Evidence Synthesis in Conservation Science: Summer Research Report

Thalia Eleni Seale

September 2021

Introduction

With drastic loss of biodiversity, there is an increasing need for scientifically informed conservation interventions. Evidence synthesis, combining multiple sources of information to communicate the state of knowledge on a certain issue is integral to this effort. However, poor quality study designs and other sources of bias pervade the conservation evidence base, presenting a major methodological problem for providing evidence-based recommendations [1] [3] .

This report builds on previous work by quantifying the bias of different experimental design features by comparing the results of studies with different designs concerning the same question. [Summary of results.]

Bias and study design

In this section, we cover the key concepts relating to bias and study design.

What is bias?

Bias of an estimator is the difference between the estimator's expected value and the true value of the parameter.

In the context of evidence synthesis to answer a specific question, there are two sources of bias that may be found in a study:

- **Internal bias**, reflects problems within the study that could cause the results to differ from the true value of the parameter that was intended to be measured. A study that has low internal bias is **rigorous**.
- **External bias**, reflects well the study answers the question target question. A study with low external bias is **relevant**. [11]

Some examples of internal bias include [11] :

- **Selection bias**: Systematic differences between comparison group and baseline
- **Performance bias**: Systematic differences other than the study comparison, such as inadequate blinding.
- **Attrition bias**: Systematic differences between comparison groups in exclusions and drop-outs.
- **Detection bias**: Systematic differences between groups in outcome assessment.

Some examples of external bias include[11] :

- **Population bias**: Differences in age, sex or health status etc of idealised study participants vs target population.
- **Intervention bias**: Differences in treatment vs target method.

- **Control bias**: Differences in control strategy.
- **Outcome bias**: Differences in method of measurement vs target question.

The relationship between bias, studies and the target question being considered is shown in Fig. 1.

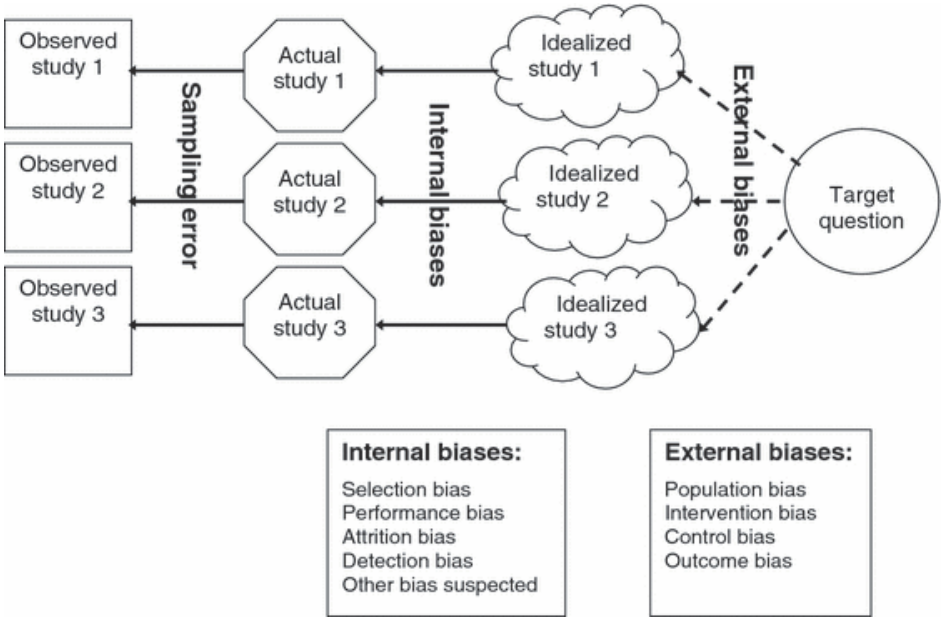


Figure 1: Diagram demonstrating the relationship between internal, external biases . [11]

Evidence synthesis may be affected further by:

- **Publication bias**: Bias caused by the phenomenon that papers are more likely to be published if show significant results.
- **Language bias:** Literature may be overlooked if due to its language of publication. [4]
- **P-hacking**: Manipulation of data in order to produce a significant result.
- **HARKing**: Hypothesising after results are known.[2]

Study Designs

The focus of the project is on the types of bias arising from study design. Fig. 2 shows the hierarchical nature of common study designs used in conservation science.

There are many potential biases that may arise from the choice of study design. Randomisation removes pre-impact differences in a stochastic sense, thus non-random samples may be biased through the influence of confounding covariates. Lack of control introduces the potential for large design bias since it is difficult to separate changes occurring as a result of the treatment vs regardless of the treatment. [2]

Prevalence of bias within the conservation, social and medical knowledge bases.

Now that we can recognises forms bias, we will discuss the high prevalence of bias within the conservation, social and medical knowledge bases, and therefore the motivation for bias mitigation in evidence synthesis.

There is substantial bias in the distribution of conservation intervention studies towards Europe, America and Australasia- large risk of extrinsic bias for evidence synthesis regarding endangered habitats. [4]

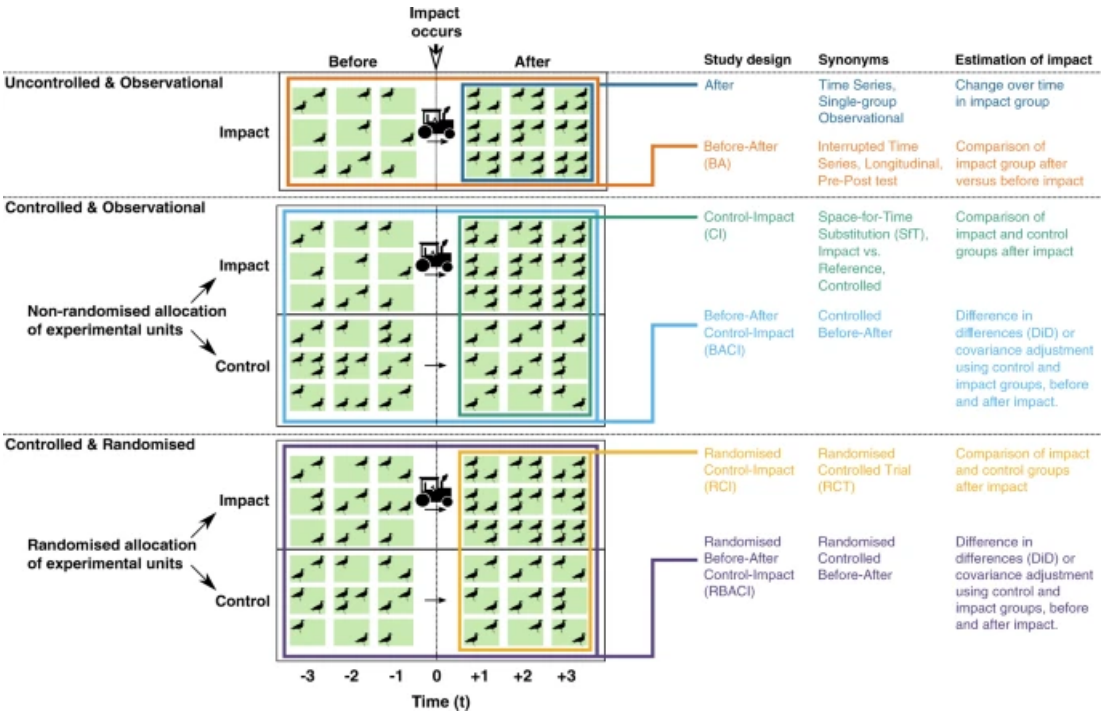


Figure 2: "A hypothetical study set-up is shown where the abundance of birds in three impact and control replicates (e.g., fields represented by blocks in a row) are monitored before and after an impact (e.g., ploughing)..." [2]

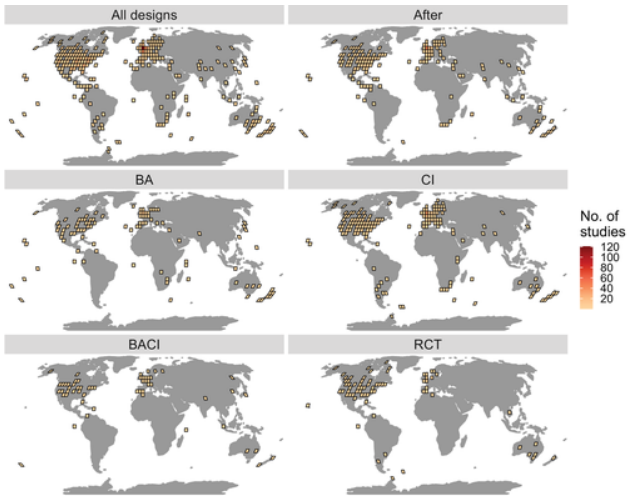


Figure 3: Spatial distribution of studies on bird conservation. [4]

Approximately, 90% of amphibian studies and 84% of bird studies were conducted in North America, Europe or Australasia. Geographical distribution varies considerably by study design. Most reliable study designs, BACI and RCT, concentrated in North America and Europe [4]. Fig. 3 depicts the global distribution of different study designs. Examining summaries of several thousand biodiversity conservation intervention studies in the Conservation evidence data base, R-BACI, R-CI and BACI designs (the most robust designs) were found to make up 23% of intervention studies [2].

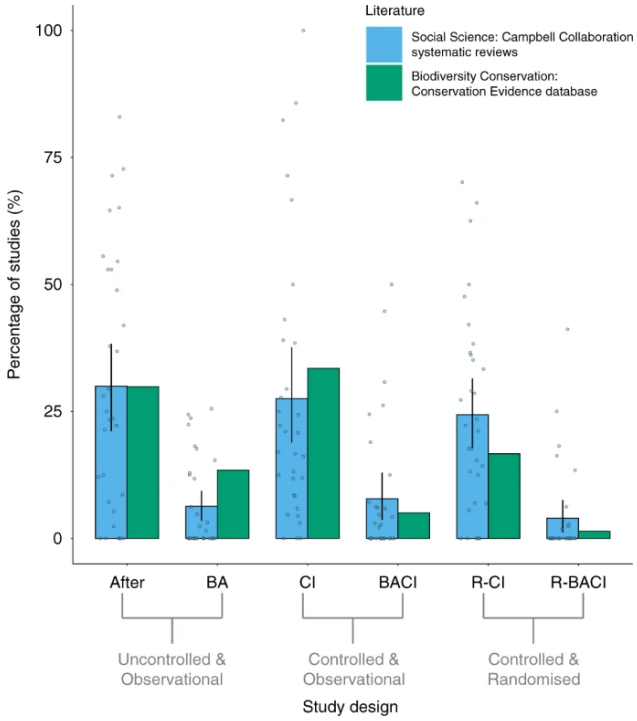


Figure 4: Percentage prevalence of study designs in conservation and social science literature databases.

In both the environmental and social sciences, logistical and ethical considerations mean that randomised study designs are often not possible. R-BACI, R-CI and BACI designs comprised 36% of intervention studies for social science determined by examining systematic reviews by the Campbell Collaboration.

In the preprint publication "Face masks and coverings for the general public: Behavioural knowledge, effectiveness of cloth coverings and public messaging", Mills et al. identify the importance of factors on public adherence to wearing face masks and coverings such as public understanding of virus transmission and risk perception and consistent, effective public messaging. Discarding lower quality study designs prevented effective communication of results. The issue of relevance also reappears regarding the transferability of results between settings; in this case, there was doubt as to how well the results in healthcare settings could be applied to the general population[8].

Methods for Meta-analyses: Random- and Mixed-effects models

Meta-analysis is a method for combing data from several different studies on the same outcome of interest to produce a single estimate. It is one of the most common forms of evidence synthesis.

Types of meta-analysis:

- **Aggregate data (AD)**: Summary data for the same outcome from each study is pooled for statistical analysis,
- **Individual participant data (IPD)**: Pooling of raw data for each participant from each included study.

IPD is ideal but it is harder and more costly to obtain raw data so AD is more common.

Suppose we have:

1. Estimate of outcome of interest γ_i,
2. The standard error of the estimate σ_i,

for every study i of n studies.

There are a few different models for meta-analyses: **Fixed effect model**

- – Assumes all observed variation caused by within-study sampling error.
  – All studies assumed to measure the same overall effect.
  – If θ is the true effect size then the model is given by

$$\gamma_i = \theta + \epsilon_i, \tag{1}$$

  where ϵ_i is the sampling error for γ_i.

Note that if a study is suffering from internal bias, this would contribute to an increase in μ_i in the random-effects model, and thus if many studies with internal bias are included in a meta-analysis, this can contribute to large τ^2, between study heterogeneity.

Quantifying Bias

In this section, we look at existing work into quantifying the bias due to study design.

Using data from R-BACI and BACI studies, Christie et al. empirically estimate bias due to so study design, due to the hierarchy of the study designs. [2]

Since (R-)BACI contains "all the possible data you could have" from the situation being studied, it is possible to ignore some of the data and analyse it as if only a (R-)CI or (R-)BA study was performed for example.

The model proposed is:

- $\hat{\beta}_{ij}$ is the true effect estimator in study i using study design j,
- β_i is the true effect for the response i,
- $\gamma_i j$ is the bias of design j in response i,
- ϵ_{ij} is the sampling noise of the statistical estimator,
- Ω is the correlation matrix for the different estimators in the same response,
- λ is a scaling factor accounting for possible over/under-estimation of SEs,

$$\hat{\beta}_{ij} = \beta_i + \gamma_{ij} + \epsilon_{ij}, \tag{2}$$

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2), \gamma_{ij} \sim \mathcal{N}(0, \sigma_j^2), \epsilon_i \sim \mathcal{N}(0, \Lambda). \tag{3}$$

λ for the non-randomised results appears to be significantly above 1, indicating systematic underestimation of statistical error, so there is some unaccounted for variation still present.

- Estimates of (R-)BACI and (R-)CI were similar while point estimates for most other designs often differed substantially in their magnitude and sign.
- Around 30% of responses in both non-randomised and randomised datasets, study design estimates differed in their statistical significance.
- Rare for 95% confidence intervals of different designs' estimates to not overlap except comparing estimates of BA to (R-)BACI and (R-)CI.

Table 3. Advantages, potential limitations, and applicability of the available methods for bias adjustment in meta-analysis

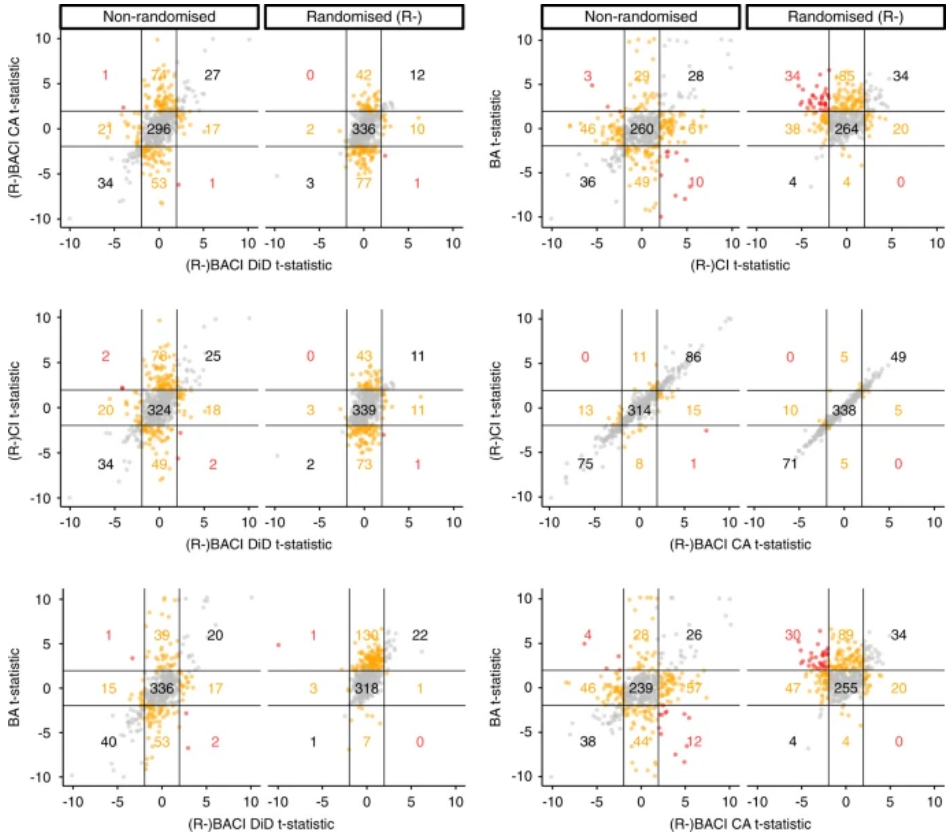
Approach	Advantages	Potential limitations
Stratification	Very easy to implement in practice.	May introduce collider stratification bias. Choice of scale may influence quality threshold.
Quality effects weights	Does not require information on the direction and magnitude of bias. Can be used even with a limited number of studies. Maintains nominal coverage. Tolerates a fair amount of variability in reproducibility related to who assesses the quality of the studies. This will not affect model-based error reduction if the study ranking retains some value in terms of risk of bias.	May not bias-adjust if studies do not vary much in quality because the adjustment is relative. Requires a comprehensive tool for study appraisal. Equal weighting in current scales diminishes its effectiveness for bias adjustment.
Meta-regression	Can model excess heterogeneity in effect size distributions. Can reveal and adjust for confounding among characteristics. Simple method to apply with available software.	Risk of misapplication and misinterpretation of models. May induce selection (collider) bias (when used in lieu of stratification). Requires enough studies to be available. Does not have a mechanism for reduction in statistical certainty caused by poor quality.
Elicitation by experts	Simple methodology.	Magnitude and direction of bias cannot be empirically confirmed. Results are usually not reproducible. Highly subjective and may be prone to psychological biases.
Prior distributions	Based on empirical evidence.	Study specific bias is context-dependent and not always exchangeable between studies. Requires empirical data on the impact of different bias domains to be available. Large effort to collate information unless researchers have already done this and the data sets are available for public use.

Figure 5: Table summarising advantages and disadvantages of some proposed methods for bias mitigation. [10]

Table 2 Results of hierarchical Bayesian model for randomised and non-randomised datasets.

Term	Posterior mean	95% Credible Interval
Randomised (R-)		
σ_β	0.746	[0.679, 0.813]
λ	1.119	[0.980, 1.276]
σ[BACI DiD]	0.029	[0.005, 0.097]
σ[BACI CA]	0.005	[0.002, 0.008]
σ[CI]	0.005	[0.002, 0.008]
σ[BA]	0.773	[0.699, 0.846]
Ω[BACI DiD, BACI CA]	0.268	[0.152, 0.379]
Ω[BACI DiD, CI]	0.239	[0.122, 0.354]
Ω[BACI DiD, BA]	0.849	[0.770, 0.914]
Ω[BACI CA, CI]	0.995	[0.994, 0.996]
Ω[BACI CA, BA]	−0.168	[−0.332, 0.002]
Ω[CI, BA]	−0.184	[−0.349, −0.015]
Non-randomised		
σ_β	0.700	[0.628, 0.776]
λ	1.822	[1.595, 2.098]
σ[BACI DiD]	0.017	[0.004, 0.049]
σ[BACI CA]	0.049	[0.005, 0.128]
σ[CI]	0.091	[0.008, 0.137]
σ[BA]	0.645	[0.573, 0.720]
Ω[BACI DiD, BACI CA]	0.140	[0.010, 0.263]
Ω[BACI DiD, CI]	0.036	[−0.106, 0.176]
Ω[BACI DiD, BA]	0.798	[0.718, 0.865]
Ω[BACI CA, CI]	0.939	[0.923, 0.954]
Ω[BACI CA, BA]	−0.127	[−0.285, 0.026]
Ω[CI, BA]	−0.229	[−0.397, −0.061]

In randomised datasets, BACI and CI terms refer to R-BACI and R-CI designs (denoted by 'R-'). The σ terms are the standard deviations of the bias of each design, so larger σ values correspond to more biased designs. σ_β refers to the standard deviation of the true effect across all datasets. Ω represents the within-response correlations between study design estimates, and λ models systematic underestimation ($\lambda > 1$) or overestimation ($\lambda < 1$) of the statistical error using GLM (M)s. See methods for more details on the model.
BA before-after, BACI before-after-control-impact, CI control-impact.



Field	Description
`publication_id`	Link to publication summary on the metadataset website
`intervention`	Standardised intervention publication has tested
`population`	Broad group or system studied
`outcome`	Outcome measured by publication
`Design`	Terms denoting study design used
`response_ratio`	Response ratio calculated based on the study design
`log_response_ratio`	log(`response_ratio`)
`selected_v`	Variance for the effect size

Table 1: A summary of the data fields and there significance.

Overview of Project Methods

In the previous section, we discussed methods for quantifying bias and one of the methods proposed suffered from the issue that the population correlation matrix between different kinds of estimates for the same data was hard to deduce. This may be worked around by comparing independent studies with different designs measuring the same underlying value to deduce the relationship between study features and bias.

The data

The data consists of 258 publication results, which may be grouped based on attributes such as the population studied, treatment used and outcome measured.

Main features of the data

The Table 1 details the main fields of information.

We consider results from different publications to be comparable if the intervention, population and outcome are the same.

Preparing the data

Code for preparing the data may be found via the following link: Data preparation and Optimisation. Main package used to handle data frames was Tidyverse [12].

Initial observations about the data

Code for observations of the data may be found via the following link: Data preparation and Optimisation

The model

Let $\hat{\beta_{ij}}$ be the log response ratio estimate for study j in group i, where

$$i = 1, \ldots, m, j = 1, \ldots, n_i, \tag{4}$$

where $m \in \mathbb{N}$ is the number of comparison groups and $n_i \in \mathbb{N}$ is the number of publications in each comparison group.

In our model, the log response ratio estimate $\hat{\beta_{ij}} \in \mathbb{R}$ is modelled by

$$\hat{\beta_{ij}} = \beta_i + \gamma_{ij} + \epsilon_{ij}, \tag{5}$$

where $\beta_i \in \mathbb{R}$ is the log response ratio for comparison i, $\gamma_{ij} \in \mathbb{R}$ is design bias in study j in comparison i and $\epsilon_{ij} \in \mathbb{R}$ is the random error in study j in comparison i. These components are in turn modelled by:

$$\beta_j \sim \mathcal{N}(0, \sigma_\beta^2), \tag{6}$$
$$\gamma_{ij} \sim \mathcal{N}(0, \sigma^2(X_{ij})), \tag{7}$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2). \tag{8}$$

where $X_{ij} \in \mathbb{R}^d$ is a vector of design features of study j in group i, $\sigma_{ij} \in \mathbb{R}^+$ is the standard error, and

$$\log \sigma^2(X_{ij}) = X_{ij}^T \theta, \theta \in \mathbb{R}^d. \tag{9}$$

We also assume weakly informative prior distributions:

$$\sigma_\beta \sim \text{InvGamma}(1, 0.02), \tag{10}$$
$$\theta_i \sim \mathcal{N}(0, 1.52). \tag{11}$$

For now we are interested in the following design features:

- Randomisation,
- Controls,
- Before-after,
- Presence of all three previous features, i.e. BACI design.

So we set:

$$X_{ij} = \begin{pmatrix} 1 \\ x_{\text{randomised}} \\ x_{\text{controlled}} \\ x_{\text{before-after}} \\ x_{\text{BACI}} \end{pmatrix}, \tag{12}$$

where the first element of the vector is the intercept, and $x_{\text{design feature}}$ is a dummy coding variable for the features we are considering. X_{ij} may be determined from the `Design` field of the data.

Study Design Optimisation

As mentioned in the previous section, in publications may appear in at most one comparison. Thus, to be able to use as much of the data that we have available as possible. To do this, we used linear and quadratic optimisation tools to find a good combination of studies and comparisons.

The different optimisation methods attempted can be found at Data preparation and Optimisation. We mainly use the Gurobi package to perform the optimisation [7].

The best trade off between complexity and quality of results appeared to be between "345" method and the quadratic method. The comparison is explored in Optimisation comparison. Ultimately, we settled on the "345" method.

Implementing the model in Stan

Code for the implementation of the model in Stan can be found at Stan model [9]. We check that the code is functioning correctly using a simulated data set, using the Nimble package to simulate inverse gamma variables [5].

Results

A discussion of the implementation of the data and the results may be found here: Stan model. To check that the MCMC chains successfully converged, we utilise the Bayesplot package [6].

References

[1] Alec P Christie et al. "Poor availability of context-specific evidence hampers decision-making in conservation". In: *Biological Conservation* 248 (2020), p. 108666.

[2] Alec P Christie et al. "Quantifying and addressing the prevalence and bias of study designs in the environmental and social sciences". In: *Nature communications* 11.1 (2020), pp. 1–11.

[3] Alec P Christie et al. "The challenge of biased evidence in conservation". In: *Conservation Biology* 35.1 (2021), pp. 249–262.

[4] Alec P Christie et al. "The challenge of biased evidence in conservation". In: *Conservation Biology* 35.1 (2021), pp. 249–262.

[5] P. de Valpine et al. "Programming with models: writing statistical algorithms for general model structures with NIMBLE". In: *Journal of Computational and Graphical Statistics* 26 (2017), pp. 403–417. DOI: 10.1080/10618600.2016.1172487.

[6] Jonah Gabry et al. "Visualization in Bayesian workflow". In: *J. R. Stat. Soc. A* 182 (2 2019), pp. 389–402. DOI: 10.1111/rssa.12378.

[7] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2021. URL: https://www.gurobi.com.

[8] M Mills, C Rahal, and E Akimova. "Face masks and coverings for the general public: Behavioural knowledge, effectiveness of cloth coverings and public messaging". In: *The Royal Society* 26 (2020).

[9] Stan Development Team. *RStan: the R interface to Stan*. 2020. URL: http://mc-stan.org/.

[10] Jennifer C Stone et al. "Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches". In: *Journal of clinical epidemiology* 117 (2020), pp. 36–45.

[11] Rebecca M. Turner et al. "Bias modelling in evidence synthesis". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172.1 (2009), pp. 21–47. DOI: https://doi.org/10.1111/j.1467-985X.2008.00547.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-985X.2008.00547.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2008.00547.x.

[12] Hadley Wickham et al. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.