

RESEARCH ARTICLE

Twenty Years of Evidence-Based Conservation: Progress, Promise, and Future Directions

Temporal and spatial trends in study designs used to understand what works in conservation

Amelia Staples¹ | William H. Morgan²  | Lynn V. Dicks²  | Silviu O. Petrovan²  |
 Rebecca K. Smith²  | William J. Sutherland²  | Alec P. Christie^{1,2} 

¹Faculty of Natural Sciences, Centre for Environmental Policy, Imperial College London, South Kensington, London, UK

²Department of Zoology, University of Cambridge, Cambridge, UK

Correspondence

Alec P. Christie

Email: a.christie@imperial.ac.uk

Funding information

Imperial College London, Imperial College Research Fellowship; Cartier for Nature; Arcadia; MAVA Foundation; A.G. Leventis Foundation

Handling Editor: Paul Franklin

Abstract

1. Robust study designs are essential for estimating the effects of conservation actions, but a range of evaluation designs are used that differ markedly in their robustness to bias. However, it is unclear whether robust designs are becoming more frequently used or how usage varies across subject areas and geographies.
2. We analysed 7356 English and 1183 non-English language primary studies that tested the effects of conservation actions from the Conservation Evidence database. We classified each study's methodological design to one of five broad types (After, Before–After [BA], Control–Impact [CI], Before–After–Control–Impact [BACI] and randomised experiments) and fitted a Bayesian multinomial logistic regression, with study-level random effects, to test how design usage varies by year, national income, biodiversity intactness, language and subject area.
3. The evidence base has shifted gradually from dominance by simple After designs (1970–2023: 74% to 33%) towards greater use of CI (15% to 39%) and more robust designs: randomised experiments (0% to 15%) and BACI (1% to 6%). Design use was uneven across subjects, with randomised experiments more common in habitat and soil-focused topics (e.g. farmland), and After or BA designs more common in animal-focused topics (e.g. birds and reptiles). Studies in non-English languages and lower-income countries were more likely to use After and BA designs and less likely to use randomised experiments and CI designs.
4. Geographical coverage of robust designs was limited: randomised experiments and BACI studies occurred in few countries, and many biodiverse and environmentally degraded countries had few or no evaluations using any design. Despite modest improvements over time, several structural and contextual constraints (linked to subject, language and resources) continue to shape where rigorous evaluation is feasible.
5. *Synthesis and applications:* Conservation is shifting towards using more robust designs to evaluate actions, but these remain rare. We suggest embedding study design choice more strategically within adaptive management and evidence-based

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Journal of Applied Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

conservation practice, encouraging cost-efficient, context-dependent upgrades to more robust designs. This will require targeted capacity building, researcher–practitioner partnerships and longer-term funding—especially in biodiverse, lower-income countries with few studies of any design, let alone more robust ones.

KEYWORDS

biodiversity conservation, causal inference, conservation action, conservation evidence, evidence-based conservation, experimental design, impact evaluation, study design

1 | INTRODUCTION

Rapid biodiversity loss is an urgent global challenge requiring efficient use of resources to implement effective conservation actions (Ceballos et al., 2017; Sutherland, 2022). To inform better practitioner and policymaker decisions, it is crucial that relevant and reliable evidence is made readily accessible on which actions work, where, how and under what conditions (Ferraro & Pattanayak, 2006; Pullin & Knight, 2003; Sutherland et al., 2004). However, evidence is only as reliable as the study designs used to generate it (Rubin, 2008): less robust designs can produce misleading estimates of the effects of actions that, in turn, could lead to selecting and implementing ineffective or even harmful actions (Christie, Abecasis, et al., 2020; Cooke et al., 2023; Lemieux et al., 2021). As a result, there have been growing calls to raise standards in evaluating the effects of conservation actions (Asquith, 2020; Ferraro et al., 2023; Jones & Shreedhar, 2024; Ockendon et al., 2021).

Several broad types of study designs are typically used to evaluate conservation actions (see Figure 1), each relying on different causal assumptions and thus different vulnerabilities to bias (Christie, Abecasis, et al., 2020; Christie et al., 2019; De Palma et al., 2018). The fundamental goal of most tests of conservation actions is to establish the effect they had on an outcome of interest (Ferraro, 2009)—this requires estimating the ‘counterfactual’ or what would have happened without the action.

Randomised experiments (typically called ‘Randomised Controlled Trials’ or RCTs; Figure 1) randomly assign experimental units to treatment and control groups, and when feasible to implement, provide the strongest basis to reliably estimate the counterfactual and thus for inferring causal effects (Pynegar et al., 2021). Eliminating potential bias from confounding variables (potentially associated with the treatment and outcomes of interest; O’Garra et al., 2025) is a crucial foundation of causal inference—without randomisation, other types of designs must use alternative approaches to minimise this bias.

Non-randomised designs in conservation include the Before–After–Control–Impact (BACI) design (Figure 1), which compares temporal change in treated and control units (Thiault et al., 2017; Underwood, 1994) and estimates the treatment effect via a difference-in-differences approach. BACI designs rely on the parallel-trends assumption that post-treatment trends in the control and treatment groups are parallel (Wauchope et al., 2021); when implemented

well, these are generally considered more robust than simple designs—see below (Christie, Abecasis, et al., 2020; Christie et al., 2019). Control–Impact (CI) designs compare treated and untreated units after the action occurs but do not use pre-treatment data (Figure 1), making them vulnerable to selection bias unless confounders are carefully controlled, for example through matching or blocking (Christie, Abecasis, et al., 2020; Morgan & Winship, 2007; Pynegar et al., 2025). Before–After (BA) designs compare outcomes in treated units before and after an action occurs without a control group, implicitly assuming that no change would have occurred in the absence of the action (Figure 1). Finally, After designs measure only post-treatment outcomes in treated units and attribute all observed differences to the action (Figure 1). The assumptions of BA and After designs are rarely satisfied in dynamic ecological and socio-ecological systems (Adams et al., 2019; Christie et al., 2019).

Choosing among these designs is a central decision in evaluating actions, especially when they have large ecological, social or financial consequences. More robust designs (e.g. randomised experiments, BACI) typically require greater investment, planning and sometimes difficult ethical or political decisions. On the other hand, simple designs (e.g. After, BA) are often easier to implement but provide much weaker evidence about causal effects. In practice, conservation practitioners frequently operate under conditions which can limit the feasibility of randomisation, baselines or long-term studies (Curzon & Kontoleon, 2016; Pynegar et al., 2021), including working with: tight budgets and time constraints; complex systems and rare, elusive or mobile species (e.g. marine mammals; Hordern et al., 2024); and actions embedded in people’s livelihoods. Adaptive management and decision-analytic frameworks therefore emphasise designing evaluations that are as rigorous as is realistically possible in a given context—whilst recognising that design choices must be made explicitly and transparently, considering key constraints and the potential consequences of misinforming decisions (Mapstone, 1995).

Previous work has highlighted persistent reliance on simple designs in conservation and the limited use of more robust designs (Bladon et al., 2025; Christie, Amano, et al., 2020) relative to disciplines such as medicine and parts of the social sciences (Christie, Abecasis, et al., 2020). Barriers include ethical concerns about withholding actions, difficulties in random allocation, short-term funding cycles and limited analytical capacity, particularly outside

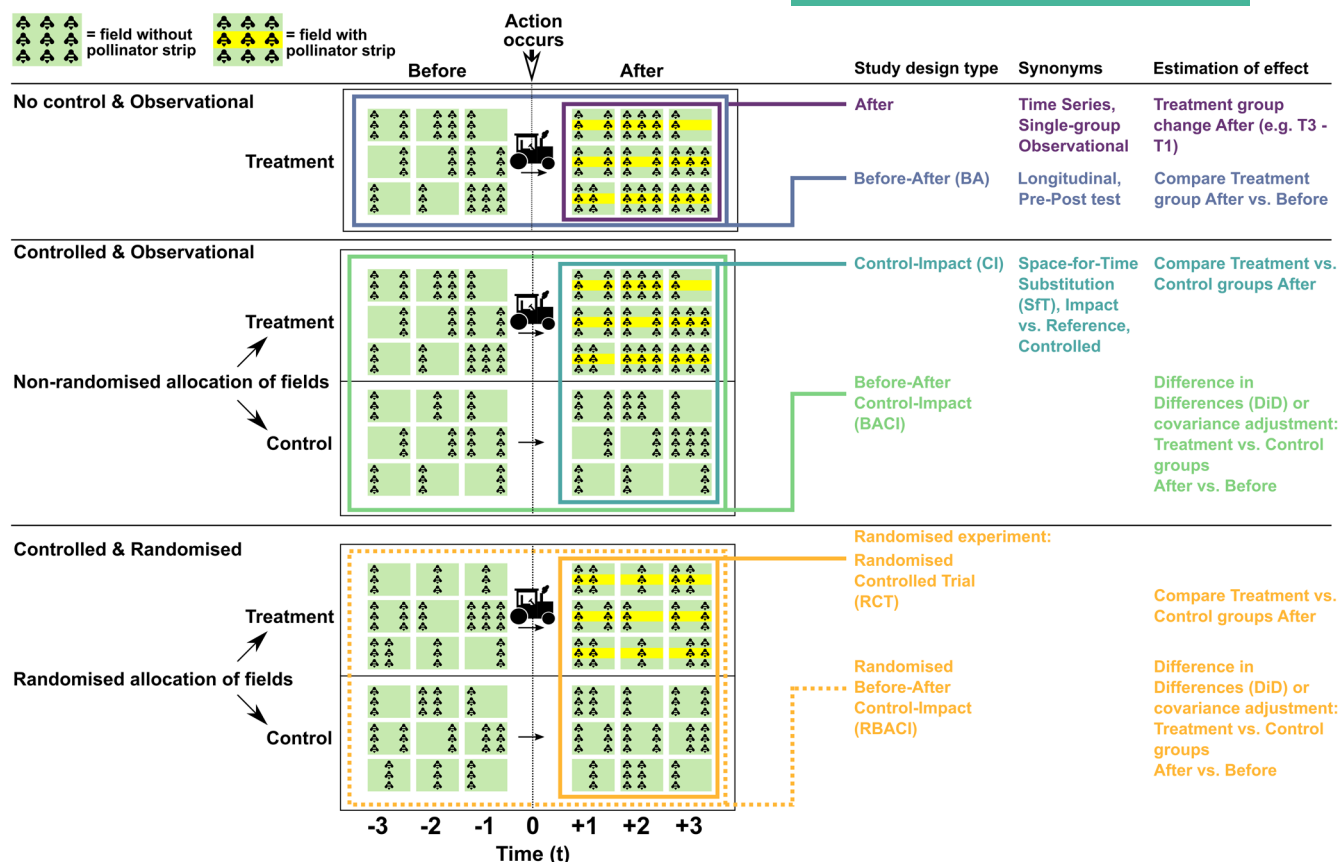


FIGURE 1 Comparison of common study design types used to estimate the effect of conservation actions. The hypothetical example shows a study evaluating how creating pollinator strips in arable fields affects the abundance of bees and different study designs that could be used. Fields assigned to receive the action (treatment) and those without the action (controls) may be surveyed at multiple time points before and after the action occurs (at time $t=0$). Each coloured box outlines the subset of data that would typically be collected and analysed under a particular design type (After, Before-After, Control-Impact, Before-After-Control-Impact or randomised experiment). Adapted from Christie, Abecasis, et al. (2020).

well-resourced institutions and countries (Pynegar et al., 2025; Tinsley-Marshall et al., 2022). There is also some evidence that study design use varies by subject areas and geography; Christie et al. (2021) found that tests of amphibian and bird actions using BACI and RCT designs were typically concentrated in the Global North, whilst Amano et al. (2021) found differences in the prevalence of different study designs between the English and non-English language scientific literatures. However, we currently lack a comprehensive, quantitative assessment of how study design use has changed over time and how it varies among conservation subjects, countries and languages.

Here, we use the Conservation Evidence database—at time of writing, comprising 7356 English language (Sutherland et al., 2019) and 1183 non-English language primary studies (Amano et al., 2021) that quantitatively test conservation actions—to explore the temporal and spatial patterns in study design use across conservation actions. We classify each study according to five broad design types (After, BA, CI, BACI, randomised experiments) and fit a hierarchical multinomial model to: (1) quantify how the relative use of these designs has changed over time, (2) identify differences between language types, subject areas and geographies, and where certain

designs are particularly common or scarce and (3) explore how these patterns can inform decisions about study design within the context of adaptive management and evidence-based conservation. This study will also explore whether study design use aligns with conservation needs, where evidence gaps and weaknesses are, and how strategic improvements in study design could most improve the reliability and usefulness of tests of conservation actions.

2 | MATERIALS AND METHODS

2.1 | Conservation Evidence database

We used data from the Conservation Evidence project (<https://www.conservationevidence.com>), a publicly available database to inform conservation decisions by collating studies that test the effectiveness of conservation actions (Sutherland et al., 2019). Studies are identified from extensive manual searches of both broad and subject-specific conservation journals, both English and non-English language, using a standardised protocol. This process involves scanning of titles and abstracts for an entire journal series, rather than

keyword search terms, in a process of discipline-wide literature searches (Sutherland et al., 2019). Relevant studies are grouped by subject area into 'synopses' (e.g. bird, reptile, marsh and swamp conservation) and by the conservation action that is tested—in a process called 'subject-wide evidence synthesis' (Sutherland et al., 2019). Studies are reported in a short narrative summary (~200 words) that includes the main results, methods and study design(s), plus other key metadata (e.g. location, taxa)—a single study may be summarised several times if different study designs are used or different conservation actions are tested.

For this study, we first excluded all secondary studies (i.e. reviews and meta-analyses). The non-English language component of the database includes studies from searches in 16 languages: Arabic, French, German, Hungarian, Italian, Japanese, Korean, Persian, Polish, Portuguese, Russian, simplified Chinese, Spanish, traditional Chinese, Turkish and Ukrainian (Amano et al., 2021). These studies are currently being integrated into different Conservation Evidence 'synopses' so were not yet assigned to a specific subject area—we therefore assigned these to their own subject area: 'non-English language studies'.

The Conservation Evidence database already included structured data for studies on their publication year, subject area (synopsis), the country where they were conducted and publication language.

2.2 | Study design classification

Conservation Evidence uses a standardised glossary of methodological terms (e.g. 'controlled', 'before–after', 'randomised') to describe each study's design(s)—for example see Thornton et al. (2025, p. 18). Whilst the non-English language studies had already been labelled with a broad study design type (after work by Amano et al. (2021)), we needed to add labels for the English language studies in the database. We applied two complementary approaches (Table S1) to do this by searching for key terms in the narrative summaries of studies using regular expressions in R statistical software (R Core Team, 2024). First, we detected key terms and their combinations in the first sentence of each study summary as they should include a standardised description here. Second, as older study summaries sometimes lack these standardised descriptions, we identified where key terms typically appeared within these summaries and restricted searches to the first 150 characters—which we determined captured all relevant design terms, whilst reducing false matches. If terms were not detected with either approach, the study summary was classified as an 'After' design. Using both approaches helped to avoid misclassifying terms that referred to other aspects of a study (e.g. 'control' describing the control of an invasive species rather than a control group) and accounted for stylistic differences in summaries between older and more recent synopses.

We compared design labels assigned by the two approaches, resolving any discrepancies by manually checking study summaries and then mapped the resulting terms to five broad design categories

following Christie et al. (2019) and Christie, Abecasis, et al. (2020). To assess classification accuracy, we manually checked a random sample of 850 unique studies (981 study summaries—approximately 10% of all studies). Only 12 summaries had misclassified designs (affecting 11 unique studies), giving an overall accuracy of 99%; we corrected these observed errors before analysis. These corrections were: three After designs reclassified as CI, four After, one CI and one BA reclassified as randomised experiments, and two After and one CI reclassified as reviews or meta-analyses—which were excluded accordingly.

2.3 | Analysis

We analysed how study design use varied with publication year, national income (GDP per capita), levels of biodiversity degradation (Biodiversity Intactness Index, BII; Newbold et al., 2016), publication language and subject area. To explore how study design use varied geographically, we chose to use GDP per capita (constant 2015 US Dollars from the World Bank via the WDI package in R; Arel-Bundock, 2025) and BII estimates for studied countries as explanatory variables because these data were widely available on a global scale and represented proxies for levels of 'research capacity' and 'conservation need', respectively.

We obtained country-level estimates of BII from Phillips et al. (2021). Historical BII estimates were available for 1970, 1980, 1990 and 2000–2014, so we linearly interpolated these values to annual resolution (1970–2014). For 2015 onwards, only projections of BII were available, and so, we chose the SSP2-RCP4.5 projection as this is widely used as a moderate socio-economic pathway scenario case in climate impact assessments (Riahi et al., 2017).

We fitted a Bayesian multinomial logistic regression with a random intercept for study ID because some studies used more than one design or appear in multiple synopses (453 studies and 516 studies, respectively). The response variable was study design, coded as a nominal factor with five levels ('After', 'BA', 'CI', 'BACI', 'Rand.exp.') and using 'After' as the reference category. The continuous explanatory variables were publication year, log-transformed GDP per capita and national-level BII, each centred and scaled to mean zero and unit variance. We also included publication language (English vs. non-English) and subject area (Conservation Evidence synopsis) as categorical explanatory variables. We excluded five subject areas (Sustainable Aquaculture, Bee Conservation, Grassland Conservation, Natural Pest Control and Primate Conservation) that did not include all five study designs from the multinomial model so that all response variable levels were present in each subject.

We fitted the hierarchical multinomial model in R using brms (Bürkner, 2017) and cmdstanr (Gabry et al., 2025), specifying a categorical family with a multinomial logit link. For each non-reference study design category (*c*), the linear predictor included fixed effects of the covariates and a study-level random intercept to account for dependence among multiple designs reported by the same study:

$$\begin{aligned}\text{logit}\{P(Y_i=c)\} = & \alpha_c + \beta_{c,\text{year}} \cdot \text{year_sc}_i \\ & + \beta_{c,\log_gdp_scaled} \cdot \log_gdp_scaled_i \\ & + \beta_{c,bii_scaled} \cdot bii_scaled_i \\ & + \gamma_c^T \mathbf{x}_i, \text{language, synopsis} + u_{j[c]},\end{aligned}$$

where Y_i is the study design for observation i and $u_{j[c]} \sim \mathcal{N}(0, \sigma_u^2)$ is a random intercept for study j . We used weakly informative normal priors for fixed effects (Lemoine, 2019) and exponential priors for the standard deviation of the study-level random effects (Betancourt, 2017; Gelman et al., 2008).

Models were estimated using four Markov chains with 4000 iterations per chain (2000 warm-up). Before fitting, we assessed collinearity among predictors using variance inflation factors from an ordinary least-squares approximation; VIF values ranged from 1.1 to 4.6, indicating that multicollinearity was low. We confirmed chain convergence using the Gelman–Rubin diagnostic ($\hat{R} < 1.05$) and visual inspections of trace plots, and adequate parameter estimation required an effective sample size >100 per parameter. Posterior predictive checks were conducted by comparing the observed outcomes with simulated datasets from the posterior predictive distribution. We treated effects as statistically significant if their 95% credible intervals (CrIs) did not overlap with zero.

To aid interpretation, we obtained marginal predicted probabilities for each explanatory variable by computing posterior expected probabilities for each study design whilst averaging over the observed distributions of the other variables and study-level random effects. We summarised these posterior draws as means and 95% CrIs. To compare English and non-English language studies, we calculated, for each design, the posterior mean difference in predicted proportion (non-English–English), its 95% CrI, and the posterior probability that this difference exceeded zero.

All analyses were conducted in R (version 4.4.0; R Core Team, 2024). The code and data relating to these analyses is freely available on Zenodo: <https://doi.org/10.5281/zenodo.17192808>. Ethics approval was not required for this study as there were no animal or human research subjects involved.

3 | RESULTS

3.1 | Temporal trends

We fitted a Bayesian multinomial logistic regression, with a random intercept for study, to estimate how study design (After, BA, CI, BACI, randomised experiment) varied by publication year, national income (GDP per capita), BII, language type (English vs. non-English) and subject area (synopsis). All Markov chains mixed well ($\hat{R} \leq 1.03$ for all parameters; effective sample sizes typically >1000 ; Table S2). Study-level random-effect standard deviations were clearly non-zero (Table S2), indicating substantial between-study heterogeneity.

After studies had historically made up the greatest raw proportion of studies but this has since declined (1970: 73.8%; 2023:

33.4%—see Figure 2 and Figure S1) whilst the raw proportion of CI studies has substantially increased, becoming the most common design in recent years (1970: 15.0%; 2023: 39.0%). The raw proportion of BA studies has stayed relatively unchanged (1970: 12.5%; 2023: 12.9%), whilst there has been a modest increase for BACI studies (1970: 1.3%; 2023: 6.2%) and a substantial increase for randomised experiments (1970: 0%; 2023: 15.0%), although all these three designs still make up a relatively small proportion of all studies (Figure 2; Figure S1).

These temporal changes to the composition of the evidence base were mirrored in the raw proportion of studies published each year (Figure 2) and by estimates from the multinomial model, averaging over other variables—all changes over time were statistically significant (both 95% CrIs >0 ; Table S2). Predicted proportions of After studies declined between 1970 and 2023 from 49.3% (95% CrIs: 43.7%–54.9%) to 30.7% (27.9%–33.5%), whilst CI studies increased from 34.8% (29.1%–40.6%) to 44.9% (40.9%–48.3%) over the same period. Predicted proportions of BA studies increased marginally over time (8.6%–9.2%), as did BACI studies (4.0%–5.0%), whereas randomised experiments rose substantially from 3.3% (1.9%–5.0%) to 10.2% (8.0%–12.7%).

3.2 | Language differences

In the raw data, CI studies made up the greatest proportion of studies across languages and for English language studies (Table 1), whereas After studies were most common in non-English language studies. The raw proportion of CI studies was substantially lower in non-English language studies, and both BACI studies and randomised experiments represented only a small proportion of non-English language studies. The raw proportion of BACI studies was similar between English and non-English language studies, but randomised experiments were more common in English language studies, with the opposite pattern for BA (Table 1).

Language differences in study design use persisted after adjusting for other variables and multiple designs per study in the multinomial model (Table S2). Compared to English language studies, non-English language studies had higher predicted proportions for After studies (difference: 21.7%, 95% CrI: 17.3%–25.9%) and BA studies (difference: 7.2%, 4.7%–10.1%). In contrast, non-English language studies had lower predicted proportions of CI studies (difference: –20.0%, 95% CrI: –23.5% to –16.0%) and randomised experiments (difference: –7.6%, –8.7% to –6.6%). Differences for BACI studies were small and only marginally different from zero (difference: –1.3%, –2.5% to –0.12%), consistent with raw proportions between language types.

3.3 | Differences between subject areas

Study design use differed strongly among subject areas (Figure 3). Raw proportions showed that several animal-focused subjects

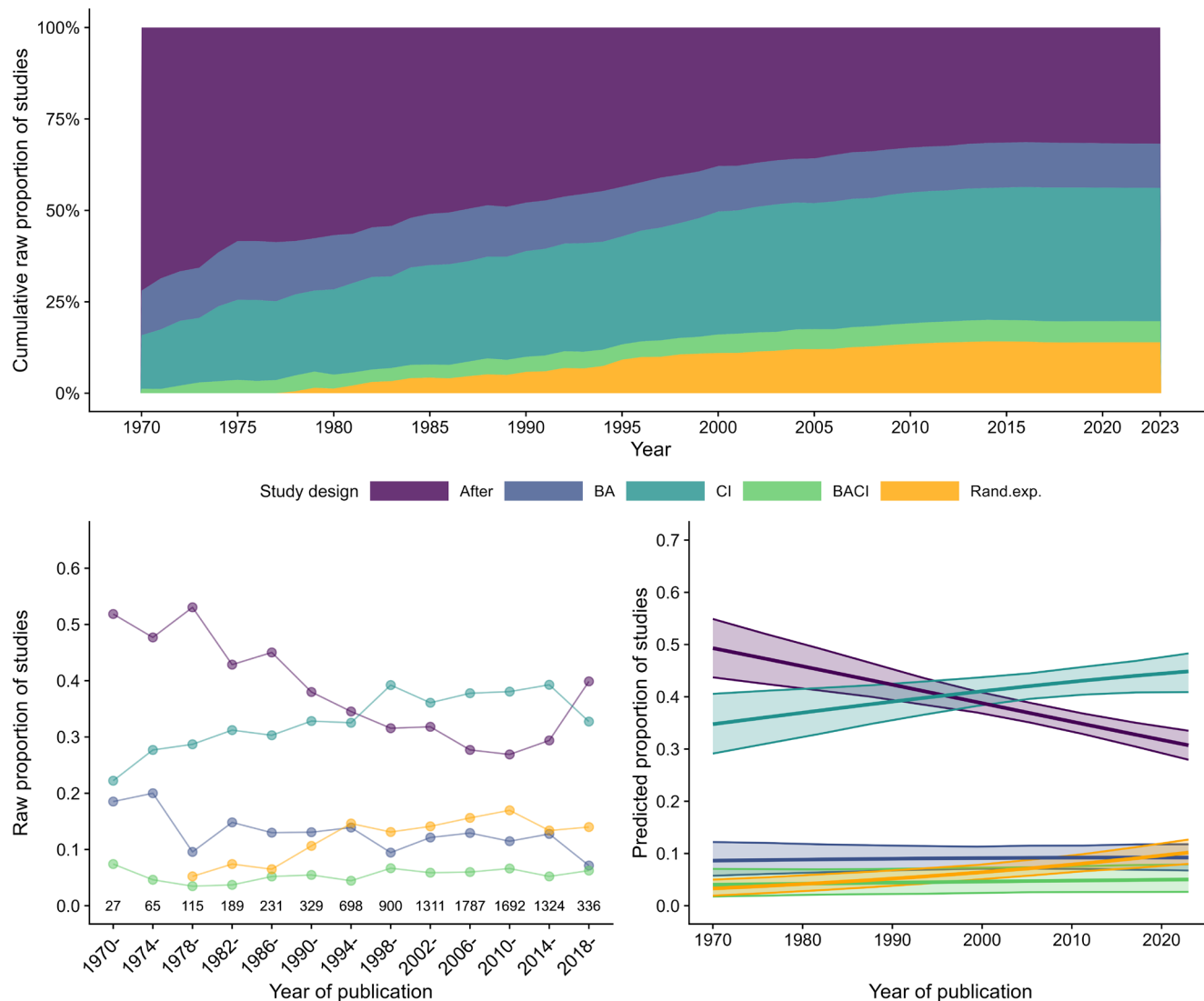


FIGURE 2 Temporal trends in the use of study designs in the Conservation Evidence database (1970–2023). Top: Cumulative raw proportion of all studies using a given design in each year—full dataset (1912–2023) in Figure S1. Bottom left: Raw proportions of designs in 4-year bins (note that the last bin is: 2018–2023) with numbers at the bottom showing the total number of studies per bin. Bottom right: Predicted proportion of studies from the multinomial model (adjusting for other variables and multiple designs per study) with 95% CIs as shaded bands.

Study design	Overall		English language		Non-English language	
	% Studies	Studies	% Studies	Studies	% Studies	Studies
After	33.8	2884	31.9	2346	45.5	538
Before–After	12.9	1098	11.8	869	19.4	229
Control–Impact	38.7	3308	40.9	3009	25.3	299
Before–After–Control–Impact	6.1	524	6.4	468	4.7	56
Randomised experiment	14.8	1265	16.3	1202	5.3	63
All	8539		7356		1183	

TABLE 1 Raw proportions and numbers of studies that used each study design in the Conservation Evidence database, both overall and split by whether studies were published in English or non-English languages. Note studies may use more than one design.

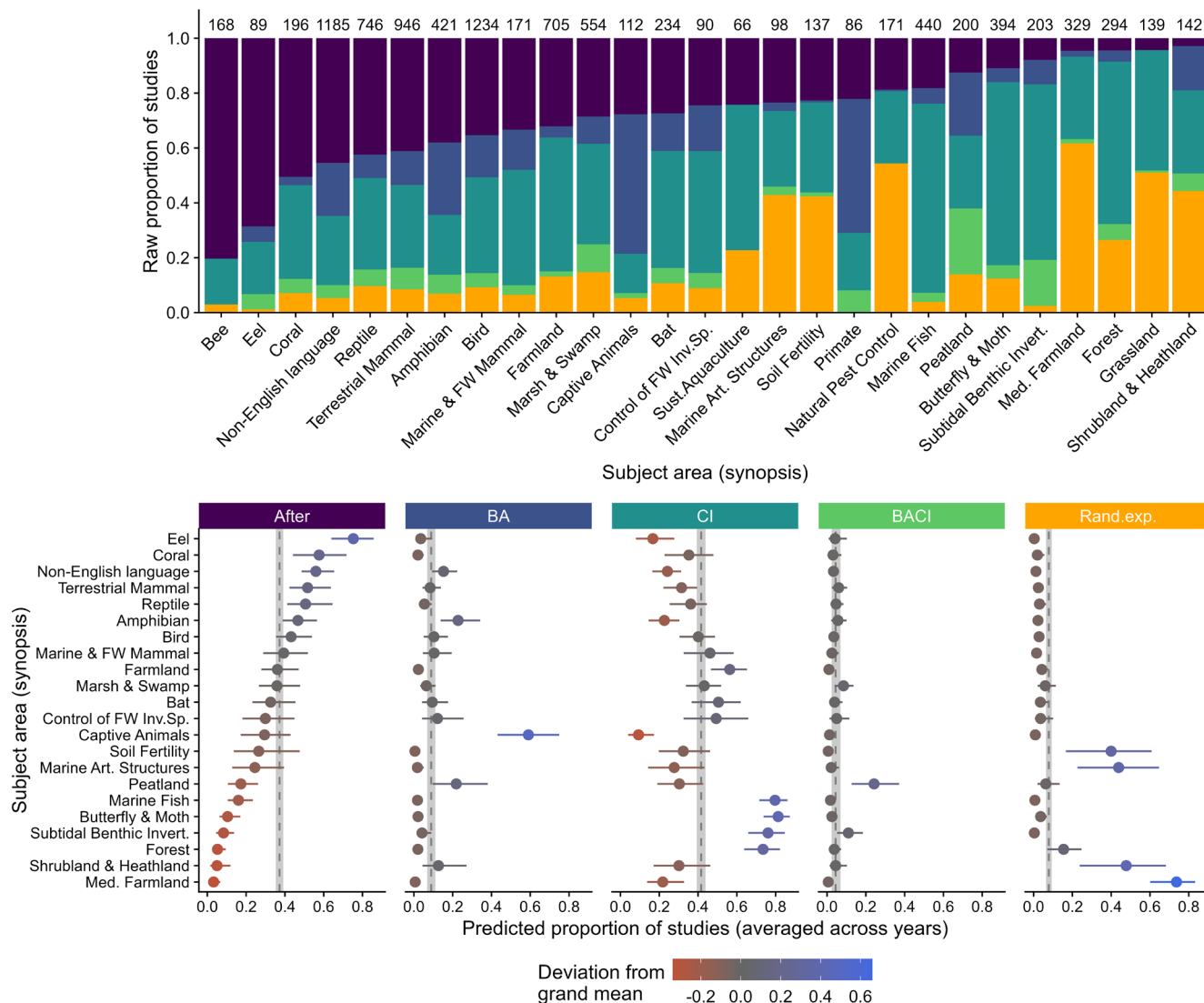


FIGURE 3 Raw and predicted proportions of each study design by subject area (synopsis). Top panel shows raw proportions of designs (see colour labels in headers for bottom panel) in order of highest to lowest proportion of After studies from the Conservation Evidence database. BA=Before–After, CI=Control–Impact, BACI=Before–After–Control–Impact, Rand.exp. = randomised experiments. The total number of studies in each subject is displayed at the top of each bar. Bottom panel shows the predicted proportion of studies for each design across each subject area with 95% Credible Intervals (Cris) from the multinomial model—dashed lines and grey bands indicate the grand mean proportion of studies for each design and its 95% Cris across all subjects. Colours of points and 95% Cris represent direction and magnitude of deviation from overall grand means (red=negative deviation and blue=positive deviation).

(e.g. Bee, Eel, Reptile, Terrestrial Mammal, Amphibian and Bird Conservation) were dominated by After studies as the most common design (36%–81% of studies) with relatively few BACI or randomised experiments. In contrast, randomised experiments were far more common in habitat- and soil-focused subjects (e.g. Mediterranean Farmland, Natural Pest Control, Soil Fertility, Shrubland and Heathland, Grassland, Forest, Marine Artificial Structures). BACI designs were also relatively common in Peatland, Subtidal Benthic Invertebrate and Marsh and Swamp Conservation.

Multinomial model predictions confirmed these differences (Table S2; Figure 3), with the predicted proportion of After studies being higher than the grand mean in subjects such as Eel, Terrestrial Mammal and Reptile Conservation and lower than average in

habitat-focused subjects, particularly Mediterranean Farmland, Shrubland & Heathland and Forest Conservation. For randomised experiments, we observed the opposite pattern: proportions were markedly above the grand mean in Mediterranean Farmland and Shrubland & Heathland Conservation, as well as Soil Fertility and Marine Artificial Structures, and below average in many animal-focused subjects (e.g. Eel, Subtidal Benthic Invertebrates, Marine Fish and Captive Animals; Figure 3). BACI study proportions deviated only marginally from the grand mean, but were clearly above average in Peatland Conservation. CI study proportions were greatly above average in Marine Fish, Butterfly and Moth, Subtidal Benthic Invertebrates and Forest Conservation, whilst below average in Captive Animals, Amphibian, Mediterranean Farmland

and Eel Conservation (Figure 3). BA studies were above average in Captive Animal Conservation and to a lesser extent in Amphibian Conservation, but below average in a range of animal and habitat-focused synopses (e.g. Mediterranean Farmland, Forest, Marine Fish, Coral and Butterfly & Moth Conservation).

The relative contribution of different subject areas to the evidence base changed over time, but in a heterogeneous way (Figure S2)—some subject areas have been updated more recently with studies than others in the Conservation Evidence database. Birds (studies searched up to 2010) and Terrestrial Mammals (up to 2017) represented the largest subject area in terms of numbers of studies, with the former declining in relative share across the time series. Several other taxa-focused subjects, including Amphibians (studies searched up to 2012), Reptiles (2018) and Marine Fish (2018) increased in proportion, whilst some specialised subjects remained rare—for example Eels (2023), Sustainable Aquaculture (2012) and Control of Freshwater Invasives (2016). Habitat- and soil-focused subjects such as Farmland (studies searched up to 2010), Soil Fertility (2012), Mediterranean Farmland (2015), Peatland (2016),

Shrubland and Heathland (2016) and Subtidal Benthic Invertebrates (2017) also showed gradual increases in proportions.

3.4 | Geographical differences

The spatial distribution of study designs was relatively similar to the overall distribution of studies, with major concentrations in North America, Western Europe and Australasia (Figure 4). However, global representation differed as out of the 160 countries with studies, After studies were present in 82% of them, CI in 75%, BA in 60%, randomised experiments in 44% and BACI in 38% (Figure 4).

The overall temporal trends described earlier were also found across continents and national income groups (Figure 5), with a decline in the raw proportion of After studies over time, whilst CI studies and randomised experiments increased in proportion (although the latter remains relatively rare) and BACI and BA proportions remained relatively unchanged. The multinomial model estimated that higher-GDP countries had slightly higher predicted proportions of

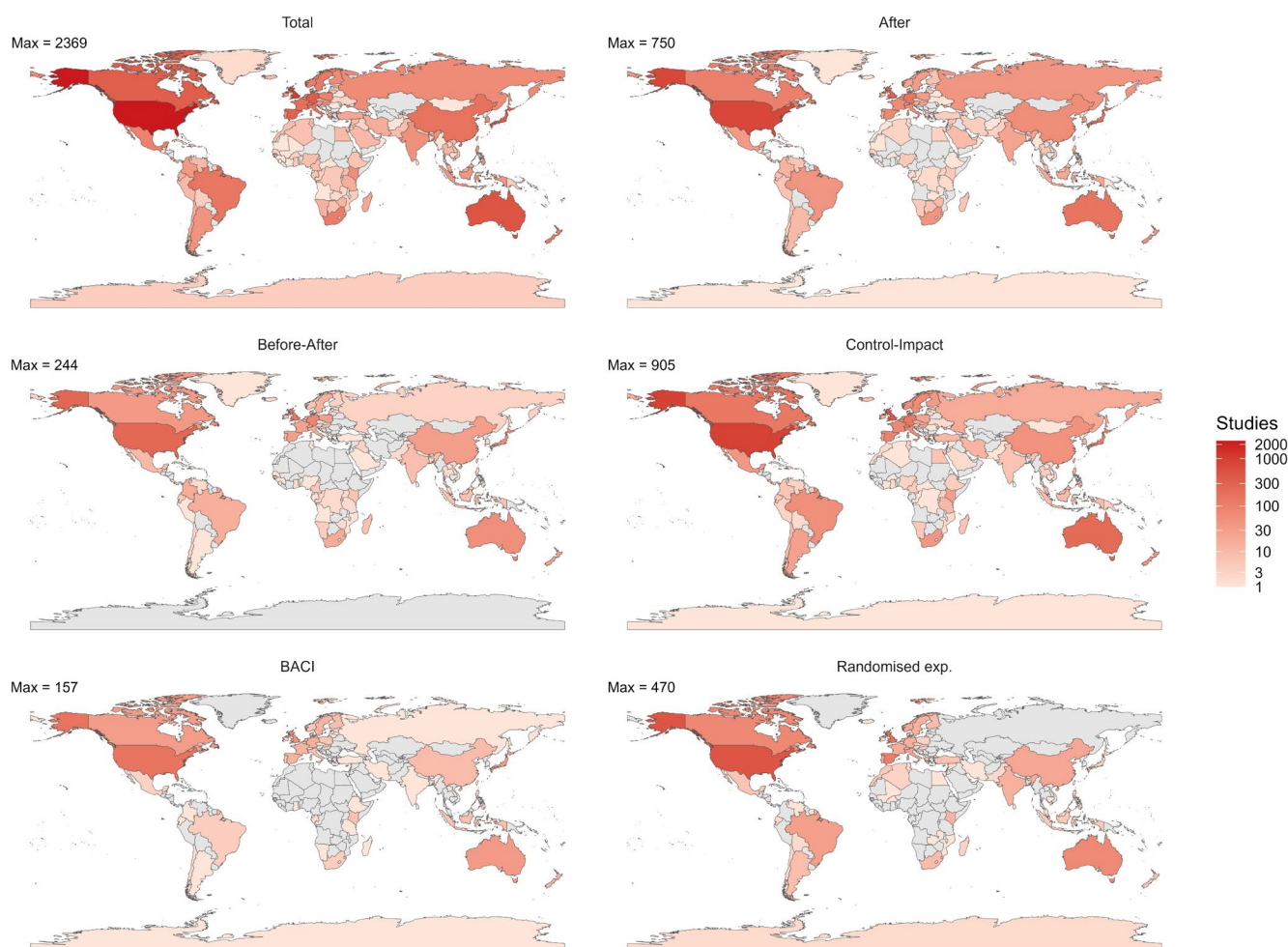


FIGURE 4 Global distribution of conservation action study designs showing the number of studies per country on a log-transformed scale for all designs combined (Total), After, Before-After, Control-Impact, Before-After-Control-Impact (BACI) and Randomised experiments. Countries with zero studies are coloured grey, whilst countries with studies are mapped from pale to dark red; annotations for each map show the maximum number of studies per country for each design.



FIGURE 5 Raw proportions of study designs from 1970 to 2023 by continent (top) and World Bank national income groups (bottom) binned into 5-year intervals (except the last interval: 2015–2023). Total numbers of studies for all designs for each continent or national income group are shown in the top left of each panel. BA = Before–After, CI = Control–Impact, BACI = Before–After–Control–Impact, Rand. exp. = randomised experiments.

CI studies—and particularly BACI studies and randomised experiments—than After studies. However, the magnitude of these effects was small and the predicted proportions of each design are broadly similar across GDP groups, with wide overlapping 95% Crls (Table S2).

The multinomial model suggested that the predicted proportion of designs was similar across levels of BII, with wide overlapping 95% Crls and small model coefficients for BII (Table S2). In countries with higher BII, predicted proportions were marginally higher for CI and marginally lower for BA relative to After studies. Figure 6 shows the lack of a clear relationship with study design, given that randomised experiments and BACI studies tend to be concentrated in relatively well-studied countries with moderate BII. In contrast, several countries with no data on BII, or relatively low or high BII, have few or no studies of any design, especially in Africa (Figure 6). Similar patterns were found for other study designs (Figure S3), with CI studies matching the broad trends found for all designs combined. There was more coverage for BA and After studies than BACI studies and randomised experiments but gaps remained in several countries, particularly in Africa and to a lesser extent parts of Asia and South America.

4 | DISCUSSION

4.1 | Overall patterns in study design use

Our analyses suggest that although the global evidence base on the effects of conservation actions has gradually shifted from being dominated by After designs towards greater use of controlled designs, and to a lesser extent randomised experiments, most studies still rely on simple designs previously shown to be less robust to bias (Christie, Abecasis, et al., 2020; Christie et al., 2019). For example, predicted proportions of After designs fell from about half of all studies in 1970 to a third by 2023, whilst CI designs and randomised experiments both increased. Despite this shift, randomised experiments and BACI studies remain scarce with predicted proportions in 2023 of only 10% and 5%, respectively. We found that patterns in study design use vary over time, between English and non-English language literatures, across subject areas and to a marginal extent between countries of different national income groups.

Together, these findings show that the reliability of evidence available to support conservation decisions is improving but remains highly variable. In many contexts, the evidence that decision-makers

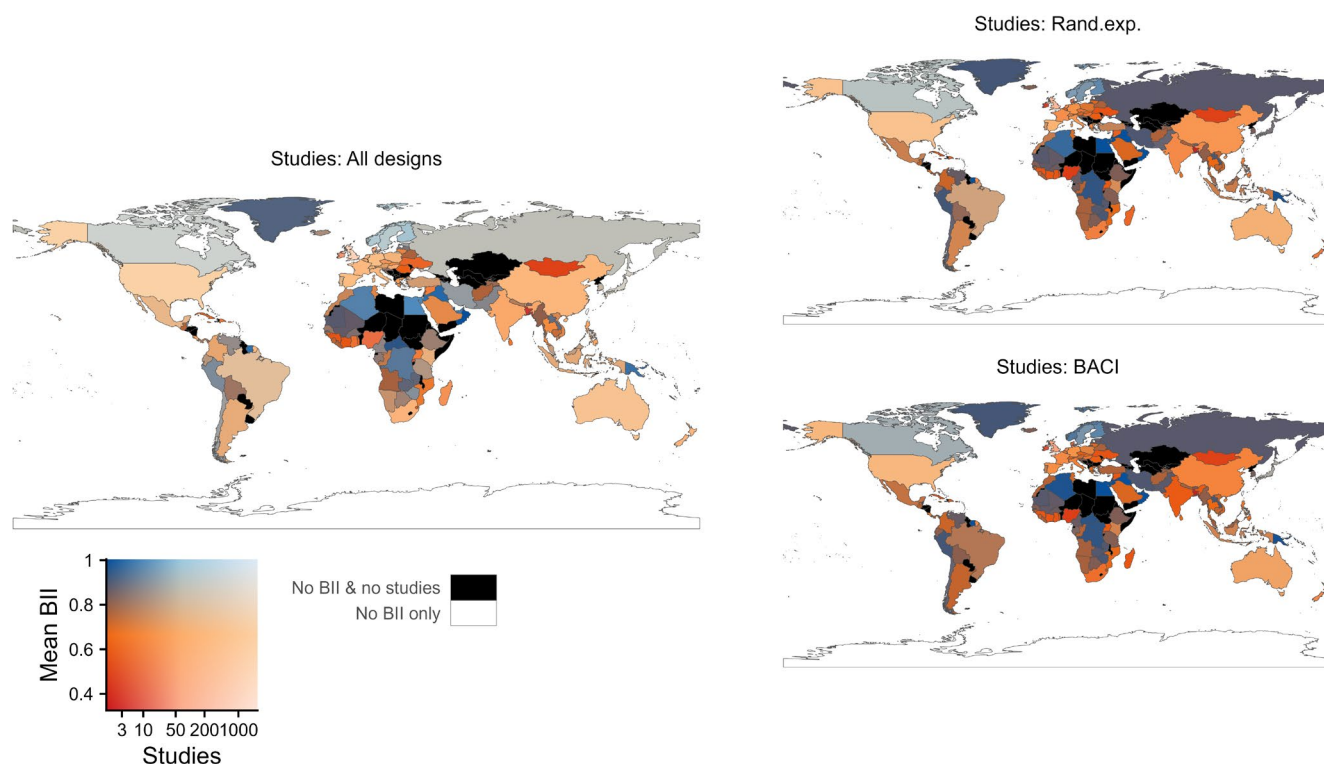


FIGURE 6 Bivariate choropleth map highlighting where there are few or no studies (darker colours or black) of all designs, Before-After-Control-Impact (BACI) and randomised experiments (see Figure S3 for other designs) alongside the mean Biodiversity Intactness Index (BII) of each country. Countries are coloured using a bivariate scale, where darker red colours represents areas with few studies and low mean BII—that is, decreasing numbers of studies ($\log_{10}$ -scaled) from pale (more studies) to darker shades (fewer studies) and decreasing mean BII from blue (high) to orange (moderate) to red (low). Countries with no BII estimates and no studies are shown in black and those with studies but no BII estimates are in white (Antarctica).

can draw on is constrained to study designs that are poorly suited for estimating counterfactual outcomes. In other words, the perceived 'causal revolution' in conservation (Jones & Shreedhar, 2024) is happening but is far from being achieved.

These overall patterns matter because study design type directly limits what can be inferred about action effectiveness. Simple After designs provide, at best, noisy trend descriptions and rely on strong assumptions about unobserved counterfactuals, whereas randomised experiments and to some extent BACI studies can, in principle, estimate the impacts of actions with substantially less bias when implemented well (Christie, Abecasis, et al., 2020; Christie et al., 2019). BA studies can be seen as an improvement over After studies because they include an explicit comparison that may be valid under certain conditions (i.e. if the measured outcome would not change in the absence of the action). However, BA studies are still vulnerable to bias unless they use specific approaches to estimate the counterfactual, such as the Interrupted Time Series (ITS) design (Gottman & Glass, 1978; Wauchope et al., 2021)—although these are most suited to uncommon situations where long-term time series data is available (O'Garra et al., 2025; Wauchope et al., 2021). CI studies represent an improvement over After designs, but they are also vulnerable to bias unless rigorous techniques to estimating counterfactuals are applied, such as matching and pairing to account for confounders (O'Garra et al., 2025; Wauchope et al., 2021). Our results therefore suggest that many published evaluations of actions must be treated with caution when informing high-stakes conservation decisions, even as the relative share of more robust designs grows.

4.2 | Subject-specific and geographical variation

We found that study design use was highly uneven across conservation subjects, suggesting strong context-dependence in the types of designs considered feasible or acceptable. Habitat and soil-focused topics (e.g. Mediterranean Farmland, Shrubland and Heathland, Soil Fertility, Marine Artificial Structures) showed higher predicted probabilities of randomised experiments than the overall mean, whilst many animal-focused topics relied disproportionately on After and BA studies (e.g. on eels, marine fish, captive animals). This pattern is consistent with intuitive constraints on data collection. For example, actions targeting sessile taxa or manipulable substrates more readily support replication, blocking and randomisation (Pynegar et al., 2025). In contrast, actions involving mobile, rare, elusive (e.g. nocturnal or cryptic species) or ethically sensitive species often pose logistical and resource challenges and therefore less robust designs are more commonly used (Hordern et al., 2024).

Since study design use differs by subject area, the shifts we observed in the numbers of studies within each subject area (Figure S2) may help explain why temporal trends are more pronounced for some designs than others (Figure 2)—partly due to how the coverage of the Conservation Evidence database has expanded and been

updated over time. However, these shifts can only partly account for the overall decline in After designs and rise in CI and randomised experiments; both changes in design use within subject areas and changes in subject area representation contributed to the temporal patterns we observed.

Geographical and language patterns suggested that structural inequalities also shape who can implement more rigorous designs and where. Geographically, studies remained heavily concentrated in North America, Western Europe and Australasia, with many bio-diverse and highly impacted countries having few or no evaluations of actions. These findings match previous findings of geographic publication bias in studies testing conservation actions (Anjan et al., 2025; Christie, Amano, et al., 2020; Hordern et al., 2024; Speight et al., 2025) and conservation studies more widely (Watson et al., 2017). In addition, randomised experiments and BACI studies had lower geographical coverage than other designs, being present in 44% and 38%, respectively, of the 160 countries represented in the Conservation Evidence database—whilst After, CI and BA studies were present in 82%, 75% and 60% of studied countries, respectively.

Non-English language studies had substantially higher proportions of After and BA designs than English language studies, and lower proportions of CI and randomised experiments, even after accounting for other variables in the multinomial model. Similarly, lower-income countries showed a slight overrepresentation of After and BA designs, and underrepresentation of randomised experiments and BACI designs, relative to higher-income countries. These findings are consistent with previous work showing that the non-English language and Global South literature often fills taxonomic and geographic gaps, but tends to do so with less robust designs (Amano et al., 2021). These patterns likely reflect broader political, social and economic constraints on evaluating conservation actions, such as limited resources, institutional incentives and publication practices that vary among countries and languages (Amano et al., 2021). It is also likely that many other determinants of design choice (e.g. levels of local institutional support, short-term versus long-term funding and practitioner capacity) remain unmeasured at the national scale. Together, these subject- and geography-specific patterns suggest that universal prescriptions on how to improve study designs in conservation are not appropriate. Instead, our findings point towards the urgent need for targeted, context-sensitive strategies in specific regions, including a greater focus on training and building capacity for implementing more robust study designs in countries that lack studies generally (e.g. especially low-income countries in Africa, Asia and South America).

4.3 | Implications for adaptive management and decision-making

From a joint adaptive management and evidence-based conservation perspective (Gillson et al., 2019), our study's broad trends

suggest that conservation is slowly moving towards using study designs that support iterative learning and building a stronger evidence base. However, our findings also highlight that many actions are still evaluated in ways that make it difficult to learn from past failures and successes to improve conservation actions over time.

Our findings also have direct implications for how impact evaluations are embedded within adaptive management and broader decision processes. In practice, conservation decision-makers typically cannot delay action until an ideal randomised experiment is in place, particularly for rapidly spreading invasive species or critically endangered populations (Pynegar et al., 2025). Conversely, whilst designing robust studies may cause short-term delays, it can also save time by preventing the adoption or prolonged use of ineffective actions based on misleading or unreliable results from poorly designed studies (Alpizar & Ferraro, 2020; Pynegar et al., 2025). Examples include European agri-environment schemes that were once thought to be effective based on simple CI studies with weak causal inference (Batáry et al., 2015), but more robust studies have since shown that these schemes are only effective at reversing farmland bird population declines in some contexts (Sharps et al., 2023).

However, rather than framing robust designs as in tension with urgent action, adaptive management emphasises iteratively updating actions in light of new evidence, which can and should include upgrading study designs over time. Within this adaptive framework, our temporal and spatial analyses can help identify where marginal improvements in design are most feasible and most valuable. First, before any design choice is made, evidence-based conservation principles highlight that the wider evidence base should be consulted first to check whether an evaluation is necessary or useful (Sutherland, 2022). If an evaluation is warranted, then our findings suggest that modest shifts towards CI designs (with appropriate controls, matching or blocking; Schleicher et al., 2020), in subject areas and regions currently dominated by After and BA designs, could substantially improve causal inference at relatively low additional cost. This is especially true if fully randomised experiments remain unrealistic (O'Garra et al., 2025; Wauchope et al., 2021). Even in challenging settings there are often realistic opportunities to incorporate controls, baselines or randomisation and to avoid the use of After designs—alternatively, there are qualitative impact evaluation methods that may be better suited to the situation (Cheek et al., 2023).

Moreover, in subjects and countries where CI designs are already common and logistical constraints are less acute, there may be greater scope to embed randomisation or pre-treatment baselines (i.e. a BACI design) into routine monitoring. Prioritising design improvements should, at least in part, be guided by the consequences of decision errors (Mapstone, 1995): higher minimum design standards and a greater willingness to invest in them are warranted when actions may be highly costly (financially and otherwise), irreversible and/or politically sensitive (e.g. where there is high human or animal welfare concern).

Critically, adaptive management decisions must recognise the trade-off between design robustness and other aspects of

study quality, notably replication and duration—potentially trading greater internal validity for reduced precision and/or generalisability (external validity). More complex designs often demand greater resources per study unit, which can lead to reduced sample sizes or shorter time series; robust designs cannot compensate for extremely low power or insufficient study durations. Although our dataset derived from the Conservation Evidence database did not quantify sample sizes or study duration, this trade-off should be explicitly considered when planning evaluations. For example, in some cases, a well-replicated CI design using carefully selected controls and matching may offer more reliable guidance for adaptive decisions than a severely underpowered or poorly implemented BACI study.

4.4 | Barriers and opportunities for improving study designs

The persistent dominance of simple designs, especially in non-English language studies and lower-income countries, highlights the importance of geographically-variable practical, political and cultural barriers to using more robust study designs. Ethical concerns over withholding actions, difficulties in implementing randomised experiments for socially embedded or livelihood-related actions, short funding cycles and lack of analytical support all constrain what is achievable in practice for many projects.

These patterns point to at least four actionable opportunities. First, funders and organisations implementing actions can explicitly recognise study design as part of project quality (e.g. Smith et al., 2026; Tinsley-Marshall et al., 2022). Providing dedicated resources and longer funding durations to support long-term robust impact evaluation is particularly important for policy or large-scale investments (Baylis et al., 2015), especially in lower-income countries. This includes explicitly triaging when any kind of evaluation is required—that is, checking the evidence base for previous studies rather than requiring projects to evaluate actions as a 'tick-box' exercise.

Second, capacity-building initiatives—especially those targeting non-English-speaking and lower-income countries—could focus on practical, cost-efficient design upgrades (e.g. adding matched controls, obtaining some baseline data or embedding simple randomisation where feasible) rather than only promoting 'gold-standard' randomised experiments.

Third, stronger partnerships between researchers and practitioners can help co-design evaluations that fit within operational constraints whilst still moving 'up' the decision tree of design options—for example, by integrating counterfactual thinking into existing monitoring programmes. Although some schemes exist to encourage the uptake of more rigorous study designs to test conservation actions (e.g. Evidence Champions Programme and Society for Conservation Biology's Impact Evaluation Group seminars), further increasing and deepening collaborations between academics and practitioners would be beneficial.

Fourth, encouraging practitioners to consider where a traditional quantitative evaluation may not be appropriate or feasible and instead looking to qualitative impact evaluation methods (Cheek et al., 2023). This is particularly relevant for where projects cannot disentangle the effects of individual actions or specifically wish to evaluate a combined set of actions together (Sharkey et al., 2025). However, as discussed earlier, greater training and awareness-raising of such methods are clearly needed given that many conservation practitioners, researchers and policymakers may be unfamiliar with their use.

4.5 | Future research

Our study highlights several gaps that future work could address to better understand and improve the design of conservation evaluations globally. Linking study designs to sample sizes, duration and effect size estimates would allow formal tests of hypothesised trade-offs between robustness and replication and would clarify when more complex designs deliver meaningful gains in decision-relevant precision. Qualitative and mixed-methods research—such as interviews and surveys with practitioners, funders and policymakers—could identify the specific institutional and socio-political barriers that drive reliance on simple designs in particular subjects, languages or regions.

Extending design analyses beyond the five categories considered here to incorporate specific counterfactual methods (e.g. different matching approaches; Schleicher et al., 2020) would also provide a more granular picture of how causal inference tools are being used in conservation. Furthermore, investigating taxon-specific trends in study design use using trait-based data could help to understand whether robust designs are more commonly used to study certain species groups or types (e.g. rare or threatened species).

Finally, there is a need for comparative work that evaluates how study designs influence the conclusions of synthesis and decision-support tools, such as systematic reviews, evidence maps and prioritisation exercises, especially when evidence is unevenly distributed across subjects, geographies and languages. Such research would directly inform how evidence from weaker designs should be weighted or supplemented when informing adaptive management and policy.

5 | CONCLUSION

Our analysis of more than 8500 English and non-English language studies in the Conservation Evidence database shows that studies testing conservation actions are gradually shifting away from simple designs (that make more causal assumptions) to more robust designs—but that the latter remain relatively rare and unevenly distributed. These disparities mean that the credibility and usefulness of evidence for guiding conservation decisions varies systematically.

There are particularly acute challenges in the non-English language literature where After and BA designs are more common, and typically low-income countries in Africa, Asia and South America where few studies of any design are present.

Rather than advocating a one-size-fits-all gold standard, our results support a decision-tree approach in which design choices are treated as explicit, context-dependent decisions within adaptive management and broader decision processes. In many cases, meaningful improvements can be made by moving from After to BA or CI designs with better controls, matching or baselines, whilst reserving the greater investment required for BACI or randomised experiments for decisions where the consequences of error are highest. Achieving this will require coordinated efforts by funders, practitioners and researchers to invest in design capacity, reduce structural barriers and integrate rigorous evaluation into routine conservation practice, particularly in lower-income countries that currently bear a disproportionate share of the world's biodiversity and conservation challenges.

AUTHOR CONTRIBUTIONS

Amelia Staples and Alec P. Christie conceived the ideas and designed methodology; Amelia Staples, William H. Morgan, Lynn V. Dicks, Silviu O. Petrovan, Rebecca K. Smith, William J. Sutherland and Alec P. Christie collected the data; Amelia Staples and Alec P. Christie analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

Alec P. Christie was funded by Imperial College London through an Imperial College Research Fellowship. We are grateful to the generous funders of the Conservation Evidence project listed here: <https://www.conservationevidence.com/content/page/24#support-received> and all other collaborators. We thank Julia Jones for providing advice on drafts of the manuscript.

CONFLICT OF INTEREST STATEMENT

Alec P. Christie serves as an Associate Editor for *Journal of Applied Ecology* but had no role in the editorial handling or decision of this manuscript.

DATA AVAILABILITY STATEMENT

Data are available from Zenodo: <https://doi.org/10.5281/zenodo.17192808> (Christie et al., 2026).

ORCID

William H. Morgan  <https://orcid.org/0000-0002-7594-6453>

Lynn V. Dicks  <https://orcid.org/0000-0002-8304-4468>

Silviu O. Petrovan  <https://orcid.org/0000-0002-3984-2403>

Rebecca K. Smith  <https://orcid.org/0000-0003-3294-7592>

William J. Sutherland  <https://orcid.org/0000-0002-6498-0437>

Alec P. Christie  <https://orcid.org/0000-0002-8465-8410>

REFERENCES

- Adams, V. M., Barnes, M., & Pressey, R. L. (2019). Shortfalls in conservation evidence: Moving from ecological effects of interventions to policy evaluation. *One Earth*, 1(1), 62–75. <https://doi.org/10.1016/j.oneear.2019.08.017>
- Alpizar, F., & Ferraro, P. J. (2020). The environmental effects of poverty programs and the poverty effects of environmental programs: The missing RCTs. *World Development*, 127, 104783. <https://doi.org/10.1016/j.worlddev.2019.104783>
- Amano, T., Berdejo-Espinola, V., Christie, A. P., Willott, K., Akasaka, M., Baldi, A., Berthinussen, A., Bertolino, S., Bladon, A. J., Chen, M., Choi, C. Y., Kharrat, M. B. D., De Oliveira, L. G., Farhat, P., Golivets, M., Aranzamendi, N. H., Jantke, K., Kajzer-Bonk, J., Cisel Kemahli Aytekin, M., ... Sutherland, W. J. (2021). Tapping into non-English-language science for the conservation of global biodiversity. *PLoS Biology*, 19(10), e3001296. <https://doi.org/10.1371/JOURNAL.PBIO.3001296>
- Anjan, A., Geldmann, J., Harfoot, M., & Christie, A. P. (2025). The spatial distribution of tests of conservation interventions does not align with global conservation needs. *Biological Conservation*, 309, 111313. <https://doi.org/10.1016/j.biocon.2025.111313>
- Arel-Bundock, V. (2025). *WDI: World development indicators and other World Bank data. R package version 2.7.9*. <https://CRAN.R-project.org/package=WDI>
- Asquith, N. (2020). Large-scale randomized control trials of incentive-based conservation: What have we learned? *World Development*, 127, 104785. <https://doi.org/10.1016/j.worlddev.2019.104785>
- Batáry, P., Dicks, L. V., Kleijn, D., & Sutherland, W. J. (2015). The role of agri-environment schemes in conservation and environmental management. *Conservation Biology*, 29(4), 1006–1016. <https://doi.org/10.1111/COBI.12536>
- Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2015). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9, 58–64.
- Betancourt, M. (2017). *How the shape of a weakly informative prior affects inferences*. <https://mc-stan.org>
- Bladon, E. K., Christie, A. P., Smith, R. K., Sutherland, W. J., & Bladon, A. J. (2025). Butterfly and moth conservation: Results from a global synopsis of evidence. *Journal of Insect Conservation*, 29(2), 1–10. <https://doi.org/10.1007/s10841-024-00646-4/FIGURES/4>
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, 114(30), E6089–E6096. <https://doi.org/10.1073/pnas.1704949114>
- Cheek, J. Z., Eklund, J., Merten, N., Brooks, J., & Miller, D. C. (2023). A guide to qualitative attribution methods for evaluation in conservation. *Conservation Biology*, 37, e14071. <https://doi.org/10.1111/cobi.14071>
- Christie, A., Staples, A., Morgan, W., Dicks, L., Petrovan, S., Smith, R., & Sutherland, W. (2026). *alecchristie888/study_design_time: Code and data for revised manuscript (v0.1.1)*. Zenodo. <https://doi.org/10.5281/zenodo.17192808>
- Christie, A. P., Abecasis, D., Adjeroud, M., Alonso, J. C., Amano, T., Anton, A., Baldigo, B. P., Barrientos, R., Bicknell, J. E., Buhl, D. A., Cebrian, J., Ceia, R. S., Cibils-Martina, L., Clarke, S., Claudet, J., Craig, M. D., Davoult, D., De Backer, A., Donovan, M. K., ... Sutherland, W. J. (2020). Quantifying and addressing the prevalence and bias of study designs in the environmental and social sciences. *Nature Communications*, 11(1), 6377. <https://doi.org/10.1038/s41467-020-20142-y>
- Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. R., & Sutherland, W. J. (2020). Poor availability of context-specific evidence hampers decision-making in conservation. *Biological Conservation*, 248, 108666. <https://doi.org/10.1016/j.biocon.2020.108666>
- Christie, A. P., Amano, T., Martin, P. A., Petrovan, S. O., Shackelford, G. E., Simmons, B. I., Smith, R. K., Williams, D. R., Wordley, C. F. R., & Sutherland, W. J. (2021). The challenge of biased evidence in conservation. *Conservation Biology*, 35(1), 249–262. <https://doi.org/10.1111/COBI.13577>
- Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., & Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, 56(12), 2742–2754. <https://doi.org/10.1111/1365-2664.13499>
- Cooke, S. J., Cook, C. N., Nguyen, V. M., Walsh, J. C., Young, N., Cvitanovic, C., Grainger, M. J., Randall, N. P., Muir, M., Kadykalo, A. N., Monk, K. A., & Pullin, A. S. (2023). Environmental evidence in action: On the science and practice of evidence synthesis and evidence-based decision-making. *Environmental Evidence*, 12(1), 10. <https://doi.org/10.1186/s13750-023-00302-5>
- Curzon, H. F., & Kontoleon, A. (2016). From ignorance to evidence? The use of programme evaluation in conservation: Evidence from a Delphi survey of conservation experts. *Journal of Environmental Management*, 180, 466–475.
- De Palma, A., Sanchez-Ortiz, K., Martin, P. A., Chadwick, A., Gilbert, G., Bates, A. E., Börger, L., Contu, S., Hill, S. L. L., & Purvis, A. (2018). Challenges with inferring how land-use affects terrestrial biodiversity: Study design, time, space and synthesis. *Advances in Ecological Research*, 58, 163–199. <https://doi.org/10.1016/BS.AEER.2017.12.004>
- Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, 2009(122), 75–84. <https://doi.org/10.1002/EV.297>
- Ferraro, P. J., Cherry, T. L., Shogren, J. F., Vossler, C. A., Cason, T. N., Flint, H. B., Hochard, J. P., Johansson-Stenman, O., Martinsson, P., Murphy, J. J., Newbold, S. C., Thunström, L., van Soest, D., van't Veld, K., Dannenberg, A., Loewenstein, G. F., & van Boven, L. (2023). Create a culture of experiments in environmental programs. *Science*, 381(6659), 735–737. <https://doi.org/10.1126/science.adf7774>
- Ferraro, P. J., & Pattanayak, S. K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology*, 4, e105.
- Gabry, J., Češnovar, R., Johnson, A., & Bröder, S. (2025). *cmdstanr: R interface to CmdStan* (R package version 0.9.0).
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4), 1360–1383. <https://doi.org/10.1214/08-AOAS191>
- Gillson, L., Biggs, H., Smit, I. P. J., Virah-Sawmy, M., & Rogers, K. (2019). Finding common ground between adaptive management and evidence-based approaches to biodiversity conservation. *Trends in Ecology & Evolution*, 34(1), 31–44. <https://doi.org/10.1016/j.tree.2018.10.003>
- Gottman, J. M., & Glass, G. V. (1978). Analysis of interrupted time-series experiments. In *Single subject research* (pp. 197–235). Elsevier. <https://doi.org/10.1016/B978-0-12-425850-1.50011-9>
- Hordern, E., White, T. B., Berthinussen, A., Smith, R. K., Sutherland, W. J., & Christie, A. P. (2024). Prioritizing future evidence needs for marine and freshwater mammal conservation action. *Animal Conservation*, 28, 542–552. <https://doi.org/10.1111/acv.13003>
- Jones, J. P. G., & Shreedhar, G. (2024). The causal revolution in biodiversity conservation. *Nature Human Behaviour*, 8(7), 1236–1239. <https://doi.org/10.1038/s41562-024-01897-6>

- Lemieux, C. J., Halpenny, E. A., Swerdfager, T., He, M., Joyce Gould, A., Den Hoed, D. C., Bueddefeld, J., Hvenegaard, G. T., Joubert, B., & Rollins, R. (2021). Free Fallin'? The decline in evidence-based decision-making by Canada's protected areas managers. *Facets*, 6(1), 640–664. https://doi.org/10.1139/FACETS-2020-0085/SUPPL_FILE/FACETS-2020-0085_SUPPLEMENT1.DOCX
- Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7), 912–928. <https://doi.org/10.1111/OIK.05985>
- Mapstone, B. D. (1995). Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors. *Ecological Applications*, 5(2), 401–410. <https://doi.org/10.2307/1942031>
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research* (pp. 1–319). Cambridge University Press. <https://doi.org/10.1017/CBO9780511804564>
- Newbold, T., Hudson, L. N., Arnell, A. P., Contu, S., de Palma, A., Ferrier, S., Hill, S. L. L., Hoskins, A. J., Lysenko, I., Phillips, H. R. P., Burton, V. J., Chng, C. W. T., Emerson, S., Gao, D., Pask-Hale, G., Hutton, J., Jung, M., Sanchez-Ortiz, K., Simmons, B. I., ... Purvis, A. (2016). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, 353, 288–291. <https://doi.org/10.1126/science.aaf2201>
- Ockendon, N., Amano, T., Cadotte, M., Downey, H., Hancock, M. H., Thornton, A., Tinsley-Marshall, P., & Sutherland, W. J. (2021). Effectively integrating experiments into conservation practice. *Ecological Solutions and Evidence*, 2(2), e12069. <https://doi.org/10.1002/2688-8319.12069>
- O'Garra, T., Martin, R., Pynegar, E., Polo-Urrea, C., & Eklund, J. (2025). Selecting among counterfactual methods to evaluate conservation interventions. *Conservation Science and Practice*, 7, e70066. <https://doi.org/10.1111/CSP2.70066>
- Phillips, H. R. P., De Palma, A., Gonzalez, R. E., Contu, S., Hill, S. L. L., Baselga, A., Archer, F., Ball, J. A., Brookes, C. L. L., Chalk, K., Gough, R. L., Harwood, V., & Purvis, A. P. (2021). *The Biodiversity Intactness Index - Country, region and global-level summaries for the year 1970 to 2050 under various scenarios* [Data set]. Natural History Museum. <https://doi.org/10.5519/he1eqmg1>
- Pullin, A. S., & Knight, T. M. (2003). Support for decision making in conservation practice: An evidence-based approach. *Journal for Nature Conservation*, 11(2), 83–90. <https://doi.org/10.1078/1617-1381-00040>
- Pynegar, E., Booth, H., Doulton, H., Ferraro, P. J., Mohamed, M., Rakotonarivo, O. S., & Jones, J. P. G. (2025). RCTs in the wild: Designing and implementing conservation programs as randomized control trials. *Conservation Science and Practice*, 7, e70029. <https://doi.org/10.1111/CSP2.70029>
- Pynegar, E., Gibbons, J., Asquith, N., & Jones, J. (2021). What role should randomized control trials play in providing the evidence base for conservation? *Oryx*, 55(2), 235–244. <https://doi.org/10.1017/S0030605319000188>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, 2(3), 808–840. <https://doi.org/10.1214/08-aos187>
- Schleicher, J., Eklund, J. D., Barnes, M., Geldmann, J., Oldekop, J. A., & Jones, J. P. G. (2020). Statistical matching for conservation science. *Conservation Biology: The Journal of the Society for Conservation Biology*, 34(3), 538–549. <https://doi.org/10.1111/COBI.13448>
- Sharkey, W., Jeha, L., Lo, M. G. Y., Simkins, A. T., Llopis, J. C., Felipe Cadiello, M. M., Clarke, C., Brown, M. R. C., Sharma, P., Serem, E. K., Adekola, O. E., Ariando, W., & Verissimo, D. (2025). The value of qualitative approaches to impact evaluation in biodiversity conservation. *Conservation Science and Practice*, 7(11), e70102. <https://doi.org/10.1111/csp.270102>
- Sharps, E., Hawkes, R. W., Bladon, A. J., Buckingham, D. L., Border, J., Morris, A. J., Grice, P. V., & Peach, W. J. (2023). Reversing declines in farmland birds: How much agri-environment provision is needed at farm and landscape scales? *Journal of Applied Ecology*, 60(4), 568–580. <https://doi.org/10.1111/1365-2664.14338>
- Smith, R. K., Ockendon, N., Al-Fulaij, N., Badalotti, A., Beattie, R., Byrne, J., Cooke, S., de Zoeten, T., Frick, W. F., Funk, S. M., Gibbs, D., Gupta, L., Hendrix, T., Holmes, B., Johnstone, N., Kinghorn, J., Mickleburgh, S., Miller, F., Muir, M., ... Sutherland, W. J. (2026). Role of funders in embedding tests in conservation practice. *Conservation Biology*, e70309. <https://doi.org/10.1111/cobi.70309>
- Speight, O., Morgan, W. H., White, T. B., Sainsbury, K. A., Bouskila, A., Rotem, G., Smith, R. K., Sutherland, W. J., Watson, M. J., & Christie, A. P. (2025). Exploring gaps, biases, and research priorities in the evidence for reptile conservation actions. *Conservation Biology*, 39, e70073. <https://doi.org/10.1111/cobi.70073>
- Sutherland, W. J. (2022). *Transforming conservation: A practical guide to evidence and decision making* (pp. 1–413). Open Book Publishers. <https://doi.org/10.11647/OBP.0321>
- Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19(6), 305–308. <https://doi.org/10.1016/J.TREE.2004.03.018>
- Sutherland, W. J., Taylor, N. G., MacFarlane, D., Amano, T., Christie, A. P., Dicks, L. V., Lemasson, A. J., Littlewood, N. A., Martin, P. A., Ockendon, N., Petrovan, S. O., Robertson, R. J., Rocha, R., Shackelford, G. E., Smith, R. K., Tyler, E. H. M., & Wordley, C. F. R. (2019). Building a tool to overcome barriers in research-implementation spaces: The conservation evidence database. *Biological Conservation*, 238, 108199. <https://doi.org/10.1016/J.BIOCON.2019.108199>
- Thiault, L., Kernaléguen, L., Osenberg, C. W., & Claudet, J. (2017). Progressive-change BACIPS: A flexible approach for environmental impact assessment. *Methods in Ecology and Evolution*, 8(3), 288–296. <https://doi.org/10.1111/2041-210X.12655>
- Thornton, A., Morgan, W. H., Bladon, E. K., Smith, R. K., & Sutherland, W. J. (2025). Coral conservation: Global evidence for the effects of actions. In *Synopses of conservation evidence series*. University of Cambridge, Open Book Publishers. <https://doi.org/10.11647/OBP.0453>
- Tinsley-Marshall, P., Downey, H., Adum, G., Al-Fulaij, N., Bourn, N. A. D., Brotherton, P. N. M., Frick, W. F., Hancock, M. H., Hellon, J., Hudson, M. A., Kortland, K., Mastro, K., McNicol, C. M., McPherson, T., Mickleburgh, S., Moss, J. F., Nichols, C. P., O'Brien, D., Ockendon, N., ... Sutherland, W. J. (2022). Funding and delivering the routine testing of management interventions to improve conservation effectiveness. *Journal for Nature Conservation*, 67, 126184. <https://doi.org/10.1016/j.jnc.2022.126184>
- Underwood, A. J. (1994). On beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecological Applications*, 4(1), 3–15. <https://doi.org/10.2307/1942110>
- Watson, J. E. M., di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., Maina, J. M., Possingham, H. P., Rogalla von Bieberstein, K., & Venter, O. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, 10, 32–42. <https://doi.org/10.1016/J.GECCO.2017.01.008>

Wauchope, H. S., Amano, T., Geldmann, J., Johnston, A., Simmons, B. I., Sutherland, W. J., & Jones, J. P. G. (2021). Evaluating impact using time-series data. *Trends in Ecology & Evolution*, 36(3), 196–205. <https://doi.org/10.1016/j.tree.2020.11.001>

SUPPORTING INFORMATION

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

Table S1. A cross-walk table showing how the detection of key terms in study paragraphs was used to define the study design used in each English language study.

Table S2. Posterior summaries for parameters from the Bayesian multinomial logistic regression of study design (After, BA, CI, BACI and randomised experiment) on publication year (scaled), log GDP per capita (scaled), Biodiversity Intactness Index (BII; scaled), article language (English or non-English) and subject area (synopsis), with a random intercept for study ID.

Figure S1. Cumulative raw proportion of studies using each study design in the Conservation Evidence database published in each year.

Figure S2. The raw proportion of studies over time (in 4-year intervals—note the last interval is: 2018–2023) for all studies (all designs) in each subject area (synopsis).

Figure S3. Bivariate choropleth map highlighting where there are few or no studies (darker colours or black) of CI, BA and After designs (see Figure 6 for other designs) alongside the mean Biodiversity Intactness Index (BII) of each country.

How to cite this article: Staples, A., Morgan, W. H., Dicks, L. V., Petrovan, S. O., Smith, R. K., Sutherland, W. J., & Christie, A. P. (2026). Temporal and spatial trends in study designs used to understand what works in conservation. *Journal of Applied Ecology*, 63, e70478. <https://doi.org/10.1111/1365-2656.70478>