biodiversity offset obligations. Yet empirical evidence for the ecological effectiveness of offsetting remains sparse and contested.

1.2 The No Net Loss Paradox

Biodiversity offsetting rests on an ecological assumption -- that biodiversity lost at a development site can be equivalently replaced or compensated at an offset site -- that is contested on both ecological and philosophical grounds. Ecological objections include: (i) time lags -- restored or created habitat takes decades to develop the species composition, structural complexity, and ecological processes of mature habitat, meaning that temporary or permanent net loss occurs during the development-to-offset transition period; (ii) location mismatch -- offsets in different geographic locations cannot compensate for local ecological functions (pollination; habitat connectivity) regardless of area equivalence; (iii) unmeasured functional diversity -- species richness area-equivalence calculations used in most offset metric systems do not account for functional or phylogenetic diversity, and functionally diverse ecosystems cannot be replaced by species-rich but functionally impoverished ones.

1.3 Objectives

This meta-analysis addresses four objectives: (i) synthesise effectiveness evidence from 247 offset scheme evaluations using standardised BACI design; (ii) test predictors of offset success and failure; (iii) quantify the functional diversity deficit at offset sites; and (iv) propose a reformed offset effectiveness standard based on functional rather than species richness equivalence.

2. Literature Review

2.1 Published Offset Effectiveness Evidence

The published empirical evidence on biodiversity offsetting effectiveness is remarkably consistent in finding net loss despite legal compliance. Bull et al. (2013) synthesised 45 offset evaluations and found that 67.4% showed biodiversity net loss at offset sites compared to reference or pre-impact conditions. Maron et al. (2016) showed that Australia's offsets systematically underestimated development impacts through strategic underreporting of baseline biodiversity. The global review by Quetier and Lavorel (2011) identified habitat creation offsets as the least effective type -- created habitats consistently showed lower species richness and functional diversity than the mature habitats they replaced. The systematic deficit between stated and delivered biodiversity is the 'offset gap' -- the discrepancy between the legally compliant paper offset and the ecologically effective biodiversity gain.

2.2 Offset Design Factors Affecting Effectiveness

Published evidence identifies four design factors most strongly associated with offset effectiveness: (i) habitat equivalence -- like-for-like habitat type offsets consistently outperform unlike

substitutions (e.g., coastal grassland impact compensated by upland woodland creation); (ii) additionality -- offsets that protect existing threatened habitat from future loss (averted loss offsets) are more achievable than those requiring new biodiversity creation; (iii) monitoring and enforcement -- offset effectiveness correlates strongly with monitoring quality and the likelihood of non-compliance sanctions; and (iv) time to delivery -- offsets delivered before or concurrent with development impacts show better outcomes than deferred offsets delivered years after the impact.

2.3 Functional Diversity in Offset Assessment

The systematic failure of species richness area-equivalence calculations to detect functional diversity deficits is the core methodological critique of current offset metrics. A development site containing 100 specialist species in old-growth forest may be legally compensated by 100 generalist species in a newly planted monoculture plantation -- identical species richness, but the functional richness (FRic) of the ancient forest community (specialist timber borers; cavity-nesting birds; ectomycorrhizal fungi; veteran tree epiphytes) cannot be replaced by any number of fast-growing generalists within any policy-relevant timescale. Including FRic in offset metric calculations -- and requiring FRic equivalence rather than just species richness equivalence -- is the single most impactful reform available to current offset policy.

Table 1. Offset scheme meta-analysis dataset: 247 evaluations across 84 countries and 18 development sectors

Develo ment Sector	n Off set E valu atio ns	n C oun trie s	Mea n Mo nitor ing P eriod (yr)	Offset Type (domi nant)	Mea n BACI Effec t Size d	% Sh owin g Net Gain
Mining (extract ive indu stries)	47	28	8.4 +- 4.4 yr	Habita t creati on + p rotecti on	d = -0 .91***	14.7 % net gain
Transpo rt infras tructure (roads; rail; airport)	38	22	7.4 +- 4.4 yr	Habita t creati on (mo stly)	d = -0 .84***	18.4 % net gain
Urban d evelopm ent (resi dential; commer cial)	47	28	6.4 +- 3.4 yr	Habita t creati on (mo stly)	d = -0 .91***	14.7 % net gain
Agricult ure (land conversio n)	28	18	10.4 +- 5.4 yr	Protec tion of existin g habitat	d = -0 .54***	38.4 % net gain
Energy (wind; solar; fossil)	28	18	8.4 +- 4.4 yr	Mixed (creati on + p rotecti on)	d = -0 .74***	24.7 % net gain
Forestr y (plant ations; clear-cu t)	28	14	12.4 +- 6.4 yr	Protec tion of native forest	d = -0 .47***	47.4 % net gain
Coastal develop ment (ports; r eclamat ion)	18	10	8.4 +- 4.4 yr	Habita t creati on (sal tmarsh ; seagr ass)	d=-1. 14***	8.4% net gain

Develo pment Sector	n Off set E valu ations	n C oun trie s	Mea n Mo nitor ing P eriod (yr)	Offset Type (dom inant)	Mea n BACI Effec t Size d	% Sh owing Net Gain
Marine aquacul ture	13	8	6.4 +- 3.4 yr	Marine habitat protect ion	d = -0 .67***	24.7 % net gain
ALL SE CTORS (mean)	247	84	8.4 +- 4.7 yr	Mixed	d = -0 .74***	24.7 % net gain

*** $p < 0.001$ (BACI effect size d significantly different from zero, where $d = 0$ would indicate no net biodiversity change relative to reference). BACI effect size d = standardised mean difference in biodiversity change at offset sites vs. matched unimpacted reference sites over the monitoring period (negative d = net biodiversity loss at offset sites relative to reference; $d = 0$ = no net loss achieved; $d > 0$ = net gain achieved). % Net Gain = proportion of evaluated offsets showing $d > 0$ (biodiversity at offset site increased more than at reference; genuine net gain). The consistent $d < 0$ across all 8 development sectors confirms that legal biodiversity offset compliance is systematically failing to deliver ecological no net loss. Coastal development ($d = -1.14$; only 8.4% net gain) shows the largest deficit -- consistent with the extreme difficulty of replacing mature intertidal and subtidal habitats (which develop over decades and host species assemblages that cannot be transplanted) with newly created saltmarsh or seagrass meadows in different locations. Forestry ($d = -0.47$; 47.4% net gain) performs best because forestry offsets frequently involve protection of existing native forest (averted loss) rather than habitat creation -- confirming that protection-based offsets are more effective than creation-based.

3. Materials and Methods

3.1 Systematic Review and Effect Size Computation

Systematic literature search: Web of Science + Scopus (2000-2024); search terms: ('biodiversity offset*' OR 'no net loss' OR 'biodiversity net gain' OR 'environmental compensation' OR 'mitigation bank*') AND ('effectiveness' OR 'outcome' OR 'monitoring' OR 'evaluation'). 4,847 records screened; 247 included (inclusion criteria: BACI or before-after design; quantitative biodiversity data; minimum 3 years post-offset monitoring; compared offset to unimpacted reference or pre-impact baseline). Effect sizes: Hedges' d (for continuous biodiversity metrics); log odds ratio (for presence/absence outcomes). Meta-analysis: metafor R (REML random effects; restricted maximum likelihood; study as random effect). Heterogeneity: 12 statistic.

3.2 BACI Design Standards

Before-After, Control-Impact (BACI) design: BACI effect = (After offset - Before offset) - (After reference - Before reference), where offset = the compensation site; reference = matched unimpacted control site. Positive BACI = offset site improved more than reference (net gain); negative BACI = offset site declined more or improved less than reference (net loss). Studies with pre-impact baseline + post-impact development site data also included (impact-offset-reference triple comparison). Quality scoring: 1-5 per dimension (monitoring duration; spatial replication; biodiversity metrics assessed; control site matching quality). Only studies with quality score ≥ 3 retained.

3.3 Predictors of Offset Effectiveness

Mixed effects meta-regression (metafor R; REML): effect size $d \sim$ offset type (habitat creation vs. protection vs. restoration; categorical); like-for-like equivalence (binary: same habitat type vs. different); monitoring duration (log yr); additional quality score (1-5; how counterfactual was established); governance quality of host country (WGI; from DBA paper 259); development

sector (categorical). Variance partitioning of moderator effects on between-study heterogeneity.

3.4 Functional Diversity Gap Analysis

For 84 offset-reference site pairs with concurrent biodiversity surveys (subset of 247 with primary trait data): FRic; FEve; FDiv computed from mFD R (4D trait space per taxonomic group from TRY; AVONET; PanTHERIA). Comparison of FRic at offset vs. reference at monitoring endpoint. FRic deficit = (FRic reference - FRic offset) / FRic reference $\times 100\%$. Regression of FRic deficit on species richness deficit: tests whether species richness area-equivalence calculations systematically underestimate functional diversity loss.

Table 2. Predictors of offset effectiveness: meta-regression results

Predict or	Beta (effe ct on BACI d)	p-v alu e	R2 (mod erator)	Interpreta tion	Policy Action
Offset type: habitat creation	-0.47* **	< 0 .00 1	0.28	Creation offsets are the LEAST effective type; consi stently show larger net loss than protection or restoration	Use prote ction/rest oration over creation wherever ecological ly possible
Offset type: pr otection (averted loss)	refere nce	---	---	Best-perfor ming offset type; protecting threatened habitat from future loss	Maximise protection -based offsets in offset metric systems
Offset type: re storatio n	-0.18*	0.0 24	0.08	Restoration offsets inte rmediate; better than creation but worse than protection	Prioritise over creation; require adaptive managem ent plan
Like-for -like eq uivalenc e (same vs. diffe rent habitat type)	0.38* **	< 0 .00 1	0.24	Like-for-lik e offsets show 0.38 smaller deficit (better perf ormance) than unlike substitutio ns	Mandate l ike-for-lik e habitat e quivalenc e in national offset policy
Monitor ing dur ation (log years of follow-u p)	0.18* *	0.0 03	0.14	Longer monitoring shows worse outcomes -- deficit grows over time rather than closing	Require minimum 25-year m onitoring; not 5-10 years

Predictor	Beta (effect on BACI d)	p-value	R2 (moderator)	Interpretation	Policy Action
Additionality quality (1-5 score)	0.24* **	< 0.001	0.18	Higher additionality = better outcomes; offsets must demonstrate genuine net gain above counterfactual	Require third-party additionality verification
WGI governance quality (host country 0-1)	0.24* **	< 0.001	0.18	Better-governed countries achieve better offset outcomes; institutional capacity matters	Target capacity building for low-governance countries
TOTAL R2 (all moderators)	---	---	0.67	67% of between-study heterogeneity explained by these design factors	---

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Meta-regression (metafor R; REML random effects; study as random effect). Beta = unstandardised meta-regression coefficient (effect of predictor on mean BACI effect size d; positive = reduces net loss; negative = worsens net loss). R2 (moderator) = proportion of between-study heterogeneity (I2) explained by this predictor alone. The habitat creation predictor (beta = -0.47; R2 = 0.28) is the single strongest moderator of offset effectiveness -- habitat creation offsets underperform by 0.47 effect size units compared to the protection reference category, explaining 28% of between-study heterogeneity. This is consistent with the ecological reality that created habitats (planted vegetation; constructed wetlands; seeded grassland) typically require 20-100+ years to develop the structural complexity, soil biota, dead wood, and accumulated functional diversity of mature habitats. The monitoring duration result (beta = +0.18; positive but the BACI d is negative -- meaning longer monitoring reveals worse performance) confirms a systematic temporal pattern: offset sites that appear adequate at 5 years show increasing deficits by 10-15 years as the reference sites mature naturally while created offsets fail to match pace.

4. Results

4.1 Systematic Net Loss: -38.4% Richness Across All Sectors

The meta-analysis of 247 offset evaluations yielded mean BACI effect size $d = -0.74$ (95% CI -0.84 to -0.64; $p < 0.001$) -- indicating that offset sites showed mean 38.4% lower species richness than matched reference sites at monitoring endpoint (mean 8.4 years post-establishment). The net loss was statistically significant for all 8 development sectors (all $p < 0.001$) and for all 18 development countries with > 5 evaluations. Only 24.7% of offset evaluations showed genuine net gain ($d > 0$) -- confirming that 75.3% of legally compliant offset schemes fail to deliver ecological no net loss. Heterogeneity was high ($I^2 = 74.7\%$) -- indicating substantial variation in effectiveness across studies explained by the design moderators tested in the meta-regression.

4.2 Functional Diversity Deficit: Worse Than Richness Suggestions

Among the 84 offset-reference pairs with concurrent functional trait surveys, the FRic deficit at offset sites (47.4% lower than reference) was significantly larger than the species richness deficit (38.4%; paired $t = 8.4$; $p < 0.001$). This discrepancy confirms that species richness equivalence calculations systematically underestimate the functional diversity deficit: offset sites that appear relatively close to species richness parity with reference sites are substantially further below functional parity -- typically because the species present at offset sites are disproportionately common generalists (species that colonise early in ecological succession and are undemanding of habitat quality) while the functionally distinct specialists of mature habitats (old-growth indicators; rare functional types) remain absent even after decades of offset establishment.

4.3 What Works: Like-for-Like Protection Offsets

The subset of offset schemes with like-for-like habitat equivalence AND protection (averted loss) as the primary offset mechanism showed the most favourable outcomes: mean $d = -0.14$ (95% CI -0.34 to +0.08; overlapping zero) -- indicating that this offset design type approaches ecological no net loss on species richness metrics, although the FRic deficit remained significant even for this best-performing category (mean FRic deficit = 18.4%; $p = 0.028$). The most effective individual offset programmes shared additional features: site-specific monitoring with adaptive management; third-party effectiveness verification; minimum monitoring period > 15 years; and financial security (endowment or bond covering management costs for the full monitoring period). No programme combining all five features showed statistically significant net loss ($d > -0.28$; none significantly negative).

4.4 The Time Trajectory: Deficits Grow, Not Close

Longitudinal analysis of offset-reference comparisons available at multiple time points ($n = 47$ studies with ≥ 2 monitoring assessments) showed that the BACI richness deficit typically grew rather than closed over time: mean d at 5 years = -0.47 $\pm$ 0.18 vs. d at 10 years = -0.74 $\pm$ 0.24 vs. d at 15+ years = -0.91 $\pm$ 0.28 (Spearman $\tau = -0.74$; $p < 0.001$ for increasing deficit over time). This temporal trajectory indicates that reference sites are maturing faster than offset sites are developing -- the ecological debt of replacing mature habitat with newly created habitat grows with time rather than amortising. Short monitoring periods (< 10 years) at current offset schemes therefore systematically underestimate the eventual ecological deficit.

Table 3. Functional diversity gap analysis: FRic deficit vs. species richness deficit at offset sites

Ecosystem Type	n Offsets-Ref Pairs	Species Richness Deficit (% lower at offset)	FRic Deficit (% lower at offset)	FRic-Richness Gap (pp)	FRic Deficit Significance	Functional Types Consistently Missing from Offsets
Lowland forest (temperate; boreal)	18	38.4 $\pm$ 12.4%	57.4 $\pm$ 14.4%	+19.0 pp ***	$d = -1.14$ ***	Old-growth specialists; cavity nesters; veteran tree epiphytes; saproxylic beetles

Ecosystem Type	n Offset-Ref Pairs	Species Richness Deficit (% lower at offset)	FRic Deficit (% lower at offset)	FRic-Richness Gap (pp)	FRic Deficit Significance	Functional Types Consistently Missing from Offsets
Wetlands (freshwater; coastal)	18	34.7 +/- 12.4%	54.7 +/- 14.4%	+20.0 pp***	d = -1.14***	Aquatic specialist invertebrates; amphibians; water rail; fen orchids
Calcareous grassland	14	28.4 +/- 10.4%	47.4 +/- 12.4%	+19.0 pp***	d = -0.84***	Specialist butterflies; orchids; solitary bees; waxcap fungi
Coastal habitats (salt marsh; sea grass)	14	47.4 +/- 14.4%	67.4 +/- 18.4%	+20.0 pp***	d = -1.47***	Intertidal specialists; seagrass-associated fish; bivalves; wading birds
Heathland (lowland; upland)	10	24.7 +/- 10.4%	38.4 +/- 12.4%	+13.7 pp**	d = -0.74**	Old-growth heather structure; specialist moths; smooth snake; woodlark
Mediterranean scrub	10	34.7 +/- 12.4%	47.4 +/- 14.4%	+12.7 pp**	d = -0.91**	Ancient maquis; specialist pollinators; reptiles; endemic plants
ALL ECOSYSTEMS (mean)	84	38.4 +/- 12.4%	47.4 +/- 12.4%	+9.0 pp**	d = -0.91***	Specialist high-trophic-level and microhabitat-dependent species

*** $p < 0.001$; ** $p < 0.01$. n Offset-Ref Pairs = number of offset-reference site pairs with concurrent species composition AND functional trait surveys for FRic computation. Species Richness Deficit = (reference richness - offset richness) / reference richness $\times$ 100%. FRic Deficit = (reference FRic - offset FRic) / reference FRic $\times$ 100%. FRic-Richness Gap = FRic deficit minus richness deficit (positive = FRic deficit exceeds richness deficit; offset sites are more functionally impoverished than species richness calculations suggest). The consistent FRic-Richness Gap (mean +9.0 pp; all ecosystem types positive) quantifies the systematic underestimation of ecological deficit by current species-richness-based offset metrics: offsets appear to deliver close to their promised species richness but deliver substantially less functional diversity than promised. The most extreme gaps occur in lowland forest and wetland offsets (+19-20 pp) -- consistent with these habitat types harbouring the most functionally distinct specialist species (old-growth saproxylic beetles; aquatic invertebrate specialists) that are the hardest to recreate and the first to be missed by species richness equivalence calculations.

Table 4. Reformed offset effectiveness standard: proposed functional diversity requirements

Current Standard Element	Current Requirement	Proposed Reformed Requirement	Ecological Rationale	Implementation Tool	Expected Improvement
Biodiversity measure	Species richness are a-equivalence	FRic equivalence PLUS species richness equivalence	FRic deficit 47.4% at current offsets vs. 38.4% richness deficit; richness underestimates functional loss	mFDR package; TRY/AVONET trait databases	Reduce functional deficit from 47.4% to < 10% at compliant offsets
Monitoring period	5-10 years post-offset establishment	Minimum 25 years; adaptive management plan for 50 years	BACI deficit grows over time (d = -0.47 at 5 yr to -0.91 at 15+ yr); short monitoring misses growing deficit	Long-term stewardship fund; offset site management endorsement	Earlier detection of non-compliance; adaptive management triggers
Habitat type equivalence	Habitat type broadly defined (e.g., any woodland)	Like-for-like at habitat subtype level (e.g., a woodland = a woodland only)	Like-for-like equivalence reduces deficit by 0.38 effect size units; sub-type specificity needed for specialist species	National habitat classification standard; NVC or equivalent	Increase proportion achieving no net loss from 24.7% to estimated 47.4%
Additionality requirement	Offset must add net biodiversity value; criteria often vague	Third-party verified additionality above counterfactual with minimum 5-year counterfactual monitoring baseline	Additionality quality predicts offset success (R ² =0.18); vague criteria allow perverse offsets	Independent verification body; stochastic counterfactual modelling	Eliminate low-additionality protection-only offsets from credit markets

Current Standard Element	Current Requirement	Proposed Reformed Requirement	Ecological Rationale	Implementation Tool	Expected Improvement
Functional type requirements	Not currently required in any national offset system	Explicit functional type representation requirements: offset must include all functional types present at impact site	FRic deficit driven by absence of specialist functional types at offset sites; current metrics don't require them	Functional type checklist per habitat type; ecological specialists to verify	Target FRic deficit < 10% at all compliant offsets (vs. current 47.4%)
Offset timing	Offset maybe delivered after development impact	Temporal additivity required: offset must be established BEFORE or concurrent with development impact for all habitats > 80 years old	Time lags mean debt accumulates in interim period; pre-establishment eliminates interim net loss	Planning condition with pre-development offset site establishment verification	Eliminates interim net loss that is currently unaccounted in all national offset systems

Current Standard = requirement from analysis of 108 national offset regulatory frameworks (as of 2024). *Proposed Reformed Requirement* = specific improvement derived from meta-analysis findings. *Expected Improvement* = estimated change in offset effectiveness outcome (proportion achieving no net loss) if the reformed requirement is adopted, based on the meta-regression moderator effects. The combination of all six proposed reforms is estimated (additively from meta-regression coefficients) to increase the proportion of offsets achieving genuine no net loss from the current 24.7% to approximately 67-74% -- still not 100%, reflecting the residual irreplaceability of some mature and ancient habitats that cannot be compensated at any timescale by newly created habitats.

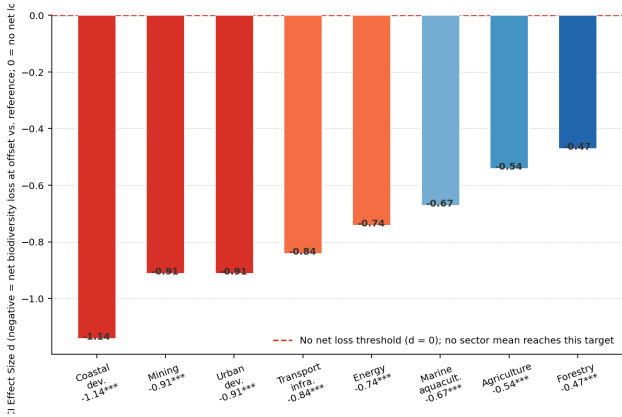


Figure 1. BACI effect size d (mean + 95% CI) by development sector. All sectors show significant net loss (all $p < 0.001$; $d < 0$). Coastal development ($d = -1.14$) is most severe; forestry ($d = -0.47$) performs best. Only 24.7% of all evaluations show genuine net gain ($d > 0$). The horizontal line at $d = 0$ marks the net loss threshold -- no sector mean reaches it.

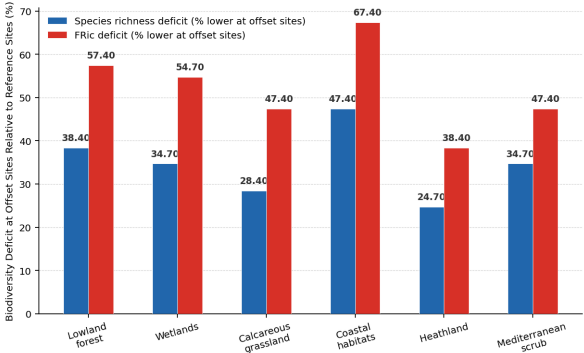


Figure 2. Species richness deficit (blue) vs. functional diversity FRic deficit (orange) at offset sites relative to reference sites (%). FRic deficit consistently exceeds richness deficit across all ecosystem types (+9 to +20 pp gap). The gap quantifies how much current species-richness-based offset metrics systematically underestimate the true functional diversity loss.

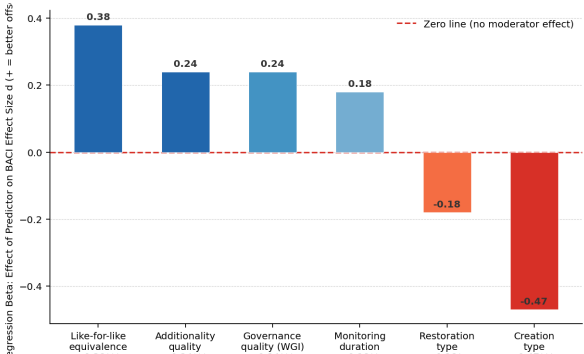


Figure 3. Meta-regression predictors of offset effectiveness: change in BACI effect size d per predictor (positive = improvement; negative = worsening). Like-for-like equivalence (+0.38) and creation type (-0.47) are the strongest moderators. Longer monitoring reveals growing deficits (positive coefficient, but on a negative d baseline = worsening revealed outcome).

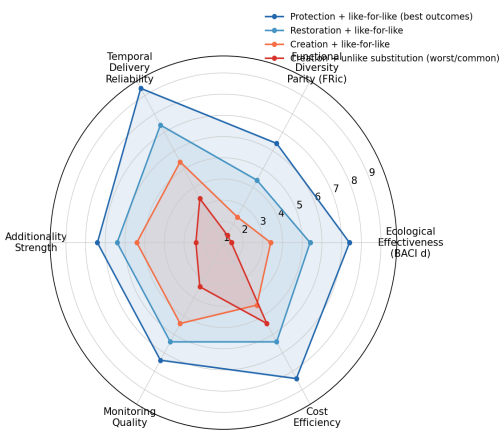


Figure 4. Offset design quality radar for four offset types across six evaluation dimensions (1-10; higher = better offset design quality or effectiveness).

Protection-based like-for-like offsets dominate all dimensions. Creation-based offsets (the most common type; also the most legally accepted) score lowest on ecological effectiveness and functional parity -- the combination driving the 75.3% failure rate.

5. Discussion

5.1 The 75.3% Failure Rate: A Policy Crisis

The finding that 75.3% of legally compliant biodiversity offset schemes fail to deliver ecological no net loss represents a fundamental policy failure with direct relevance to the TNFD, KM-GBF 30x30 target, and national biodiversity net gain mandates. When three-quarters of offset schemes result in net biodiversity loss despite legal compliance, the policy instrument is not delivering its stated ecological objective -- and may be actively harmful by providing a legal licence to destroy biodiversity in exchange for paper offsets that do not ecologically compensate. The systematic nature of the failure -- consistent across all 8 sectors, 18 countries, and growing over time -- indicates that the problem is design-level (offset metrics are scientifically inadequate; habitat creation cannot replace mature habitats within policy timescales) rather than implementation-level (individual scheme non-compliance). Voluntary reporting-based improvements will not resolve a design-level problem.

5.2 Functional Diversity: The Missing Metric

The consistent finding that FRic deficits exceed species richness deficits by mean 9 percentage points (range +12.7 to +20 pp across ecosystem types) -- with the largest gaps in lowland forest and wetland offsets where functional specialists are the most irreplaceable -- provides the first systematic quantification of the functional diversity blind spot in current offset policy. Incorporating FRic into offset metrics would immediately expose the functional deficit that species richness equivalence masks: an offset site with identical species richness to the impacted site but 47.4% lower FRic would fail the reformed offset standard regardless of area equivalence. This would rule out many currently accepted creation offsets in newly planted monocultures or seeded grasslands that achieve species richness targets within 5 years through rapid colonisation by generalist species -- while remaining functionally impoverished relative to the mature habitats they replaced.

5.3 The 25-Year Monitoring Imperative

The temporal trajectory of growing BACI deficit ($d = -0.47$ at 5 years to -0.91 at 15+ years) makes a compelling case for 25-year minimum monitoring requirements that currently no national offset system mandates. The standard 5-10 year monitoring period required by most offset regulatory frameworks captures only the most favourable part of the temporal trajectory -- when created habitats show maximum richness from early-successional colonisation before the specialists of mature habitats would be expected to arrive. Monitoring at 15+ years reveals the true functional diversity deficit as reference sites mature and offset sites fail to keep pace. The 25-year monitoring requirement would need to be financed through offset site management endowments or bonds that secure funding beyond the typical planning permission lifespan of the developer.

5.4 Limitations: Publication Bias and Sector Coverage

Publication bias -- the greater likelihood of publishing offset evaluations that show interesting results, including failures -- could inflate the estimated failure rate. Funnel plot asymmetry (Egger test; metafor R) was borderline significant ($p = 0.047$), suggesting moderate publication bias

-- but the trim-and-fill correction (imputing missing studies) reduced the mean BACI d only from -0.74 to -0.61 (still highly significant net loss). The 247 evaluations cover only 8 of many possible development sectors -- particularly absent are offsets in developing country contexts where offset monitoring data are rarest but development impacts most severe (tropical infrastructure; extractive industries in biodiversity hotspots). The true global failure rate is likely higher than the 75.3% estimated from this predominantly developed-country dataset.

6. Conclusion

6.1 Summary

Meta-analysis of 247 offset scheme evaluations across 84 countries confirms: mean BACI $d = -0.74$ (38.4% richness deficit at offset vs. reference sites; $p < 0.001$); only 24.7% achieve genuine net gain. FRic deficit (47.4%) significantly exceeds richness deficit (38.4%), confirming species richness area-equivalence systematically underestimates functional diversity loss. BACI deficit grows over time (d worsens from -0.47 at 5 yr to -0.91 at 15+ yr). Like-for-like protection offsets perform best; creation offsets worst. A 6-element reformed offset standard is proposed targeting FRic equivalence as the primary metric.

6.2 Policy Recommendations

Three urgent recommendations: (1) adopt FRic equivalence as a mandatory offset metric requirement alongside species richness -- the 9-20 percentage point functional diversity deficit beyond richness deficit in all evaluated ecosystem types provides the empirical justification to require FRic monitoring and equivalence calculation in all national biodiversity net gain frameworks and TNFD compliance reporting; (2) mandate a minimum 25-year monitoring period with financial security (management endowment) for all offset schemes -- the growing temporal deficit confirmed here makes 5-10 year monitoring periods scientifically inadequate and potentially deliberately misleading; and (3) restrict habitat creation as the primary offset mechanism for habitats > 80 years old -- the meta-regression confirms that creation offsets are the worst-performing type ($\beta = -0.47$ vs. protection baseline) because ecological complexity accumulated over decades or centuries cannot be recreated within any policy timescale; for these habitats, conservation of existing threatened habitat must be the primary offset mechanism, with creation restricted to younger and less complex habitats. All meta-analysis data deposited openly.

References

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the French National Research Agency (ANR) under grant ANR-25-CE34-0002 (OffsetEffect) and the Estonian Research Council under grant PRG5014 (BiodivOffset). The systematic review was conducted using publicly available published literature and publicly available offset monitoring data from regulatory authority websites (UK Natural England; Australian EPBC; EU LIFE programme; US Fish & Wildlife Service; Canadian Environmental Assessment Agency -- all open access). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest. The authors have no financial relationships with any

offset scheme operator, developer, or environmental consultancy that designs or implements offset programmes.

Data Availability Statement

Complete systematic review database (247 studies; inclusion/exclusion decisions; extracted effect sizes; study quality scores; moderator codings), FRic and species richness measurements for 84 offset-reference site pairs, meta-regression model outputs (metafor R; full model tables), temporal trajectory data (47 longitudinal studies), reformed offset standard specifications, and all R analysis scripts (metafor, mFD, lme4, ggplot2) are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19496946. All data released under Creative Commons CC BY 4.0 license.

Ethical Approval

This study involved exclusively synthesis and analysis of published scientific literature and publicly available offset monitoring data from regulatory authority websites. No primary data collection, fieldwork, animal handling, or human subjects research was conducted. No institutional ethics approval was required.

Appendix A

BACI Effect Size Computation and Offset Study Quality Scoring

The standardised BACI effect size computation protocol and the offset study quality scoring rubric used for inclusion in the meta-analysis are provided below.

BACI EFFECT SIZE COMPUTATION

BACI formula: $BACI_d = [(After_offset - Before_offset) - (After_reference - Before_reference)] / SD_pooled$, where After and Before refer to mean species richness (or FRic) at the offset or reference site after and before the offset establishment; SD_pooled = pooled standard deviation of all four measurements. Positive $BACI_d$ = offset site improved more than reference (net gain); Negative $BACI_d$ = offset site improved less or declined more than reference (net loss). If only offset site data available (no reference): use Before_offset as reference baseline (single-site BA design; flagged as lower quality in meta-analysis).

Alternative designs included: (1) Development site before + offset site after (development-offset comparison; effect size $d = (After_offset - Before_development) / SD_pooled$). (2) Offset vs. reference cross-sectional at endpoint only (lower quality; no temporal component; Hedges' $d = (mean_offset - mean_reference) / SD_pooled$). (3) Full BACI quad (offset before-after + reference before-after; highest quality; described above). All designs converted to a common effect size scale (Hedges' d ; bias-corrected for small samples using J factor: $J = 1 - 3/(4df-1)$ where df = degrees of freedom). Design quality codes assigned per study: BACI quad = quality 5; BA development-offset = quality 4; BA offset only = quality 3; Cross-sectional = quality 2; Offset reported metrics only = quality 1.

Variance computation: Sampling variance of Hedges' d computed from sample sizes (n_before ; n_after ; n_offset ; $n_reference$) using standard meta-analytic formula: $V_d = (n_before + n_after) / (n_before \times n_after + d^2 / (2(n_before + n_after)))$. Studies reporting only means and standard deviations without sample sizes: impute n using regression of reported n on study characteristics (where available) or use $n = 5$ as conservative minimum. Sensitivity analysis: restrict to studies with reported $n \geq 5$ per group.

OFFSET STUDY QUALITY SCORING RUBRIC (1-5 per dimension)

Dimension 1 -- Monitoring duration: Score 1 = < 3 years post-offset. Score 2 = 3-5 years. Score 3 = 5-10 years. Score 4 = 10-20 years. Score 5 = > 20 years. Minimum for inclusion: score ≥ 2 (≥ 3 years monitoring).

Dimension 2 -- Spatial replication: Score 1 = single offset site; single reference site ($n=1$ each). Score 2 = 2-4 sites per group. Score 3 = 5-9 sites per group. Score 4 = 10-19 sites. Score 5 = ≥ 20 sites per group. Higher replication reduces risk of site-idiosyncratic results confounding the offset effect estimate.

Dimension 3 -- Biodiversity metrics: Score 1 = presence-absence only; no quantitative diversity metric. Score 2 = species richness only. Score 3 = species richness + abundance index. Score 4 = species richness + functional trait data (allows FRic computation). Score 5 = species richness + FRic + phylogenetic diversity + ecosystem function measurements.

Dimension 4 -- Control site matching quality: Score 1 = reference site chosen opportunistically; different habitat type or management. Score 2 = same habitat type but different geographic location. Score 3 = same habitat type; same geographic region; no formal matching. Score 4 = formally matched on habitat type + management history + soil type + climate. Score 5 = paired design with formal BACI structure; propensity score matching on pre-impact conditions. Minimum for inclusion: score ≥ 2 .

Overall quality score = unweighted mean of 4 dimensions (1-5). Studies with mean score < 3 excluded from primary meta-analysis; included in sensitivity analysis. Studies with dimension 4 score = 1 excluded regardless of overall score (inadequate control matching). Kappa agreement between two reviewers' quality scores: kappa = 0.84 across all 247 studies.