Received: 2 March 2020 | Accepted: 15 June 2020

DOI: 10.1111/2041-210X.13445

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Comparing traditional and Bayesian approaches to ecological meta-analysis

Paula Pappalardo[1] | Kiona Ogle[2,3] | Elizabeth A. Hamman[1] | James R. Bence[4] | Bruce A. Hungate[3] | Craig W. Osenberg[1]

[1]Odum School of Ecology, University of Georgia, Athens, GA, USA

[2]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

[3]Center for Ecosystem Science and Society and Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA

[4]Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

Correspondence
Paula Pappalardo
Email: paulapappalardo@gmail.com

Funding information
Biological and Environmental Research, Grant/Award Number: DE-SC-0010632; National Science Foundation, Grant/Award Number: DEB-1655426 and DEB-1655394

Handling Editor: Robert B. O'Hara

Abstract

1. Despite the wide application of meta-analysis in ecology, some of the traditional methods used for meta-analysis may not perform well given the type of data characteristic of ecological meta-analyses.
2. We reviewed published meta-analyses on the ecological impacts of global climate change, evaluating the number of replicates used in the primary studies (n_i) and the number of studies or records (k) that were aggregated to calculate a mean effect size. We used the results of the review in a simulation experiment to assess the performance of conventional frequentist and Bayesian meta-analysis methods for estimating a mean effect size and its uncertainty interval.
3. Our literature review showed that n_i and k were highly variable, distributions were right-skewed and were generally small (median $n_i = 5$, median $k = 44$). Our simulations show that the choice of method for calculating uncertainty intervals was critical for obtaining appropriate coverage (close to the nominal value of 0.95). When k was low (<40), 95% coverage was achieved by a confidence interval (CI) based on the t distribution that uses an adjusted standard error (the Hartung–Knapp–Sidik–Jonkman, HKSJ), or by a Bayesian credible interval, whereas bootstrap or z distribution CIs had lower coverage. Despite the importance of the method to calculate the uncertainty interval, 39% of the meta-analyses reviewed did not report the method used, and of the 61% that did, 94% used a potentially problematic method, which may be a consequence of software defaults.
4. In general, for a simple random-effects meta-analysis, the performance of the best frequentist and Bayesian methods was similar for the same combinations of factors (k and mean replication), though the Bayesian approach had higher than nominal (>95%) coverage for the mean effect when k was very low ($k < 15$). Our literature review suggests that many meta-analyses that used z distribution or bootstrapping CIs may have overestimated the statistical significance of their results when the number of studies was low; more appropriate methods need to be adopted in ecological meta-analyses.

KEYWORDS

bias, confidence interval, coverage, credible interval, effect size, global climate change, log response ratio, sample size

1 | INTRODUCTION

Meta-analysis uses statistical techniques to quantitatively summarize information from different studies and is highly influential in the contemporary practice of science. To conduct a meta-analysis, an investigator gathers summary statistics from each study to calculate an effect size, with the goal of computing an overall effect size (and its uncertainty) and exploring the factors contributing to variation in effect sizes (Nakagawa, Noble, Senior, & Lagisz, 2017). The use of meta-analysis in ecology has been growing rapidly since the 1990s, and has proven particularly useful in discerning general patterns by comparing information from different species, study sites and systems (Cadotte, Mehrkens, & Menge, 2012). Advice on best methodological practices for meta-analysis is widespread in disciplines with a longer history of meta-analytic research (e.g. medical sciences) but is lagging behind in ecology (Gates, 2002). This can be problematic because ecological meta-analyses have specific challenges not necessarily typically in other disciplines.

One pervasive characteristic of ecological meta-analyses is the high heterogeneity (i.e. large among-study variation in effect sizes). Senior et al. (2016) analysed 86 meta-analyses in ecology and evolution and found that the among-study variation averaged 92% of the total variance. In contrast, a review of 509 meta-analyses in medicine found that there was no detectable among-study variation in 50% of the studies (Higgins, Thompson, & Spiegelhalter, 2009). Ecological studies also differ from many other disciplines in the typical level of within-study replication, which is fewer than 10 replicates per study (Hillebrand & Gurevitch, 2014). Such low levels of replication will influence the precision of the estimates of effect size from the primary studies (Langan, Higgins, & Simmonds, 2016). Importantly, the low level of replication typical of ecological studies is outside the range used in most simulation studies designed to assess meta-analytic methods, which typically range from dozens to hundreds (Langan et al., 2016). Thus differences between ecology and other disciplines potentially limit the insights ecologists can gain from existing simulations that compare different meta-analytic methods.

Specific advice for conducting ecological meta-analyses includes suggestions on the type of meta-analytic model and effect size calculation to use (Gurevitch & Hedges, 1999; Lajeunesse, 2015; Osenberg, Sarnelle, Cooper, & Holt, 1999), and how to deal with non-independence (Gurevitch & Hedges, 1999; Noble, Lagisz, O'dea, & Nakagawa, 2017). For example, a random-effects model is often recommended for ecological meta-analysis over a fixed-effects model (Gurevitch & Hedges, 1999), and multi-level models are increasingly being used to incorporate the non-independence commonly found in ecological meta-analyses (Nakagawa & Santos, 2012). A topic addressed in the medical literature that has received little attention in ecology (but see Adams, Gurevitch, & Rosenberg, 1997) is the choice of confidence interval (CI) used to estimate the mean effect size in a meta-analysis (Hartung & Knapp, 2001; Sánchez-Meca & Marín-Martínez, 2008; Sidik & Jonkman, 2003).

Simulation studies have shown that when the number of studies (k) in the meta-analysis is low, the CIs for a mean effect size calculated using a normal approximation are too narrow, leading to coverage below the nominal level (i.e. a 95% CI should include the true value 95% of the time; Brockwell & Gordon, 2001; Sánchez-Meca & Marín-Martínez, 2008). To avoid this problem, meta-analyses in the medical literature often use the HKSJ (Hartung–Knapp–Sidik–Jonkman; Hartung & Knapp, 2001; Sidik & Jonkman, 2003) method, which is based on a t distribution and can achieve good coverage even when k is small (Inthout, Ioannidis, & Borm, 2014). On the one hand, bootstrap techniques have been recommended for estimating CIs for means in ecological meta-analyses, due to its robustness to departures from normality (Adams et al., 1997). On the other hand, bootstrapped CIs can lead to poor coverage when estimating the among-study variance (Viechtbauer, 2007).

Bayesian methods, and the credible interval, offer an alternative approach to estimating uncertainty in meta-analyses. Although Bayesian methods may have a steep learning curve, they offer advantages in handling hierarchical models for incorporating prior information and for dealing with missing data (Ogle, Barber, & Sartor, 2013). Bayesian meta-analytic techniques produce a posterior distribution of the mean effect size and associated variance terms. Estimates of uncertainty, including credible intervals, can be directly obtained from the posterior distributions, which offer a more interpretable alternative to the frequentist-based CI (Kruschke & Liddell, 2018).

Our main goal is to compare the performance of traditional and Bayesian methods to measure the uncertainty around the estimation of a mean effect in the context of ecological meta-analysis. To achieve this goal, we conducted a two-pronged study. First, we reviewed published ecological meta-analyses to characterize the types of CI used in ecological meta-analyses, the number of replicates used in the primary studies (n_i) included in published meta-analyses and the number of studies (k) that were aggregated to calculate a mean effect size. Second, we used the n_i and k found in our literature review to inform the range of parameter values to use in conducting simulation experiments relevant to ecological meta-analyses. In particular, we determined the typical levels of n_i, k and the among-study variance, and then varied them systematically in our simulation studies. We then evaluated the performance of frequentist and Bayesian meta-analysis methods when applied to the simulated data, especially with respect to their ability to estimate the true mean effect and among-study variance, and their quantification of uncertainty intervals (i.e. confidence or credible intervals). Based on our findings, we generate recommendations on the methods to measure uncertainty that perform best for ecological meta-analysis and highlight how simple choices (sometimes overlooked by the investigators) can affect the results of meta-analyses.

2 | MATERIALS AND METHODS

2.1 | Literature review to assess characteristics of ecological datasets

2.1.1 | Literature search

We searched the Core Collection of the ISI Web of Science database in March 2017; the search string for TOPIC included (["meta-analy*"

OR "metaanaly*" OR "meta analy*"] AND ["climate change" OR "global change"]). We only included articles and reviews within the "Ecology", "Environmental Sciences", "Biodiversity Conservation" and "Plant Sciences" categories. The search resulted in 581 citations; the PRISMA diagram detailing the screening process is provided in Figure S1. After abstract screening, we checked the full text of the 205 articles published between 2013 and 2016. Of these, 96 papers satisfied the inclusion criteria for the final analysis.

2.1.2 | Criteria for inclusion

We focused on narrow sense meta-analyses: that is those that used a quantitative meta-analytic method to combine effect sizes that compared a control and a treatment group. We excluded studies that (a) only cited published meta-analyses, (b) reviewed meta-analytic methods, but did not perform a meta-analysis, (c) were identified as meta-analysis by the authors but did not use a meta-analytic model or did not calculate effect sizes, (d) used the correlation between two variables as an effect size and (e) were not 'biological meta-analyses' (as defined in Nakagawa et al., 2017), such as studies related to human health or human behaviour.

2.1.3 | Information extracted

For each paper, we extracted the number of studies (*k*) from the text, figure captions, figures and supplementary materials. Here we define a 'study' as yielding an estimate of an effect, so that a given primary paper could generate multiple effects and thus multiple studies. The *k* values were determined at three levels, (a) overall: that is, the total *k* collected by the authors (e.g. if they conducted meta-analyses on different response variables, then we summed the *k* across these variables); (b) analysis: that is, the total *k* used in a particular analysis (e.g. if an analysis examined variation among four levels of a moderator, then we summed up the number of studies in each level) and (c) category: that is, the *k* included in each category of a categorical analysis. In some cases, the authors calculated mean effect sizes for different categories separately and only compared the categories using CIs (i.e. there was no integrated analysis incorporating a category effect). In this case, we considered each categories' *k* to apply at the 'analysis' level.

When available, we also recorded the number of replicates (n_i) in the original studies. If the level of replication was unequal for the control and treatment groups, we recorded the average. Finally, from each meta-analysis, we also recorded the inferential paradigm used (frequentist vs. Bayesian) and the method used to obtain CIs for the frequentist approaches (e.g. nonparametric bootstrap, normal-based, KHSJ, etc.).

2.2 | Simulation experiments

Our literature review showed that 67% of the reported primary studies had <10 replicates. In addition, the review of meta-analyses in ecology and evolution by Senior et al. (2016) showed that among-study variation was important, and typically large, in ecological studies. Given these characteristics of ecological data, we simulated data in a full-factorial design that considered the following levels: mean number of replicates $n = \{3, 5, 10, 15, 20, 30\}$, number of studies $k = \{5, 10, 15, 25, 35, 50\}$ and among-study variance $\sigma^2_{\text{among}} = \{0.1, 0.25, 0.5, 1, 2, 5\}$. We simulated 2,000 replicated meta-analyses for each combination of n, k and σ^2_{among}. We then evaluated the performance of four meta-analytic methods applied to the simulated data: three frequentist approaches that differed in how they calculated CIs for a mean effect and a Bayesian approach.

2.2.1 | Simulating raw data for a study

We first determined the number of replicates for study i (n_i) based on a random draw from a Poisson distribution:

$$n_i^* \sim \text{Poisson}\,(n-2), \tag{1}$$

$$n_i = n_i^* + 2, \tag{2}$$

where n is the mean number of replicates representative of ecological meta-analyses. We subtracted 2 to sample from the Poisson and added 2 to the simulated n_i^* to make the minimum number of replicates for each simulated study equal 2 rather than 0. For each study, we assumed equal number of replicates for the control and treatment groups.

Individual observations ($j = 1, 2, \ldots n_i$) for the control and treatment groups were generated from a log-normal distribution (LN) such that for study i and observation j:

$$y_{C_{ij}} \sim \text{LN}\left(0, \sigma^2_{\text{rep}}\right), \tag{3}$$

$$y_{T_{ij}} \sim \text{LN}\left(0 + \mu + \varepsilon_i, \sigma^2_{\text{rep}}\right), \tag{4}$$

where σ^2_{rep} is the among-replicates variation, μ is the true overall effect, and $y_{C_{ij}}$ and $y_{T_{ij}}$ are the simulated observations for study i and observation j of the control and treatment group respectively. We set the among-replicate variation equal to 1 for both the control and treatment. For convenience, we set the location parameter for the control group equal to 0, resulting in median $(y_C) = 1$. For the treatment group in study i, we set median $(y_T) = \mu + \varepsilon_i$, where μ is the overall true treatment effect (hereafter, true effect size) and ε_i is the random effect associated with study i. We simulated ε_i as:

$$\varepsilon_i \sim N\left(0, \sigma^2_{\text{among}}\right). \tag{5}$$

Thus, the true effect size from any given study departs from μ due to its random effect (determined by ε_i), while the estimated effect size differs from the true effect size due to within-study sampling error (i.e. as influenced by n_i and σ^2_{rep}). The range of values used for σ^2_{among} were chosen to produce a similar distribution of I^2 (the proportion of variation

among effect sizes not explained by sampling error) to that reported by Senior et al. (2016) for meta-analyses in ecology and evolution (I^2 simulation results are presented in Figure S2).

2.2.2 | Estimating the effect size and within-study variance

Using the raw data simulated from each study, we computed the observed effect size for study i as the log response ratio ($\ln RR_i$), which is widely used in ecology (Nakagawa & Santos, 2012) and it is often a reasonable approximation of ecological phenomena (Osenberg, Sarnelle, & Cooper, 1997):

$$\ln RR_i = \ln\left(\frac{\bar{y}_{T_i}}{\bar{y}_{C_i}}\right), \tag{6}$$

where $\bar{y}_{T_i}$ and $\bar{y}_{C_i}$ are the sample means of the treatment and control groups respectively.

The expected sample means for each treatment in a simulated study are $E\left(y_{C_{ij}}\right) = \exp\left(\sigma^2_{\text{rep}}/2\right)$ and $E\left(y_{T_{ij}}\right) = \exp\left(\mu + \varepsilon_i + \left(\sigma^2_{\text{rep}}/2\right)\right)$. Thus, the log of the ratio of the expected values for the treatment and control groups is $\mu + \varepsilon_i$, corresponding to what we call the true study-specific effect size.

We calculated the estimated within-study variance of the log ratio (eq. 1 in Hedges, Gurevitch, & Curtis, 1999; $\sigma^2_{\text{within}_i}$) as:

$$\sigma^2_{\text{within}_i} = \frac{SD^2_{T_i}}{n_{T_i} \cdot \bar{y}^2_{T_i}} + \frac{SD^2_{C_i}}{n_{C_i} \cdot \bar{y}^2_{C_i}}, \tag{7}$$

where SD_T and SD_C are the sample standard deviations of the treatment and control groups, respectively, and $n_{T_i} = n_{C_i} = n_i$ is the simulated number of replicates in study i.

2.3 | Meta-analytic approaches

Given that we simulated independent data to highlight how the choice of uncertainty interval affects the estimation of a mean effect, we used a standard random-effects model (Gurevitch & Hedges, 1999). We comment on how our results may change with a multi-level (hierarchical) model in the Discussion section. We assume the simulated effect size for study i ($\ln RR_i$, calculated from Equation 6) follows a normal distribution with mean θ_i (the true effect for study i) and within-study variance $\sigma^2_{\text{within}_i}$:

$$\ln RR_i \sim N\left(\theta_i, \sigma^2_{\text{within}_i}\right), \tag{8}$$

$$\theta_i \sim N\left(\mu, \sigma^2_{\text{among}}\right), \tag{9}$$

We assume $\sigma^2_{\text{within}_i}$ is known, as calculated via Equation 7. Likewise, the true study-specific effect size, θ_i, is assumed to follow a normal distribution with mean μ (the true overall effect) and among-study variance, σ^2_{among} (which is sometimes referred to as τ^2 in other meta-analytic papers).

We compared different methods to construct CIs for a mean effect (at the analysis level) within the frequentist methods vs. Bayesian credible intervals. For the frequentist-based analyses, we compared: (a) a CI based on a z distribution, which is a large sample approximation, (b) a weighted CI based on the HKSJ method, which does not assume a large sample and instead uses a t distribution and (c) bootstrap methods. For the Bayesian-based analysis, we calculated the highest posterior density (HPD) credible interval.

2.3.1 | Frequentist approaches

We applied the random-effects model described by Equations 8 and 9 with inverse variance weights using the 'rma' function in the R package METAFOR (Viechtbauer, 2010), and estimated σ^2_{among} with the default restricted maximum likelihood (REML) method. To calculate the z distribution CI, we used the default settings for the random-effects model in METAFOR, which returns a 95% CI for μ based on the normal distribution. To apply the HKSJ CI, we set the option knha = T in METAFOR. The resulting CI for μ is based on both a refined estimate of σ^2_{among} and a Student's t distribution (Hartung & Knapp, 2001; Sidik & Jonkman, 2003), which accounts for the fact that σ^2_{among} is estimated and not known. For the bootstrapped CI, we estimated the bias-corrected nonparametric bootstrapped 95% CI for both μ and σ^2_{among} via the BOOT package in R (Canty & Ripley, 2017). Since the choice of HKSJ or z distribution for the μ CI does not affect the estimation of σ^2_{among}, in both cases we used METAFOR's function 'confint' to obtain the CI for σ^2_{among} ('confint' applies a Q-profile method in combination with REML).

2.3.2 | Bayesian approach

We used a 'hybrid' Bayesian framework to implement the random-effects model (Equations 8 and 9) in which we treat σ^2_{within} as known, whereas a fully Bayesian model may treat σ^2_{within} as unknown (this hybrid model is comparable to the 'empirical Bayes' method discussed in Schmid & Mengersen, 2013). Initial explorations with full and hybrid models gave qualitatively similar results and we only include the hybrid model in our analysis.

We specified relatively non-informative priors for the unknown quantities (e.g. μ and σ^2_{among}). For the mean effect size, μ, we specified a conjugate normal prior with a mean of 0 and large variance: $N(0, 10{,}000)$. Given that even diffuse priors for σ^2_{among} can influence the posterior for σ^2_{among}, particularly under small group size (Gelman, 2006), we explored five different priors for σ^2_{among} (Figures S12–S15). For the final analysis, convergence statistics and computational speed led us to focus on the Uniform(0, 10) prior for the standard deviation (σ_{among}).

The Bayesian meta-analyses were implemented in JAGS with the RJAGS R package (Plummer, 2018). For each model, we ran

three parallel Markov chain Monte Carlo (MCMC) sequences for 200,000 iterations, and discarded the first 100,000 iterations as the burn-in period. We used the $\hat{R}$ convergence diagnostic (Gelman & Rubin, 1992) to evaluate convergence of the MCMC sequences to the posterior. For the final simulations, we only included runs that had $\hat{R} < 1.1$, and checked that the proportion of discarded runs was lower than 1%. Using post-burn-in MCMC samples, we computed posterior means for quantities of interest (e.g. μ and σ^2_{among}) as point estimates. We computed 95% credible intervals as HPD intervals for both μ and σ^2_{among} using the 'HPDinterval' function in the CODA package (Plummer, Best, Cowles, & Vines, 2006).

2.4 | Implementation and assessment of the meta-analysis approaches

We ran all the analyses and simulations in the R environment (R Core Team, 2019); code is provided in the Supporting Information. For each simulated dataset, we estimated μ and σ^2_{among} via the frequentist and Bayesian methods described above. We summarized the results from the 2,000 replicated meta-analyses for each combination of factors (n, k, σ^2_{among}) and modelling approaches (i.e. frequentist and Bayesian methods to measure uncertainty). The results for the model performance associated with estimating σ^2_{among} are presented in Figures S7–S10.

We evaluated model performance using coverage, width of the uncertainty intervals, bias and efficiency. We estimated *coverage* for both μ and σ^2_{among} as the proportion (out of the 2,000 simulation replicates) of calculated 95% uncertainty intervals (CIs for the frequentist methods and credible interval for the Bayesian approach) that included the corresponding true value. Ideally, coverage should equal the nominal value of 0.95 (95%). CIs for these 'coverage proportions' were computed using the 'binom.confint' function in the R BINOM (Sundar, 2014) package, with the method 'wilson' (Agresti & Coull, 1998).

We summarized the perceived uncertainty for μ and σ^2_{among} as the mean *width of the 95% uncertainty intervals* for the 2,000 intervals for each scenario, and assessed how well the mean width was estimated using a 95% CI based on a t distribution. All else being equal, smaller uncertainty is a desirable feature, but not if it is accompanied by a reduction in coverage below the nominal level.

To evaluate *bias*, we calculated the mean difference between the point estimates for μ and σ^2_{among} and their true values based on the 2,000 simulation replicates, and report a 95% CI for this estimate based on the t distribution. Ideally, bias should be centred on zero.

Finally, to quantify the *efficiency* of the point estimates, we calculated the root mean squared error (RMSE) between the estimated and true values for μ and σ^2_{among} as:

$$\text{RMSE} = \sqrt{\frac{\sum_{s=1}^{N_{\text{sim}}} \left(\hat{a}_s - a_{\text{true}_s}\right)^2}{N_{\text{sim}}}}, \quad (10)$$

where $a = \mu$ or σ^2_{among}, $\hat{a}$ is the point estimate from each model, a_{true} is the true value used in the simulations and N_{sim} is the number of simulations.

3 | RESULTS

3.1 | Literature review to assess characteristic of ecological datasets

Of the 96 meta-analyses that satisfied our criteria (Table S1), 95 and 26 provided information on the number of studies (k) and number of replicates (n_i) associated with the original dataset respectively. Only three meta-analyses used a Bayesian approach. The majority of meta-analyses were published in *Global Change Biology* (23), followed by *Agriculture Ecosystems and Environment* (7) and *Ecology* (6) (Figure S3 displays the full list). The quality of reporting varied, and is discussed in more detail in the Supporting Information. We also provide additional information on k and n_i (by taxa, environment and topic) in the Supporting Information (Table S2; Figures S4 and S5).

3.1.1 | Number of studies

The number of studies (k) used to estimate an effect was highly skewed at the three levels we considered: overall, analysis and category (Figure 1). The overall k ranged from 25 to 32,567 (Figure 1a upper panel), with a median of 273 and with relatively few (12%) including more than 1,000 studies. For most papers, however, analyses were performed for different response variables or different moderators, and the k used for a particular analysis was considerably lower (Figure 1a middle panel), ranging from $k = 1$ (for a paper that presented all possible comparisons, even when one potential analysis was represented by only a single study) to $k = 8{,}474$, with a median of $k = 44$ (i.e. 50% of meta-analysis included 44 or fewer studies); 16% had $k \leq 10$. The number of studies included within categories ranged from $k = 1$ to 1,430, with a median of 16; 36% had $k \leq 10$ (Figure 1a lower panel).

3.1.2 | Number of replicates

The distribution of the reported number of replicates in the original studies (n_i) cited by the climate change meta-analyses was highly skewed, ranging from $n_i = 1$ to 21,600, with most studies having only a few replicates; the median was 5 (Figure 1b). The strong skewness in these data led us to inspect some of the original publications from which exceptionally large n_i values were reported. We found publications in which n_i values were likely misreported or greatly inflated by pseudoreplication (details in Table S3 and Figure S6).

3.1.3 | Analytic method to estimate the uncertainty interval for a mean effect

In 38.5% of the papers reviewed, the method used to calculate the frequentist-based CI for the mean effect was not mentioned

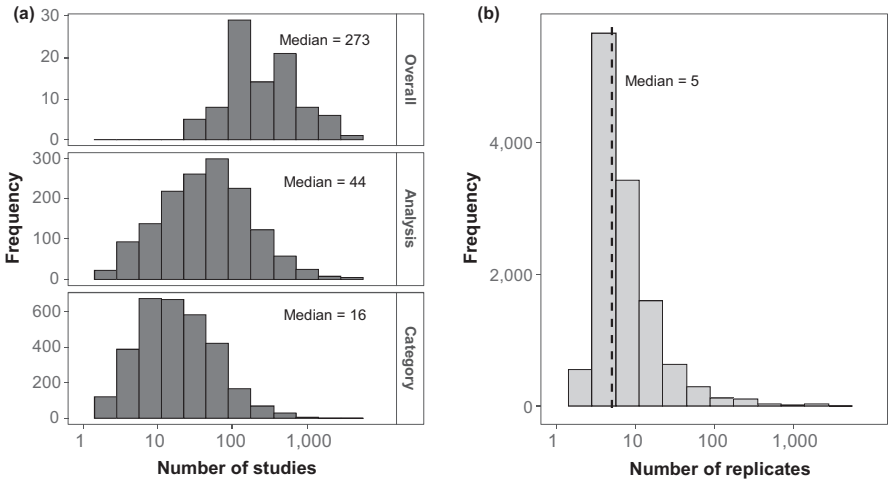


FIGURE 1 Results from the literature review of ecological meta-analyses: (a) distribution of the number of studies (k) reported for overall, analysis and category levels; the median k is indicated in each panel; (b) distribution of the number of replicates used in the original studies (n_i), as reported in each meta-analysis; the median n_i is indicated with a dashed line. Note that the x-axes are on a log scale

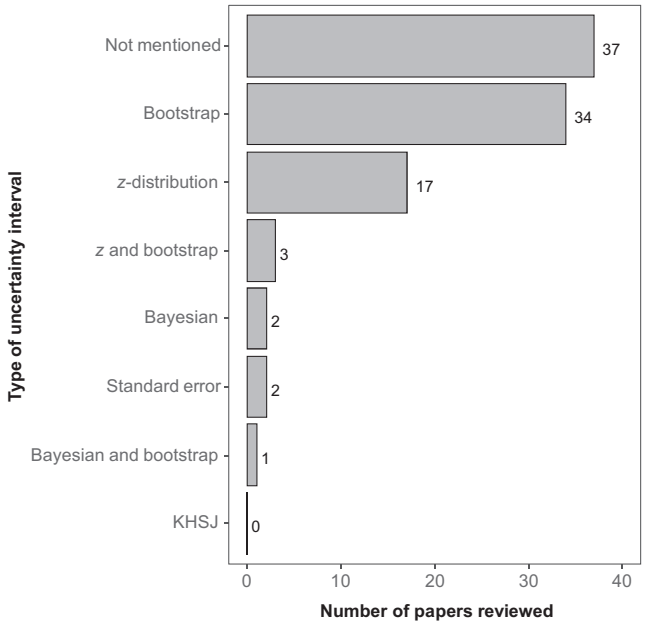


FIGURE 2 Types of uncertainty intervals reported by the ecological meta-analyses. In some cases, more than one type of uncertainty interval was reported

(Figure 2). Of the papers reporting how the CI was calculated, the majority used bootstrapped or z distribution CIs; only three papers used credible intervals (Bayesian method), and a few used a combination of methods (Figure 2). No papers reported using HKSJ method. Of the papers that did not specify the method, nine used Metawin (which defaults to a t distribution for the parametric CI, without the KHSJ refinement); 12 papers used the packages META or METAFOR in R (which default to a z distribution); and two used the Comprehensive Meta-Analysis software (which defaults to a z distribution). Assuming these 23 papers used the software defaults, then 31 papers used a z distribution and nine used a t distribution but without the KHSJ refinement. Thus, bootstrapped and z distribution CIs likely comprise the vast majority of approaches, with KHSJ CIs being entirely absent from our dataset.

3.2 | Simulation experiments: Estimation of a mean effect

The number of studies, k, used to estimate a mean effect size, μ, substantially affected the coverage of the frequentist methods, but this effect of k depended on the type of method used to estimate the 95% CIs (Figure 3a). For example, z distribution CIs for μ had coverage lower than the nominal level when $k < 40$, and coverage was appreciably lower for $k < 20$ (Figure 3a). Similarly, bootstrapped CIs had lower than nominal coverage when $k < 40$ (Figure 3a). In contrast, KHSJ CIs had close to nominal coverage over all values of k (Figure 3a). The Bayesian credible interval generally showed coverages around 95%, but when $k = 5$, coverage was >95% (Figure 3a).

Coverage can be smaller than nominal levels either because of bias or because the width of the uncertainty interval is inappropriately narrow (i.e. uncertainty is underestimated). The three frequentist methods for computing CIs for μ used the same approach for obtaining point estimates and had minimal bias centred on zero (Figure S11a,c,e). Thus, the observed differences in coverage for μ resulted from differences in the width of the uncertainty interval (Figure 3b). The Bayesian credible interval was generally wider than the frequentist-based CIs, and of the frequentist CIs, the KHSJ CI tended to be the widest; when k was small, the z distribution and bootstrapped CIs were ~1/3 smaller than they should be based on the more appropriate KHSJ CI (Figure 3b).

Increasing the mean number of replicates (n) in the primary studies did not greatly affect coverage (Figure 3b), the width of the uncertainty interval (Figure 3e), bias (Figure S11c) or RMSE (Figure S11d) for μ. Our results were likely produced because the among-study variation dominated within-study variation over the range of levels considered for the simulation factors (as determined by the review by Senior et al., 2016).

Increasing the among-study variance $\left(\sigma^2_{\text{among}}\right)$ increased the width of the uncertainty interval for μ (Figure 3f), but had only small effects on coverage (Figure 3c). Bias in the estimation of μ was negligible and unaffected by an increase in σ^2_{among} (Figure S11e), but the error in the estimation increased with the increase in heterogeneity (RMSE, Figure S11f).

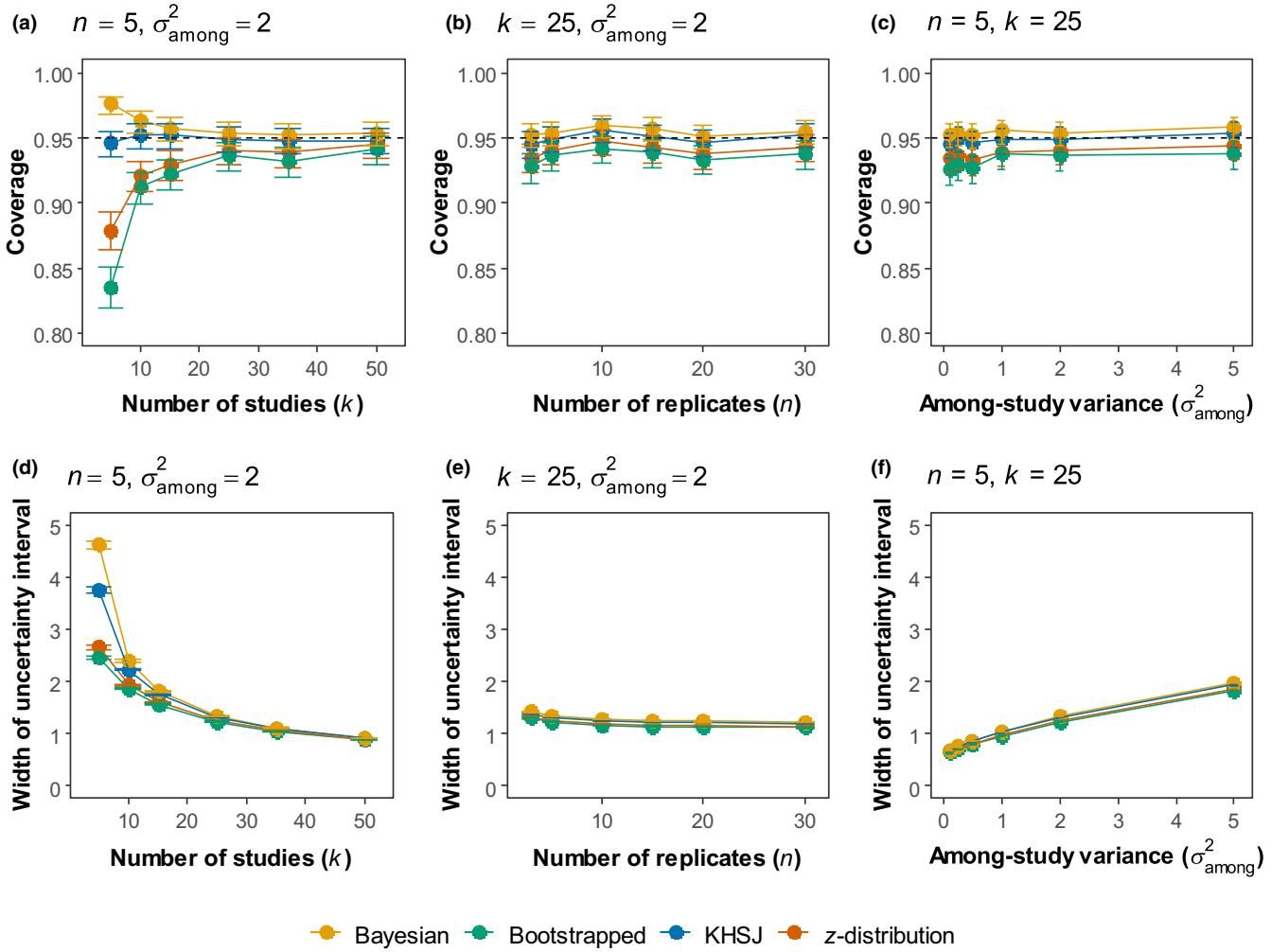


FIGURE 3 Coverage and the width of the 95% uncertainty interval for different methods used to estimate the mean effect size (μ) in a meta-analysis as a function of the number of studies (a, d), the mean number of replicates (b, e) and the among-study variance (c, f). The dashed horizontal line in panels (a)–(c) indicates the nominal value of 95%. Different colours denote the method used to estimate the uncertainty interval. Error bars provide the 95% CI

4 | DISCUSSION

Our literature review shows that ecological meta-analyses are highly variable in terms of how many studies (k) are included in the meta-analysis and the number of replicates reported in the original publications (n_i). Despite this high variability, both across and within meta-analyses, k and n_i tend to be low. The high frequency of meta-analyses with comparatively few studies ($k \leq 44$ in 50% of meta-analyses reviewed) is not unique to ecology; even lower number of studies are pervasive in medical research (Kontopantelis, Springate, & Reeves, 2013) where there has been an effort to develop methods that improve the performance of meta-analyses in such scenarios (Inthout et al., 2014). Furthermore, our simulations show that the method used to calculate an uncertainty interval greatly influences how often the interval includes the true mean effect and is very important for producing intervals with close to correct coverage when k is low. Despite its importance, a large proportion of the ecological meta-analyses we reviewed (38%) did not report the type of uncertainty interval used, and the ones that did report their methods used intervals that are problematic when k is low.

Low coverage of the z distribution CI when the number of observations (in the meta-analysis context, the number of studies, k) are low is well known in classical statistical contexts as well as in meta-analyses (Brockwell & Gordon, 2001; Hedges et al., 1999; IntHout et al., 2014). In meta-analyses, however, approaches typically default to assuming large k and thus justify the application of the z distribution. In ecology, this large sample approach is often unwarranted (Figure 1a). Furthermore, bootstrapped CIs are also well known to be problematic with small k (Hesterberg, 2015), although ecological meta-analyses tend to prioritize the potential for non-normal distributions over concerns about small k (Adams et al., 1997)—based on our results, such prioritization may be misplaced.

When k is low, the CI for a mean effect size (μ) based on the z distribution is too narrow. Some practitioners have addressed this problem by not calculating CIs when k is very small (e.g. Augusto, Delerue, Gallet-Budynek, & Achat, 2013). Others have resorted to using bootstrapped CIs (e.g. Thébault, Mariotte, Lortie, & MacDougall, 2014).

Given that bootstrapped CIs also had poor coverage when $k < 40$, this approach appears to be ill-advised. In our review, nearly half of the mean effect sizes used in an individual analysis were calculated with $k < 40$ effect sizes, where the choice of method for computing uncertainty intervals matters. As a result, many effects declared as significant probably should not have been. This is exemplified in a review of medical meta-analyses from the Cochrane Database, where of the 315 meta-analyses that yielded significant effects with the *z* distribution CI, only 79 were significant using the HKSJ CI (Inthout et al., 2014).

The default option for frequentist CIs for μ varies among software packages. For example, a *t* distribution CI (but without the HKSJ refinement) is Metawin's default, whereas the *z* distribution is the default in the Comprehensive Meta-Analysis software and in the R packages META and METAFOR (metafor is one of the most common software packages currently in use by ecologists). For those planning to conduct a random-effects meta-analysis using frequentist methods, we advise use of the HKSJ CI, which employs both a weighted estimator of the variance for the overall effect size and a *t* distribution for its associated CI (this can be set up in METAFOR using the option knha = T). Sánchez-Meca and Marín-Martínez (2008) report that the HKSJ method outperforms the simple CI-based on the *t* distribution. However, in some scenarios, coverage could be as low as 90% even using the HKSJ CI, for example, when heterogeneity is high, $k < 10$, and the number of replicates varies greatly among studies (Inthout et al., 2014). In our simulations that did not include highly uneven number of replicates, we showed that HKSJ CI's and the Bayesian credible intervals provide accurate (or at least conservative, >95%) coverage and performed best. We encourage researchers to be aware of the software defaults when calculating an uncertainty interval, and to report the method used.

The climate change meta-analyses showed exceedingly high variation in the number of replicates reported (n_i), spanning five orders of magnitude, but the majority of values were low. In fact, $n_i < 10$ in 67% of the cases, and $n_i \leq 5$ in 51% of the cases we reviewed. This pattern may be similar in other fields of ecology (Table S2; Figures S4 and S5). For example, a competition meta-analysis found n_i ranging from 1 to 1,455, with a median of 10 (Gurevitch, Morrow, Wallace, & Walsh, 1992). To obtain a more accurate estimate of μ, some authors specify a minimum n_i to calculate mean effect sizes (Gurevitch et al., 1992; Schirmel, Bundschuh, Entling, Kowarik, & Buchholz, 2016). Such censuring might improve CI performance by reducing variation in replication among studies (Inthout et al., 2014) but at the high cost of discarding important information. While one would in general expect better estimates with more replication, our simulation experiment did not show important effects of the mean number of replicates on the estimation of and inferences about μ. A similar insensitivity to the number of replicates has been observed in other studies (Sánchez-Meca & Marín-Martínez, 2008), although we included fewer replicates than most other simulations. Variation in replication among studies should produce variation in within-study variance, especially when the number of replicates is small. However, in our simulations, among-study variation was much larger than within-study variation, consistent with the characteristics of ecological meta-analyses (Senior et al., 2016), minimizing the role of variation in the number of replicates.

When the number of replicates reported (n_i) was unusually high, we checked a few of the original papers cited in each meta-analysis. Upon revisiting 17 of the original publications, we found at least 15 cases in which n_i was misreported (Table S3). This manifested in different ways. Some meta-analyses reported the total n_i in an experiment instead of the number of replicates per treatment. In other cases, authors reported the total n_i from repeated measurements or the numbers of individuals rather than the number of true replicates (e.g. plots or cages). There were also cases in which we were unable to verify the origin of the number reported in the meta-analysis. An incorrect n_i decreases the sampling variance for that effect size, which affects the weights and also the estimation of the overall heterogeneity (Noble et al., 2017). Researchers conducting a meta-analysis should be cautious when extracting data from the original studies to avoid misreporting (or inflating) the number of replicates. Publication of the data and code used to conduct a meta-analysis would also be useful to inform research on best practices for meta-analysis.

In our simulations using a random-effects model, the performance in the estimation of the among-study variance $\left(\sigma^2_{\text{among}}\right)$ was better when the true σ^2_{among} was high (Figures S4–S7). In agreement with Viechtbauer (2007), we observed that the Q-profile CI method for σ^2_{among} performed better than the bootstrap method (Figures S7–S10). The Bayesian method performed best, but had coverage above the nominal level when the number of studies was low ($k < 20$). Bayesian methods led to higher perceived uncertainty in such cases, which could be real, but this could also be a consequence of positive bias in the σ^2_{among} estimates, which was more pronounced for the Bayesian methods when $k < 20$. In this scenario, one approach to improve coverage is to use priors for σ^2_{among} that perform better when *k* is low (Gelman, 2006). Another solution is to specify more informative priors for σ^2_{among} based on a synthesis of past publications (Higgins et al., 2009). One reason to desire good estimation of σ^2_{among} is because overestimation of this variance component can lead to higher perceived uncertainty in the estimate of μ. An additional reason is that the estimates of σ^2_{among} represent real variation in effects and could be of importance in risk assessment.

In the initial explorations with the full Bayesian model, the MCMC chains for μ converged quickly, but they converged more slowly for σ^2_{among}, often falling into a 'zero variance trap' (Gelman, 2004) when the true among-study variance was close to zero. In general, convergence and mixing problems were most frequent for low *k* and low σ^2_{among}. While low σ^2_{among} is rare in ecology, low *k* is not. Of the priors we explored (Figures S12–S15), the folded-*t* and the uniform prior for the standard deviation performed best when *k* was low (we chose the uniform prior for the final simulations because it ran slightly faster). In our simulations, the hybrid Bayesian model

exhibited the practical advantages of the Bayesian methods (e.g. produces full posteriors and direct evaluation of uncertainty without approximating assumptions, among others), and was easy (and faster) to implement than the full model. In contrast, a full Bayesian approach may be more useful for multi-level models that include missing data, hierarchical structures and/or covariate effects (Ogle et al., 2013), and could benefit from informative priors for σ^2_{among}, particularly when k is low.

Our study simulated independent effect sizes. Often though, observed effect sizes are not independent (e.g. multiple observed effect sizes might be obtained from a single published article). As observed effect sizes within a group might respond similarly (due to similar methods, or similar environmental conditions), some of the among-study variation could be common to all members of a group or subgroup. Multi-level (hierarchical) models can be used to account for this. We believe that our results, including the insensitivity of our results to n, would not be materially altered in such situations, assuming the among-study variation still dominates the within-study variation. There are some challenges to be faced, however, when applying our results to more complex multi-level models. In particular, although the R package METAFOR has a function that handles multi-level models (rma.mv), the KHSJ adjustment is not available in this context, and the best that can be done with METAFOR is to construct t-based CIs of the mean (also referred to as conditional t test). For multi-level models, these t-based CIs have inflated error rates (Luke, 2017; C. Song, S.D. Peacor, C.W. Osenberg, & J.R. Bence, unpublished data), although they do outperform normal-based CIs (C. Song, pers. comm.). C. Song, S.D. Peacor, C.W. Osenberg, and J.R. Bence (unpublished data) speculated that the inflated error rates of t-based CIs resulted from not accounting for uncertainty in estimated variances. Methods exist for adjusting tests and CIs to account for uncertainty in estimated variances in multi-level models, such as the Kenward-Rogers adjustment, or simulation of null distributions (Halekoh & Hojsgaard, 2014), but to our knowledge these have not been implemented in any readily available software for conducting meta-analyses.

ACKNOWLEDGEMENTS

This research was funded by the U.S. Department of Energy (DE-SC-0010632), National Science Foundation (DEB-1655426 and DEB-1655394) and utilized Georgia Advanced Computing Resource Center resources. This is publication 2020-17 of the Quantitative Fisheries Center. We thank W. Viechtbauer for providing feedback on METAFOR computations, Natasja vanGestel and Kees Jan van-Groenigen for providing feedback during early stages of the project, Chao Song for helpful discussion and Sergio Estay for his feedback on the abstract in Spanish.

AUTHORS' CONTRIBUTIONS

All the authors conceived the idea; P.P. collected and analysed the data with contributions from C.W.O., E.A.H., J.R.B. and K.O.; P.P. led the writing; all the authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13445.

DATA AVAILABILITY STATEMENT

The data compiled in the literature review, the R code for the simulation experiment and the results from the simulation experiments are deposited in Dryad Digital Repository https://doi.org/10.5061/dryad.zw3r22863 (Pappalardo et al., 2020).

ORCID

Paula Pappalardo https://orcid.org/0000-0003-0853-7681
Kiona Ogle https://orcid.org/0000-0002-0652-8397
Elizabeth A. Hamman https://orcid.org/0000-0002-3494-6641
James R. Bence https://orcid.org/0000-0002-2534-688X
Bruce A. Hungate https://orcid.org/0000-0002-7337-1887
Craig W. Osenberg https://orcid.org/0000-0003-1918-7904

REFERENCES

Adams, D. C., Gurevitch, J., & Rosenberg, M. S. (1997). Resampling tests for meta-analysis of ecological data. *Ecology*, *78*(4), 1277. https://doi.org/10.2307/2265879

Agresti, A., & Coull, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119–126. https://doi.org/10.1080/00031305.1998.10480550

Augusto, L., Delerue, F., Gallet-Budynek, A., & Achat, D. L. (2013). Global assessment of limitation to symbiotic nitrogen fixation by phosphorus availability in terrestrial ecosystems using a meta-analysis approach. *Global Biogeochemical Cycles*, *27*(3), 804–815. https://doi.org/10.1002/gbc.20069

Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, *20*(6), 825–840. https://doi.org/10.1002/sim.650

Cadotte, M. W., Mehrkens, L. R., & Menge, D. N. L. (2012). Gauging the impact of meta-analysis on ecology. *Evolutionary Ecology*, *26*(5), 1153–1167. https://doi.org/10.1007/s10682-012-9585-z

Canty, A., & Ripley, A. (2017). *boot: Bootstrap R (S-Plus) functions*. R package version 1.3-20.

Gates, S. (2002). Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology*, *71*(4), 547–557. https://doi.org/10.1046/j.1365-2656.2002.00634.x

Gelman, A. (2004). Parameterization and Bayesian modeling. *Journal of the American Statistical Association*, *99*(466), 537–545. https://doi.org/10.1198/016214504000000458

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gurevitch, J., & Hedges, L. V. (1999). Statistical issues in ecological meta-analyses. *Ecology*, *80*(4), 1142–1149. https://doi.org/10.1890/0012-9658(1999)080[1142:SIIEMA]2.0.CO;2

Gurevitch, J., Morrow, L. L., Wallace, A., & Walsh, J. S. (1992). A meta analysis of competition in field experiments. *The American Naturalist*, *140*(4), 539–572. https://doi.org/10.1086/285428

Halekoh, U., & Hojsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – The R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–32. https://doi.org/10.18637/jss.v059.i09

Hartung, J., & Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, *20*(24), 3875–3889. https://doi.org/10.1002/sim.1009

Hedges, L. V., Gurevitch, J., & Curtis, P. S. (1999). The meta-analysis of response ratios in experimental ecology. *Ecology*, *80*(4), 1150. https://doi.org/10.2307/177062

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, *69*(4), 371–386. https://doi.org/10.1080/00031305.2015.1089789

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Hillebrand, H., & Gurevitch, J. (2014). Meta-analysis results are unlikely to be biased by differences in variance and replication between ecological lab and field studies. *Oikos*, *123*(7), 794–799. https://doi.org/10.1111/oik.01288

IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*, *14*(1). https://doi.org/10.1186/1471-2288-14-25

Kontopantelis, E., Springate, D. A., & Reeves, D. (2013). A re-analysis of the Cochrane library data: The dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE*, *8*(7), e69930. https://doi.org/10.1371/journal.pone.0069930

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206. https://doi.org/10.3758/s13423-016-1221-4

Lajeunesse, M. J. (2015). Bias and correction for the log response ratio in ecological meta-analysis. *Ecology*, *96*(8), 2056–2063. https://doi.org/10.1890/14-2402.1

Langan, D., Higgins, J. P. T., & Simmonds, M. (2016). Comparative performance of heterogeneity variance estimators in meta-analysis: A review of simulation studies. *Research Synthesis Methods*, *8*(2), 181–198. https://doi.org/10.1002/jrsm.1198

Luke, S. G. (2017). Evaluating significance in linear mixed-effect models in R. *Behavior Research Methods*, *49*, 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Nakagawa, S., Noble, D. W., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biology*, *15*, 18. https://doi.org/10.1186/s12915-017-0357-7

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*(5), 1253–1274. https://doi.org/10.1007/s10682-012-9555-5

Noble, D. W. A., Lagisz, M., O'dea, R. E., & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. *Molecular Ecology*, *26*(9), 2410–2425. https://doi.org/10.1111/mec.14031

Ogle, K., Barber, J., & Sartor, K. (2013). Feedback and modularization in a Bayesian meta–analysis of tree traits affecting forest dynamics. *Bayesian Analysis*, *8*(1), 133–168. https://doi.org/10.1214/13-BA806

Osenberg, C. W., Sarnelle, O., & Cooper, S. D. (1997). Effect size in ecological experiments: The application of biological models in meta-analysis. *The American Naturalist*, *150*(6), 798–812. https://doi.org/10.1086/286095

Osenberg, C. W., Sarnelle, O., Cooper, S. D., & Holt, R. D. (1999). Resolving ecological questions through meta-analysis: Goals, metrics, and models. *Ecology*, *80*(4), 1105–1117. https://doi.org/10.1890/0012-9658(1999)080[1105:REQTMA]2.0.CO;2

Pappalardo, P., Ogle, K., Hamman, E. A., Bence, J. R., Hungate, B. A., & Osenberg, C. W. (2020). Data from: Comparing traditional and Bayesian approaches to ecological meta-analysis. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.zw3r22863

Plummer, M. (2018). *rjags: Bayesian graphical models using MCMC*. R package version 4-8. Retrieved from https://CRAN.R-project.org/package=rjags

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*(1), 31–48. https://doi.org/10.1037/1082-989X.13.1.31

Schirmel, J., Bundschuh, M., Entling, M. H., Kowarik, I., & Buchholz, S. (2016). Impacts of invasive plants on resident animals across ecosystems, taxa, and feeding types: A global assessment. *Global Change Biology*, *22*(2), 594–603. https://doi.org/10.1111/gcb.13093

Schmid, C. H., & Mengersen, K. (2013). Bayesian meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 145–173). Princeton, NJ: Princeton University Press.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S. A., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: Its magnitude and implications. *Ecology*, *97*(12), 3293–3299. https://doi.org/10.1002/ecy.1591

Sidik, K., & Jonkman, J. N. (2003). On constructing confidence intervals for a standardized mean difference in meta-analysis. *Communications in Statistics – Simulation and Computation*, *32*(4), 1191–1203. https://doi.org/10.1081/SAC-120023885

Sundar, D.-R. (2014). *binom: Binomial confidence intervals for several parameterizations*. R package version 1.1-1.

Thébault, A., Mariotte, P., Lortie, C. J., & MacDougall, A. S. (2014). Land management trumps the effects of climate change and elevated CO_2 on grassland functioning. *Journal of Ecology*, *102*(4), 896–904. https://doi.org/10.1111/1365-2745.12236

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, *26*(1), 37–52. https://doi.org/10.1002/sim.2514

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pappalardo P, Ogle K, Hamman EA, Bence JR, Hungate BA, Osenberg CW. Comparing traditional and Bayesian approaches to ecological meta-analysis. *Methods Ecol Evol*. 2020;11:1286–1295. https://doi.org/10.1111/2041-210X.13445

Supporting Information for "Comparing traditional and Bayesian approaches to ecological meta-analysis"

Journal: Methods in Ecology and Evolution

Authors: P. Pappalardo, K. Ogle, E.A. Hamman, J.R. Bence, B.A. Hungate, & C.W. Osenberg

PRISMA diagram

For the main literature review to assess characteristics of ecological datasets we searched the Core Collection of the ISI Web of Science database in March 2017. The search string for TOPIC included (["meta-analy*" OR "metaanaly*" OR "meta analy*"] AND ["climate change" OR "global change"]). We only included articles and reviews within the "Ecology", "Environmental Sciences", "Biodiversity Conservation" and "Plant Sciences" categories.

We used a modified PRISMA diagram (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al., 2009) to describe the screening, eligibility and inclusion or studies in our literature review (Figure S1). Abstract screening was conducted using the viewer classification tool in the R package *metagear* (Lajeunesse, 2016). We only consider papers published between the years 2013-2016 for our review. The PRISMA diagram was constructed using the R package *PRISMAstatement* (Wasey, 2019):

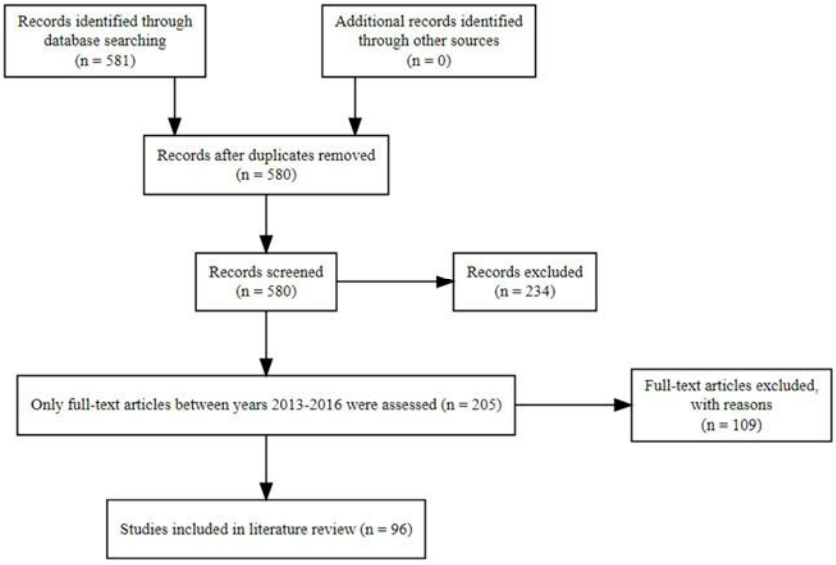


Figure S1. PRISMA diagram.

Distribution of I^2

To simulate datasets representative of ecological meta-analysis, we chose values of the among-study variance (for our combinations of mean number of replicates, *n*, and number of studies, *k*) that allowed us to generate a distribution of I^2 similar to the one reported by Senior et al. (2016) for meta-analysis in ecology and evolution. I^2 is useful for comparing variation between different studies because it measures the percentage of variation between effect sizes that it is not explained by sampling error. Senior et al. (2016) reported I^2 for 86 studies that included information on heterogeneity: median I^2= 84.67% and mean I^2 = 91.69% (Senior et al., 2016). Based on the I^2 calculated from our simulated datasets (Figure S2), an among-study variance of σ^2_{among} = 2 generated an I^2 similar to the mean I^2 observed by Senior et al. (2016) for values of *n* and *k* that are frequently found in the results of our literature review.

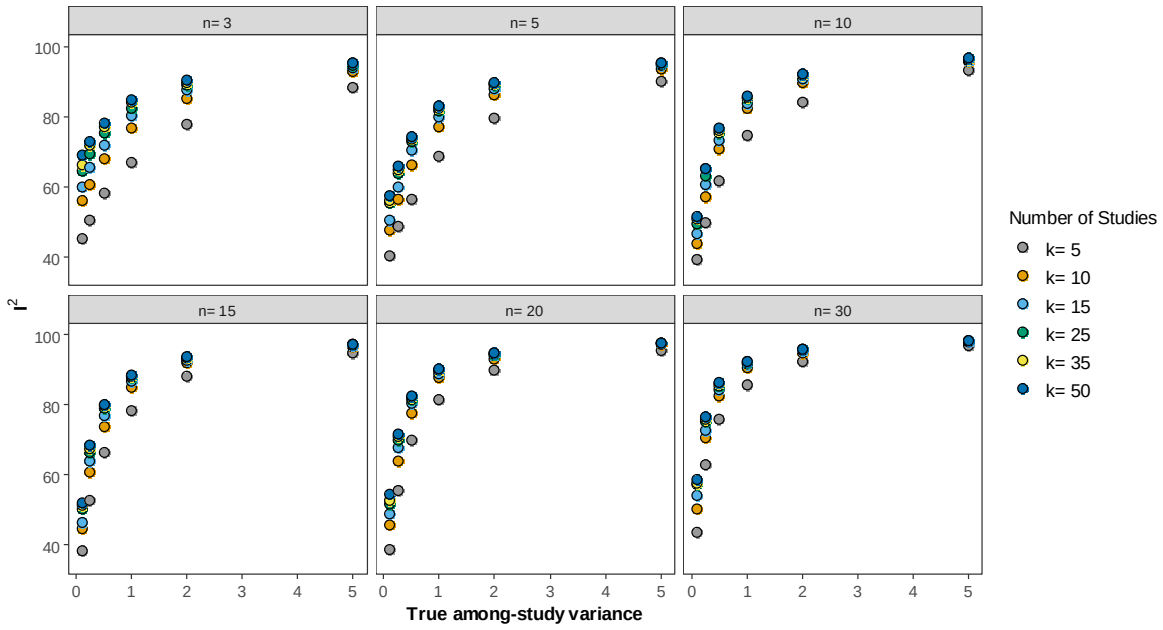


Figure S2. Mean I^2 as a function of the true (simulated) among-study variance for different combinations of the mean number of replicates, n_i, and number of studies, *k*, in the simulated datasets.

Studies included in the review of climate change meta-analyses

We reviewed a total of 96 meta-analyses on climate change experiments focused on ecological responses, from 40 journals. The journal that published most of the meta-analyses was *Global Change Biology*, followed by *Agriculture Ecosystems & Environment*, and *Ecology*. The number of climate change meta-analyses reviewed by journal is presented in Fig. S3, and a short name and DOI for each paper is presented in Table S1.

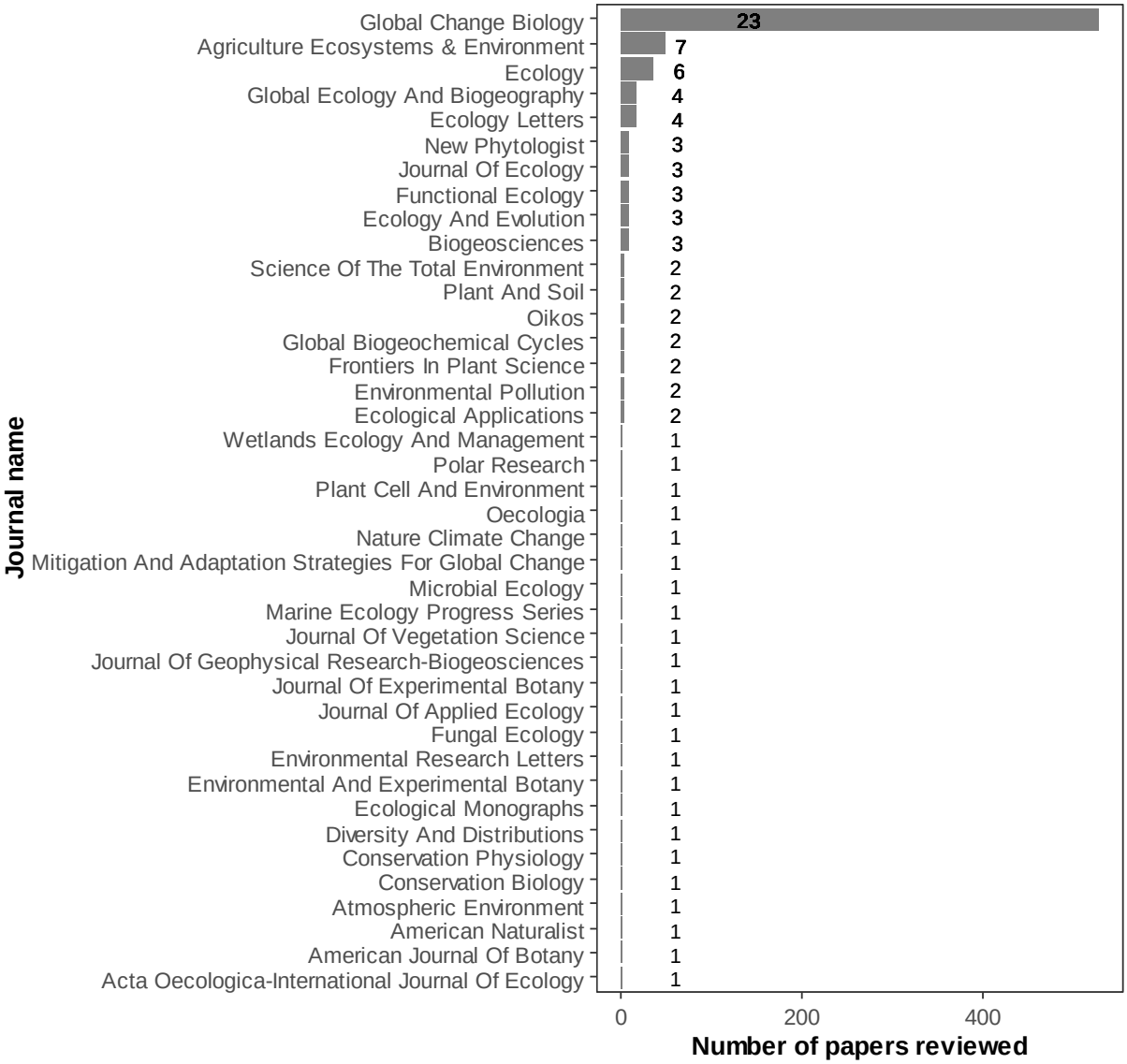


Figure S3. Number of climate-change meta-analyses reviewed, summarized by journal in which each was published, between 2013 and 2016.

Table S1. Climate change meta-analyses analyzed in the literature review.

Paper reviewed	Journal name	DOI
Abear et al 2014	Fungal Ecology	10.1016/j.funeco.2013.01.009
Aguilera et al 2013	Agriculture Ecosystems & Environment	10.1016/j.agee.2013.02.003
Albert et al 2015	Oikos	10.1111/oik.02512
Alldred and Baines 2016	Ecological Applications	10.1890/14-1525
Anderson 2016	New Phytologist	10.1111/nph.13693
Augusto et al 2013	Global Biogeochemical Cycles	10.1002/gbc.20069
Bai et al 2013	New Phytologist	10.1111/nph.12252
Baig et al 2015	Global Change Biology	10.1111/gcb.12962
Barrett and Hollister 2016	Polar Research	10.3402/polar.v35.25405
Bernhardt-Romermann et al 2015	Global Change Biology	10.1111/gcb.12993
Broberg et al 2015	Environmental Pollution	10.1016/j.envpol.2014.12.009
Chan and Connolly 2013	Global Change Biology	10.1111/gcb.12011
Chen et al 2015	Environmental Pollution	10.1016/j.envpol.2015.07.033
Deng et al 2015	Ecology	10.1890/15-0217.1
Diehl et al 2013	Journal Of Applied Ecology	10.1111/1365-2664.12032
Feng et al 2015	Global Change Biology	10.1111/gcb.12938
Finzi et al 2015	Global Change Biology	10.1111/gcb.12816
Gamfeldt et al 2015	Oikos	10.1111/oik.01549
Garcia-Palacios et al 2015	Global Change Biology	10.1111/gcb.12788
Garssen et al 2015	Global Change Biology	10.1111/gcb.12921
Goessling et al 2015	Functional Ecology	10.1111/1365-2435.12442
Gornish and Prater 2014	Journal Of Vegetation Science	10.1111/jvs.12150
Haddad et al 2014	Conservation Biology	10.1111/cobi.12323
Harvey et al 2013	Ecology And Evolution	10.1002/ece3.516
He and Dijkstra 2014	New Phytologist	10.1111/nph.12952
He and Silliman 2016	Ecological Monographs	10.1002/ecm.1221/suppinfo
Hollander and Bourdeau 2016	Ecology And Evolution	10.1002/ece3.2271
Humbert et al 2016	Global Change Biology	10.1111/gcb.12986
Iacarella et al 2015	Ecological Applications	10.1890/14-0545.1
Jackson 2015	Ecology	10.1890/15-0171.1
Jackson et al 2016	Global Change Biology	10.1111/gcb.13028
Jia et al 2016	Frontiers In Plant Science	10.3389/fpls.2016.01623
Jing et al 2016	Frontiers In Plant Science	10.3389/fpls.2016.01774
Kampf et al 2016	Science Of The Total Environment	10.1016/j.scitotenv.2016.05.067
Kivlin et al 2013	American Journal Of Botany	10.3732/ajb.1200558
Kroecker et al 2013	Global Change Biology	10.1111/gcb.12179
Kuebbing and Nunez	Global Change Biology	10.1111/gcb.12711
Lefevre 2016	Conservation Physiology	10.1093/conphys/cow009

Lemoine et al 2016	Ecology	10.1002/ecy.1506
Li et al 2016	Acta Oecologica-International Journal Of Ecology	10.1016/j.actao.2016.10.008
Liang et al 2016	Biogeosciences	10.5194/bg-13-2689-2016
Liu et al 2013	Plant And Soil	10.1007/s11104-013-1806-x
Liu et al 2014	Global Change Biology	10.1111/gcb.12517
Liu et al 2016	Global Change Biology	10.1111/gcb.13156
Llabres et al 2013	Global Ecology And Biogeography	10.1111/j.1466-8238.2012.00784.x
Loydi et al 2013	Journal Of Ecology	10.1111/1365-2745.12033
Lu 2015	Mitigation And Adaptation Strategies For Global Change	10.1007/s11027-014-9564-5
Lu et al 2013	Ecology	
Luo et al 2015	Ecology	10.1890/14-2228.1.sm
Mafongoya et al 2016	Agriculture Ecosystems & Environment	10.1016/j.agee.2016.01.017
Maillard and Angers 2014	Global Change Biology	10.1111/gcb.12438
Martinson and Fagan 2014	Ecology Letters	10.1111/ele.12305
McCary et al 2016	Ecology Letters	10.1111/ele.12562
McDevitt-Irwin et al 2016	Marine Ecology Progress Series	10.3354/meps11848
McGrath and Lobell 2013	Plant Cell And Environment	10.1111/pce.12007
Michalet et al 2014	Functional Ecology	10.1111/1365-2435.12136
Mundim and Bruna 2016	American Naturalist	10.1086/687530
Murphy and Romanuk 2014	Ecology And Evolution	10.1002/ece3.909
Oduor et al 2016	Journal Of Ecology	10.1111/1365-2745.12578
Orzechowski et al 2015	Global Change Biology	10.1111/gcb.12963
Pan et al 2016	Agriculture Ecosystems & Environment	10.1016/j.agee.2016.08.019
Paolucci et al 2013	Diversity And Distributions	10.1111/ddi.12073
Pardo et al 2015	Global Change Biology	10.1111/gcb.12806
Przeslawski et al 2015	Global Change Biology	10.1111/gcb.12833
Qiu 2015	Global Ecology And Biogeography	10.1111/geb.12360
Rodriguez-Castaneda et al 2013	Global Ecology And Biogeography	10.1111/j.1466-8238.2012.00795.x
Sasmito et al 2016	Wetlands Ecology And Management	10.1007/s11273-015-9466-7
Schirmel et al 2016	Global Change Biology	10.1111/gcb.13093
Shan and Yan 2013	Atmospheric Environment	10.1016/j.atmosenv.2013.02.009
Shantz and Burkepile 2014	Ecology	
Shantz et al 2016	Ecology Letters	10.1111/ele.12538
Shi and Han 2014	Global Biogeochemical Cycles	10.1002/2014GB004924
Skinner et al 2014	Science Of The Total	10.1016/j.scitotenv.2013.08.09

	Environment	8
Slattery et al 2013	Journal Of Experimental Botany	10.1093/jxb/ert207
Sorte et al 2013	Ecology Letters	10.1111/ele.12017
Strain et al 2014	Global Change Biology	10.1111/gcb.12619
Thebault et al 2014	Journal Of Ecology	10.1111/1365-2745.12236
Tian et al 2015	Agriculture Ecosystems & Environment	10.1016/j.agee.2015.02.008
Trap et al 2016	Plant And Soil	10.1007/s11104-015-2671-6
van der Kooi et al 2016	Environmental And Experimental Botany	10.1016/j.envexpbot.2015.10.004
van Lent et al 2015	Biogeosciences	10.5194/bg-12-7299-2015
Vicente et al 2016	Agriculture Ecosystems & Environment	10.1016/j.agee.2016.10.024
Wang et al 2013	Agriculture Ecosystems & Environment	10.1016/j.agee.2013.06.013
Wang et al 2014	Global Change Biology	10.1111/gcb.12620
Wooliver et al 2016	Functional Ecology	10.1111/1365-2435.12648
Worchel et al 2013	Microbial Ecology	10.1007/s00248-012-0151-6
Wu et al 2016	Agriculture Ecosystems & Environment	10.1016/j.agee.2016.06.028
Xu et al 2013	Biogeosciences	10.5194/bg-10-7423-2013
Yendrek et al 2013	Global Change Biology	10.1111/gcb.12237
Yoon and Read 2016	Oecologia	10.1007/s00442-016-3560-2
Yuan and Chen 2015	Nature Climate Change	10.1038/NCLIMATE2549
Yue et al 2015	Journal Of Geophysical Research-Biogeosciences	10.1002/2014JG002885
Zhang et al 2015	Global Ecology And Biogeography	10.1111/geb.12235
Zhao et al 2016	Global Change Biology	10.1111/gcb.13185
Zhou et al 2016	Global Change Biology	10.1111/gcb.13253
Zu et al 2013	Environmental Research Letters	10.1088/1748-9326/8/4/044027

Quality of reporting

The quality of reporting in the meta-analysis we reviewed varied, but was usually low. This agrees with Gates (2002) and his assessment that research on best practices for ecological meta-analyses has lagged behind other disciplines, such as medicine. We give a brief summary in this section based on the data compiled for our study (available as a an Excel file in the Dryad data repository https://doi.org/10.5061/dryad.zw3r22863). For example, only nine meta-analyses carefully explained the literature search by providing key words, explaining the screening process, and including the references of the original studies. Most papers (67 of 96) did not explain the screening process, or explained it only partially (18 of 96); only 11 papers provided a full explanation. 68 papers did and 28 did not provide specific keywords. The majority of papers (84 of 96) provided the references to the original articles, four provided partial information (Authors and year), five did not provide it, and in three cases the link to the Supporting Information (where this information may have been available) was broken and the corresponding author did not respond our email requesting the supplementary files for their paper.

Most meta-analyses (60 of 96) did not provide their original data (to replicate their effect size calculation for example), and eight provided some of the raw data collected; for eight we could not access the supplementary data (as explained above), and only 25 meta-analysis provided all the raw data compiled from the original publications. Twenty-three papers that did not provide the original data still provided the effect sizes, but only nine of these provided the variance or standard deviation of the estimated effect size, and only six of these papers provided information for the weights.

The meta-analytic model and the type of weighting used were not always specified. The type of meta-analytic model was reported more often, with random-effects and mixed-effects models being the more popular. In 13 papers a meta-analytic model was not mentioned, and in those papers, the authors appeared to use the confidence intervals to compare differences between effect sizes. Some papers combined more than one type of meta-analytic model, and others used standard, unweighted, tests (e.g., a t-test or ANOVA) to analyze the effect sizes. Forty papers weighted by inverse variance (two of this combined inverse variance with other weighting schemes), but 28 did not mention anything about weighting.

Additional data from our literature search

Exploratory literature search by sub-disciplines. At the beginning of our project, we conducted a literature search aimed at characterizing ecological meta-analyses in different ecological sub-disciplines or research areas. We searched the Core Collection of the Web of

Science database in May 2016 to obtain a set of meta-analyses from different sub-disciplines of ecology that represented highly cited meta-analyses, and thus have become some of the most influential examples in the field. The TOPIC search string was specified as (["meta-analy*" OR "metaanaly*" OR "meta analy*"] AND ["key words"]). The "key words" used to select the sub disciplines included "ocean acidification", "food limitation", "multiple stressors", or ["behavi*" AND "ecology"]. Subsequently, for each of these sub disciplines, we filtered the results by research area "Environmental Sciences and Ecology" and sorted the results by "times cited". This led to 65 meta-analyses publications that we considered further. We screened 64 abstracts and rejected 39 because they didn't meet our inclusion criterion (same as in the main literature search). We consulted the pdfs for the remaining 25 citations. The goal (not always achieved) was to have at least 10 well cited meta-analyses in each topic. For some topics such as food limitation and multiple stressors only a few meta-analyses were returned by the search; in those cases we reviewed all papers. When the same meta-analysis appeared in two sub-disciplines (e.g. "multiple stressors" and "ocean acidification"), we assigned it to one sub-discipline and only counted them once. From this preliminary search, we report here the results for the median number of replicates and the median number of studies (Table S2, Figure S4). For this preliminary search we did not collect information on type of grouping for the number of studies (overall, analysis, and category), but instead combined all types of comparisons.

Table S2. Exploratory literature search by research area. For each research area we show the number of meta-analyses reviewed, the median number of studies, and the median number of replicates.

Topic	Meta-analyses included	Median number of studies	Median number of replicates
Behavior ecology	7	14.5	5
Multiple stressors	7	11	4
Food limitation	6	9.5	4.5
Ocean acidification	5	10	No information

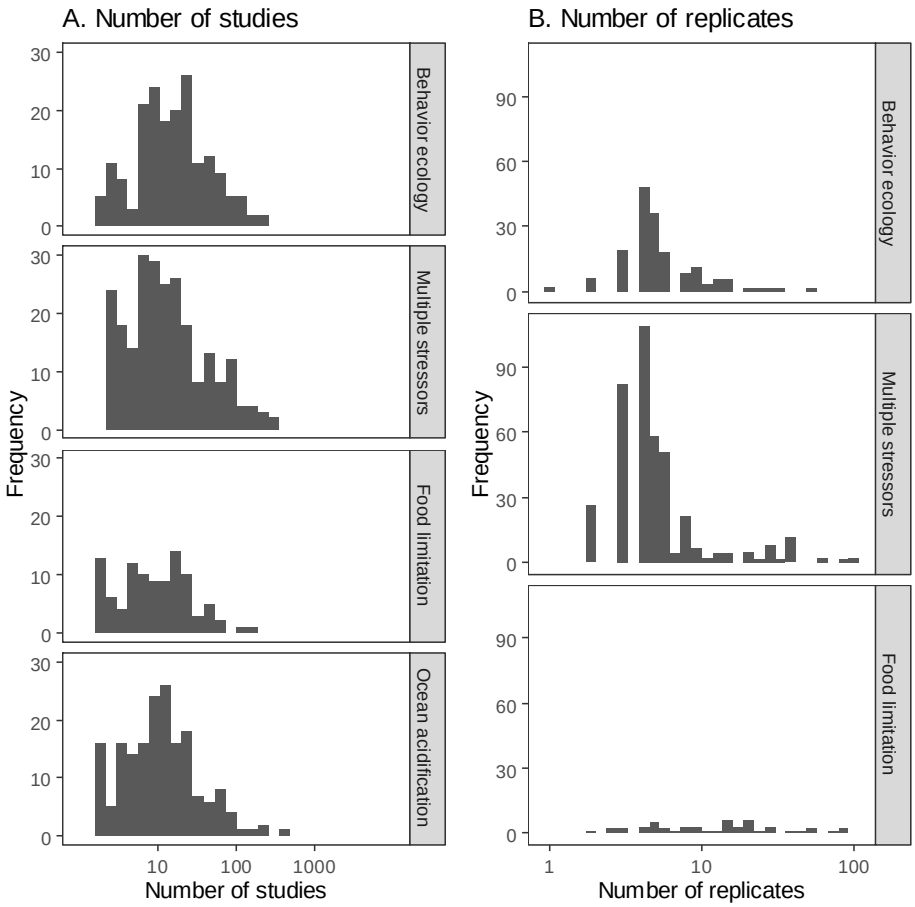
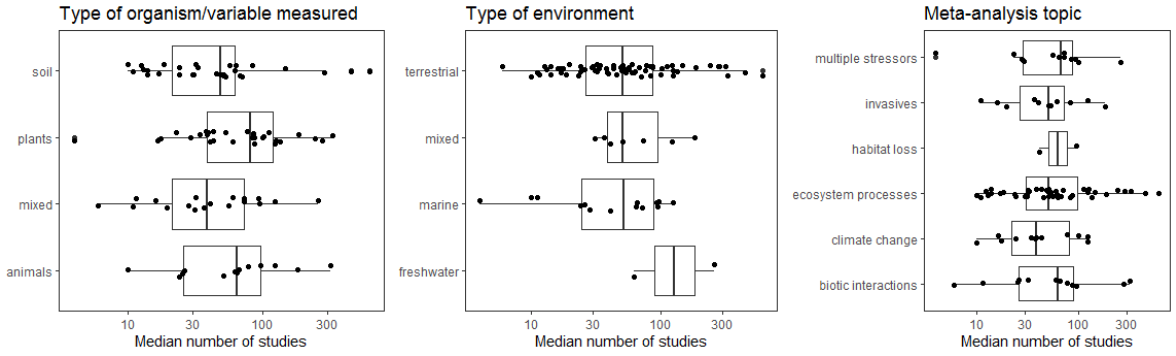


Figure S4. Results from the exploratory literature search on sub-disciplines of ecological meta-analyses. A) distribution of the number of studies (*k*) by sub-discipline; B) distribution of the number of replicates (n_i) used in the primary papers, as reported in each meta-analysis. Replication was not reported in any meta-analyses for ocean acidification. Note that the x-axes are on a log scale.

Meta-analyses on these other topics were generally consistent with the summaries for climate change (e.g., median *k*'s were all <40 and median n_i's were all <6). We decided on "climate change/global change" for the final literature search because it generated the most information and it is an important current topic that has generated many meta-analyses. Five papers found in the sub-disciplines search were also found (and included) when running the final climate change literature review.

Additional data for the global change meta-analyses. Because the climate/global change search was broad, we were able to categorize the meta-analyses into topic areas (multiple stressors, invasive species, habitat loss, ecosystem processes, climate change, biotic interactions), type of organism or variable measured, and type of environment. The patterns in these groups also were fairly consistent (Figure S5).

A. Median number of studies



B. Median number of replicates

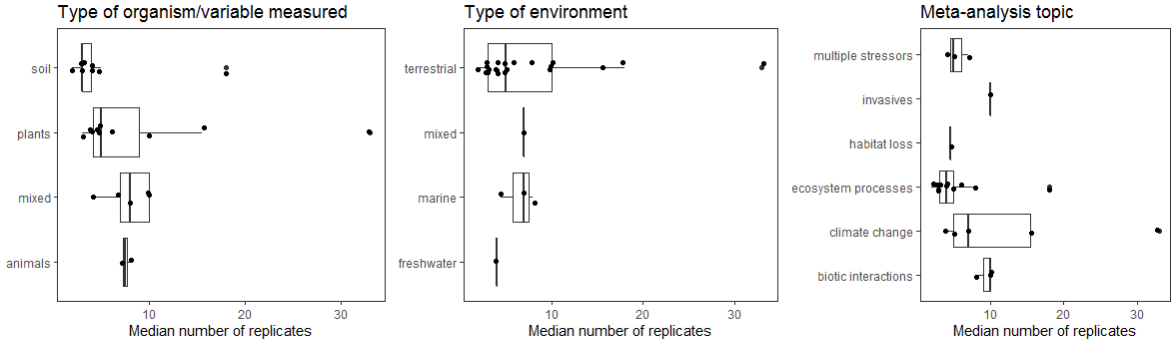


Figure S5. Additional results for the climate/global change meta-analysis. Variability on the median number of studies at the analysis level (A) and the median number of replicates (B) by type of organism (or variable) measured, type of environment, and meta-analysis topic.

Erroneous number of replicates

In some cases, the number of replicates, n_i, reported in the published meta-analyses of climate change seemed very high (i.e., in the hundreds or thousands). To examine those values, we sorted the number of replicates by mean n_i (the average number of replicates for the control and treatment groups when reported separately), and picked the top ten papers with the highest n_i and revisited the original publications. The meta-analyses were: 1) Anderson, 2016; 2) Oduor et al., 2016; 3) van Lent et al., 2015; 4) Hollander & Bourdeau, 2016; 5) Rodriguez-Castaneda et al., 2013; 6) Gornish & Prater, 2014; 7) Baig et al., 2015; 8) Shi & Han, 2014; 9) Lemoine et al., 2016; and 10) Jackson et al., 2016. From each meta-analysis paper, we aimed to double-check at least two of the original publications (if they were available to download) in which the reported number of replicates was high; however, Gornish & Prater (2014) had only one unusually high value and two of the original

publications were not accessible (reported by Shi & Han, 2014), so we checked 17 (not 20) original publications. The number of replicates appeared to have been inflated for 15 of the 17 publications (Table S3), due either to pseudoreplication in the original paper, poor reporting of the original data and/or study design, or misinterpretation on the part of the meta-analyst.

Table S3: Papers examined for evidence of misreporting in the number of replicates (n_i). We indicate the meta-analysis, the original paper cited in the meta-analysis, the number of replicates reported in the meta-analysis, and details on why we thought the number reported does not represent the actual number of replicates per treatment. The asterisk corresponds to the two sampling sizes that seem to have been reported appropriately.

Paper reviewed	Paper cited	Number of replicates reported	Details
Anderson, 2016	Pfeifer-Meister et al., 2013	2000	We were unable to verify the origin of the number reported by Anderson 2016, but 2000 is far greater than a typical number of experimental replicates.
Anderson, 2016	Gornish et al., 2015	21600	We were unable to verify the origin of the number reported by Anderson 2016, but 21600 is far greater than the number of experimental replicates.
Oduor et al., 2016	Wright & Stanton, 2007	518	518 is the minimum number of plants surviving to flower, which is greater than the experimental replicates; there were two replicates of 16 pollination crosses, 27 of those produced enough F2 seeds for the experiment and they were planted on five replicate blocks in each of six study sites.
Oduor et al., 2016	Bennington & McGraw, 1995	1280	1280 is the number of seeds from floodplain sites summed across all treatments and blocks, rather than the number of plots.
van Lent et al., 2015	Werner et al., 2006	1813	1813 is the total number of repeated measurements for a few control chambers for one experiment, rather than replicates for the control treatment.
van Lent et al., 2015	Verchot et al., 1999	256	The reported n_i is the total number of observations across all plots, rather than the number of plots per treatment.
Hollander & Bourdeau, 2016	Johnson & Black, 2008	951	We were unable to verify the origin of the number reported by Hollander & Bourdeau. Johnson & Black mentioned an average sample size of 52 snails collected from each of 83 populations.
Hollander & Bourdeau, 2016	Parsons, 1997	379*	379 is the number of snails at the Albany site used in the shell shape regression.
Rodriguez-Castaneda et	Farnsworth & Ellison, 1991	459	The value used was the number of leaves damaged from plants under a canopy versus

al., 2013			number of leaves damaged on plants growing in full sun, neglecting the limited number of plants.
Rodriguez-Castaneda et al., 2013	Delacerda et al., 1986	750	Rodriguez-Castaneda et al. reported the number of leaves damaged on 10 saplings.
Gornish & Prater, 2014	Way & Sage, 2008	400*	1000 seeds were grow to seedlings; they thinned seedlings per pot and had 400 trees per treatment per year.
Baig et al., 2015	Lavola et al., 2013	288	Baig et al. reported the total number of seeds in each chamber, rather than the number of chambers exposed to different temperature and CO_2 conditions.
Baig et al., 2015	Calfipietra et al., 2003	144	Baig et al. reported the number of trees across all three replicate plots, instead of the number of plots within each treatment (with or without CO_2).
Lemoine et al., 2016 (biodiversity)	Zhang et al., 2014	216	We were unable to verify the origin of the number reported by Zhang et al. The first experiment had 9 blocks involving 8 species. Each block had 36 plots. In total there were 324 plots. Only 2 replicates for spatial treatments were mentioned, but there were more treatments.
Lemoine et al., 2016 (biodiversity)	Allan et al., 2011	240	Allan et al. state the experiment contains 82 plots, and do not state how many plots are in each treatment.
Jackson et al., 2016	Muyssen et al., 2010	100	Jackson et al. cited the number of individuals from each clone (3) that they used in each treatment, but it is unclear how many independent setups were used.
Jackson et al., 2016	Jokinen et al., 2011	100	Jokinen et al. reported number of fish (n_i= 50) from each of two replicate cages per treatment in a tank (n_i= 100 per treatment). It appears the real number of replicates was the number of cages (n_i=2) inside each experimental tank.

Across all meta-analyses we reviewed, the number of replicates, n_i, reported varied from 1 to 2.16×10^4, with a median of 5 and a mean of 19.73. The distribution of n_i for each meta-analysis is presented in Figure S6.

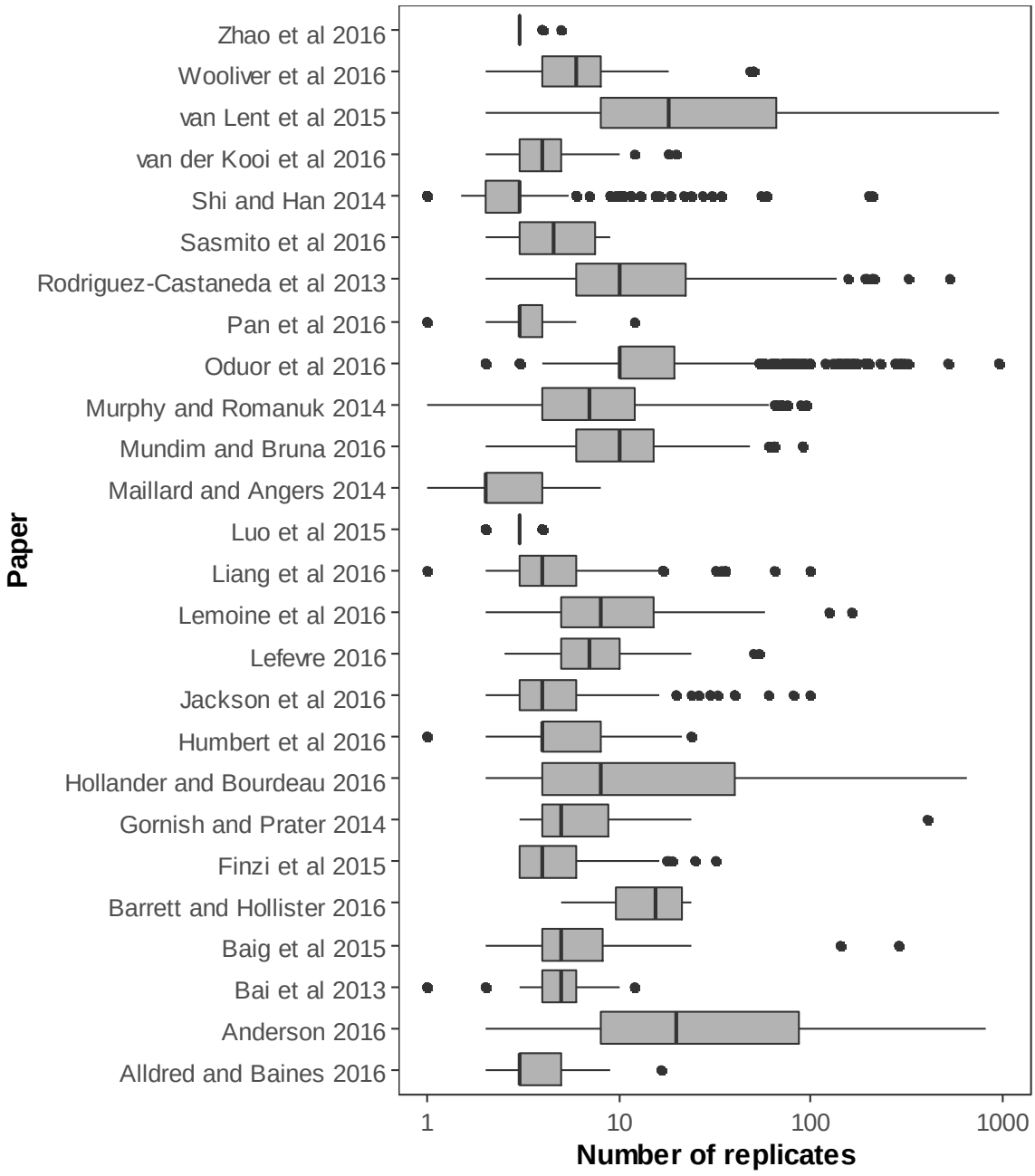


Figure S6. Distribution of the number of replicates, n_i, in the original studies for each of the 26 meta-analysis publications in our review that provided the original data. The boxplots represent the median (thick vertical line), the 25th and 75th percentiles (box), the upper whisker extends from the box to the larger value no further than 1.5xIQR, and the lower whisker extends from the box to the smallest value at most 1.5xIQR. Extreme values that exceed the whiskers are plotted individually as solid points.

Model performance associated with estimating the among-study variance (σ^2_{among})

Because other studies have explored the best methods to estimate the uncertainty intervals for the among-study variance (Viechtbauer, 2007; Veroniki et al., 2016) and it was not the main focus of this study, we only compared: 1) a parametric bootstrapped 95% CI (for the bootstrap approach), 2) a parametric CI using the Q-profile method (for both the *z*-distribution and HKSJ approaches), and 3) the HPD (high posterior density) interval of the posterior distribution of the among-study variance (for the Bayesian approach).

The ability to estimate σ^2_{among} was most affected by the true among-study variance and the number of studies. An increase in the number of studies, *k*, significantly improved the coverage of the bootstrap method to estimate the CI for σ^2_{among} in most scenarios (Figures S8A-S10A). However, an increase in the number of studies decreased coverage when the true among-study variance was very low (Figure S7A). In all combinations in which σ^2_{among} was moderate to high, the Q-profile and Bayesian methods showed coverage near or at the nominal values (Figures S8A-S10A); the Bayesian method yielded coverages >95% when $k < 20$, and the Q-profile method showed coverage <95% when $k > 20$.

Using dataset characteristics typical of ecological meta-analyses, our simulations showed that the coverage for σ^2_{among} was not significantly affected by the mean number of replicates, *n* (Figure S8B). However, when the true among-study variance was low, all methods failed to achieve good coverage for $n < 10$ (Figure S7B). When *n* and the true among-study variance were low ($n = 5, \sigma^2_{among} = 0.5$, Figure S7A), the coverage for σ^2_{among} was below the nominal level (95%) for all methods and combinations of *k* and *n*, except for the Bayesian method when $k < 25$. Coverage decreased substantially with increasing *k* (Figure S7A). On the other hand, when *n* was high, even if the among-study variance was low, there were cases in which coverage was close to the nominal level ($n = 20$, $\sigma^2_{among} = 0.5$, Figure S9A).

When the among-study variance was close to zero, coverage for σ^2_{among} was low for all methods (Figures S7C-S10C). This is likely the result of a small but positive bias in the estimates of σ^2_{among} (Figures S7I-S10I). With little real variation among studies, the perceived uncertainty was correctly viewed as low, but the bias led to CIs that frequently did not overlap the true values.

For the combination of simulation factors that we evaluated (Figures S7-S10), the Bayesian method over-estimated the among-study variance, when *k* was low or the among-study variance was large. The larger bias in the Bayesian method was also reflected in a large RMSE when *k* was low. The bootstrap approach to assessing uncertainty in the among-study variance underperformed in all combinations of parameters, rarely achieving the nominal level of coverage; this is likely because the width of the CI for the among-study variance was always low for the bootstrap approach (Figures S7-S10).

- **Simulation parameters: $n = 5, k = 25, \sigma^2_{among} = 0.5$**

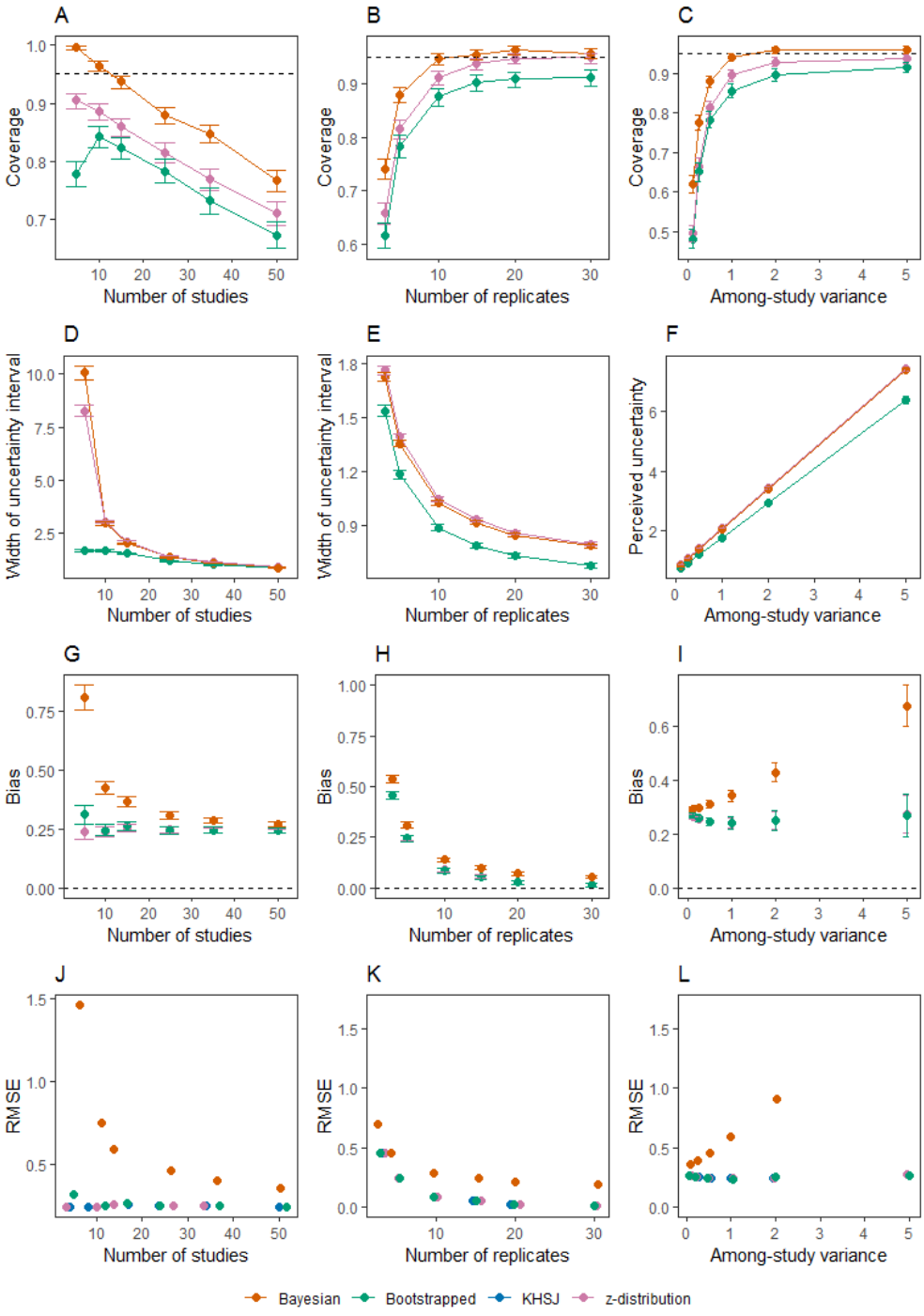


Figure S7. Performance measures of the estimation of the among-study variance as a function of the number of studies (left column), the number of replicates in the original studies (middle column) and the simulated among-study variance (right column). Performance was assessed using coverage (A, B, C), perceived uncertainty (width of the uncertainty interval) (D, E, F), bias (G, H, I), and RMSE (J, K, L). Error bars provide the 95% CI for panels A-I. Please note different scales in the y-axis for bias and width of the uncertainty interval. Simulation parameters: n = 5, k = 25, σ^2_{among} = 0.5, except for the cases in which that parameter was varied.

- **Simulation parameters: $n = 5, k = 25, \sigma^2_{\text{among}} = 2$**

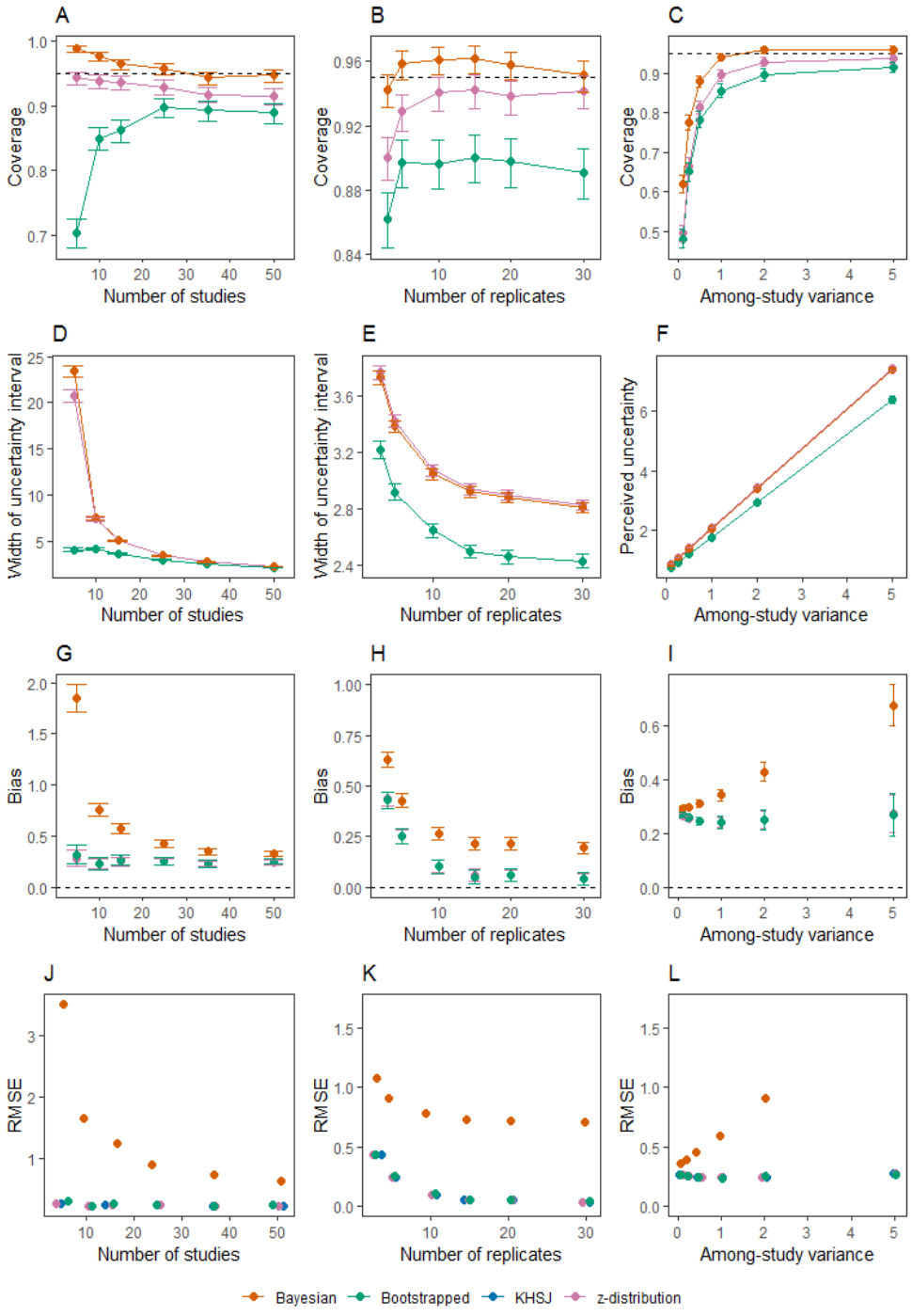


Figure S8. Performance measures of the estimation of the among-study variance as a function of the number of studies (left column), the number of replicates in the original studies (middle column) and the simulated among-study variance (right column). Performance was assessed using coverage (A, B, C), perceived uncertainty (width of the uncertainty interval) (D, E, F), bias (G, H, I), and RMSE (J, K, L). Error bars provide the 95% CI for panels A-I. Please note different scales in the y-axis for bias and width of the uncertainty interval. Simulation parameters: $n = 5, k = 25, \sigma^2_{\text{among}} = 2$, except for the cases in which that parameter was varied.

- **Simulation parameters: $n = 20, k = 25, \sigma^2_{among} = 2$**

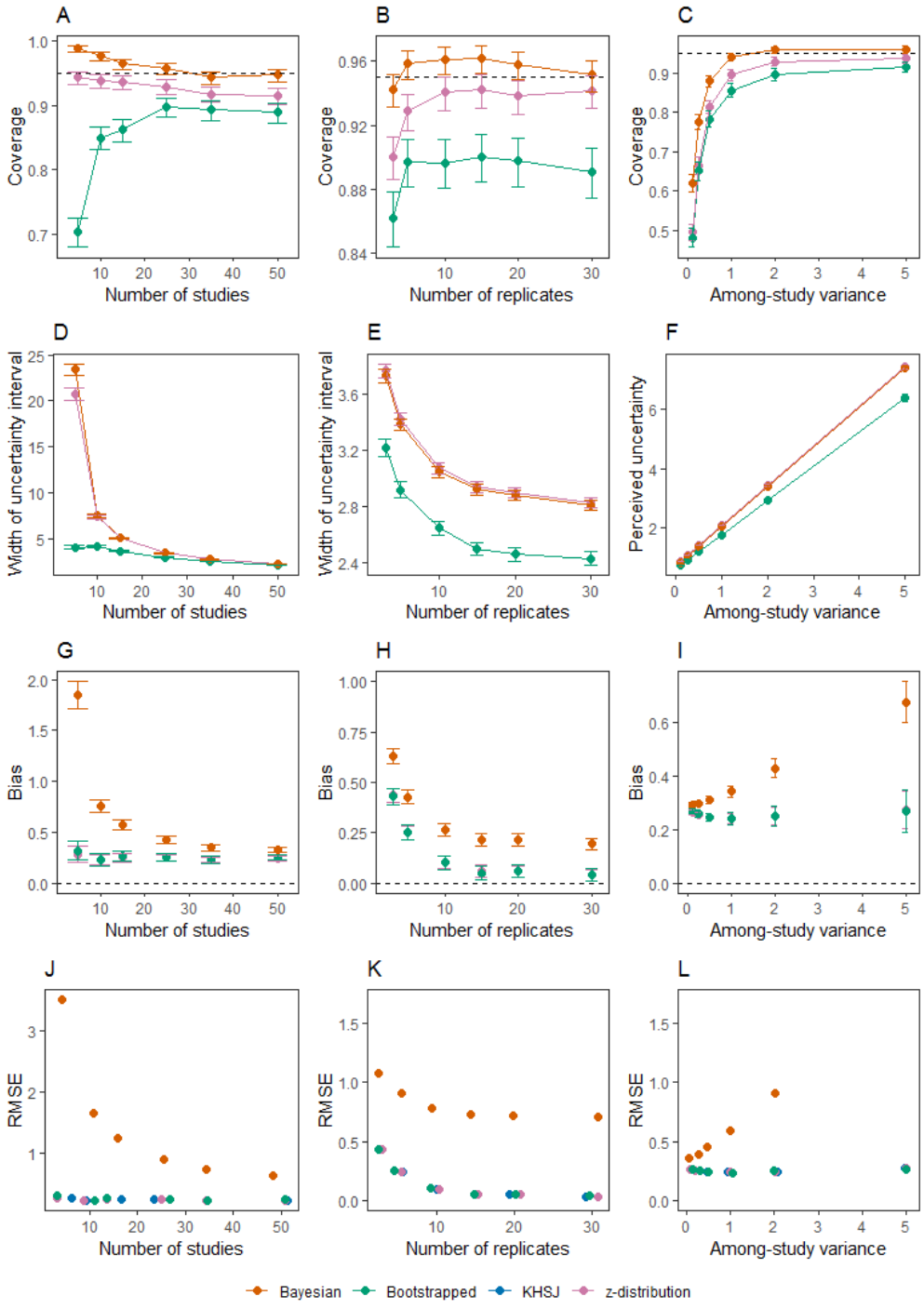


Figure S9. Performance measures of the estimation of the among-study variance as a function of the number of studies (left column), the number of replicates in the original studies (middle column) and the simulated among-study variance (right column). Performance was assessed using coverage (A, B, C), perceived uncertainty (width of the uncertainty interval) (D, E, F), bias (G, H, I), and RMSE (J, K, L). Error bars provide the 95% CI for panels A-I. Please note different scales in the y-axis for bias and width of the uncertainty interval. Simulation parameters: n = 20, k = 25, σ^2_{among} = 2, except for the cases in which that parameter was varied.

- **Simulation parameters: $n = 20, k = 25, \sigma^2_{\text{among}} = 0.5$**

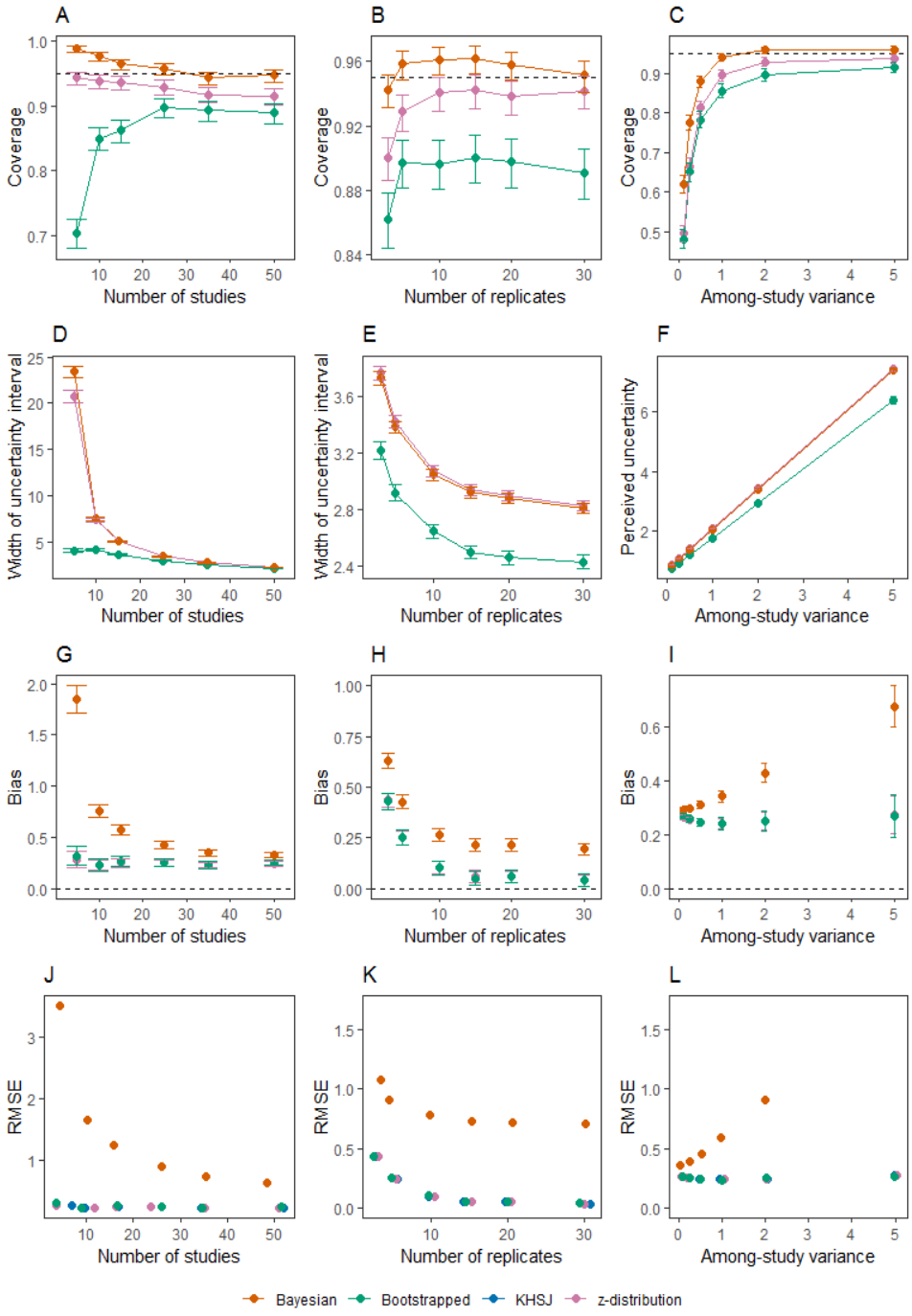


Figure S10. Performance measures of the estimation of the among-study variance as a function of the number of studies (left column), the number of replicates in the original studies (middle column) and the simulated among-study variance (right column). Performance was assessed using coverage (A, B, C), perceived uncertainty (width of the uncertainty interval) (D, E, F), bias (G, H, I), and RMSE (J, K, L). Error bars provide the 95% CI for panels A-I. Please note different scales in the y-axis for bias and width of the uncertainty interval. Simulation parameters: $n = 20, k = 25, \sigma^2_{\text{among}} = 0.5$, except for the cases in which that parameter was varied.

Bias and RMSE in the estimation of the mean effect

Here, we report the bias and RMSE for the four methods used to estimate the mean effect as a function of the number of studies (k), the mean number of replicates (n), and the among-study variance (σ^2_{among}). The RMSE was calculated using the "rmse" function from the package *Metrics* (Hamner & Frasco, 2018). All three frequentist methods yield the same point estimates for μ (and thus bias and RMSE) (Figure S11) but they may differ in their coverage, which also is affected by the width of the uncertainty interval. The Bayesian method gave very similar (but not identical) estimates of bias.

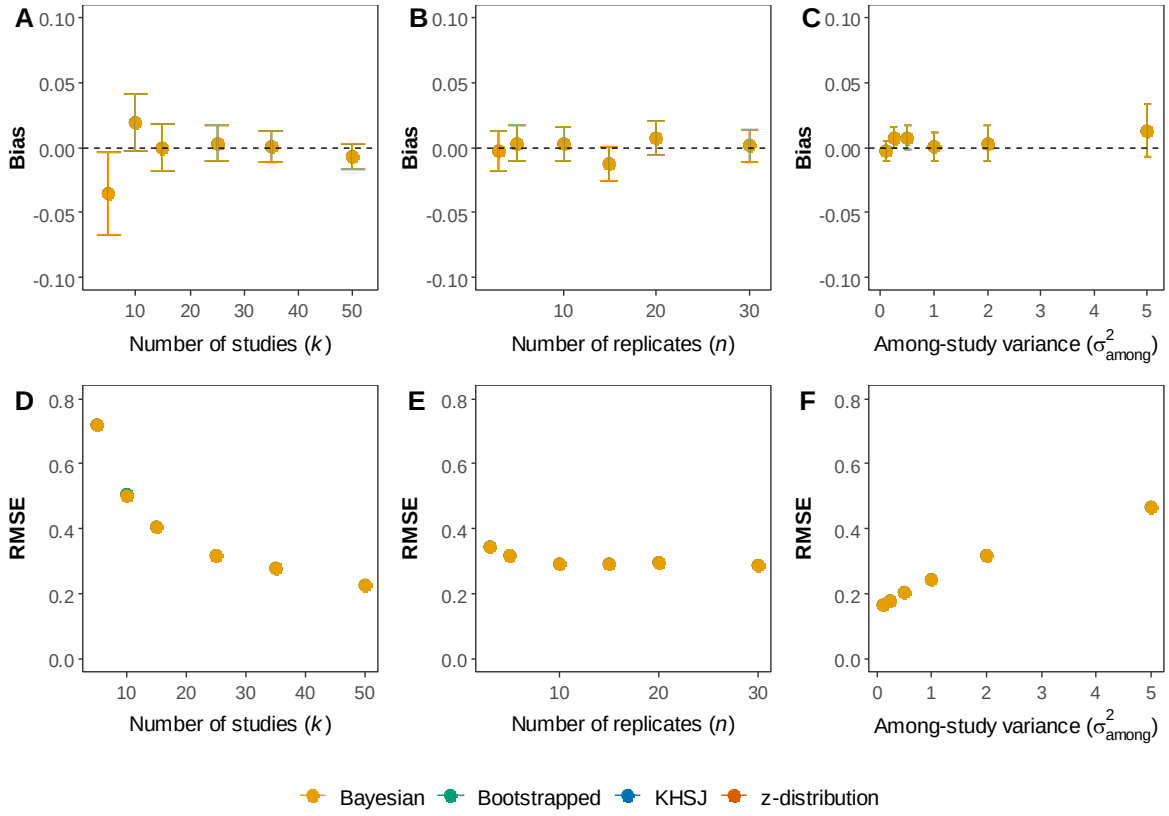


Figure S11. Bias and RMSE from the estimation of a mean effect in 2000 replicated meta-analyses as a function of the number of studies (A, B), the mean number of replicates in the original studies (C, D), and the among-study variance (E, F). Simulation parameters: $n = 5, k = 25, \sigma^2_{\text{among}} = 2$, except for the cases in which that parameter was varied. Error bars provide the 95% CI for panels A-E.

Explorations on the best priors for the among-study variance

Priors definition

There are several relatively "non-informative" priors that can be used for the among-study variance in Bayesian models. Priors can be assigned for the standard distribution, the variance, or the precision. Especially when the number of studies is small, the posterior distribution of the among-study variance can be very sensitive to the choice of prior (Gelman, 2006; Lambert et al., 2005). For these reasons, we explored five different priors for the among-study variance based on suggestions from previous studies (Gelman, 2006; Lambert et al., 2005). These included:

1) Uniform distribution for the standard deviation, we tested:
   - Uniform (0, 1)
   - Uniform (0, 10)
   - Uniform (0, 100)

Below we display the model for Uniform (0, 10):

```
model {
  for (i in 1:N){
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)     # prior for overall effect
  tau.theta <- pow(sigma.theta, -2) # define precision
  sigma.theta ~ dunif (0, 10)        # prior for standard deviation
  var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
```

2) Cauchy prior on the standard deviation:

```
model{
  for (i in 1:N){ # Likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] <- mu.theta + alpha*eps[i]
    eps[i] ~ dnorm(0, tau.eps)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3) # prior for overall effect
  # Cauchy prior for the standard deviation
  alpha ~ dnorm(0, 1)
```

```
  tau.eps~ dgamma(a, b)
  a <- 1/2
  b <- pow(AA, 2)/2
  AA <- 25
  sig.eps <- 1/sqrt(tau.eps)
  sigma.theta <- abs(alpha)/sig.eps
  var.theta <- pow(sigma.theta, 2) # set variance to monitor
}
```

3) Gamma distribution prior for the precision (i.e., 1/among-study variance):

```
model {
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for the overall effect
  tau.theta ~ dgamma(0.1, 0.1)     # Gamma prior for the precision
  sigma.theta <- 1/sqrt(tau.theta) # get the standard deviation
  var.theta <- pow(sigma.theta, 2) # set variance to monitor
}
```

4) Folded-t prior for the standard deviation:

```
model{
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)  # prior for overall effect
  # folded t prior
  A <- 5
  v <- 2
  B <- 1/(A*A)
  t.theta ~ dt (0, B, v)
  sigma.theta <- abs(t.theta)
  tau.theta <- pow(sigma.theta, -2)
  # set precision for the likelihood and variance to monitor
  var.theta <- pow(sigma.theta, 2)
}
```

5) Folded-normal prior for the standard deviation (Veroniki et al., 2015):

```
model{
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)  # prior for the overall effect
  sigma.norm ~ dnorm (0, 1)       # folded N prior for standard deviation
  sigma.theta <- abs(sigma.norm)  # fold prior
  tau.theta <- pow(sigma.theta, -2) # set precision for the likelihood
  var.theta <- pow(sigma.theta, 2)  #  set variance to monitor
}
```

Performance of the model with different priors

To analyze the prior's influence on the posterior results, we chose a small (but representative) combination of parameters for the mean number of replicates (n), the number of studies (k), and the among-study variance (σ^2_{among}). We analyzed:

- mean number of replicates: 5, 20
- number of studies: 5, 25, 50
- among-study variance: 0.5, 2, 5

The $\hat{R}$ convergence statistic can be used to evaluate convergence of the MCMC chains to the posterior distribution of the monitor parameters, in this case, the among-study variance. For each prior and combination of simulation factors, we counted the number of "bad" $\hat{R}$ values ($\hat{R} \geq 1.1$) (Gelman & Rubin, 1992). The number of replicated meta-analysis in each combination ranged from 1996 to 2000.

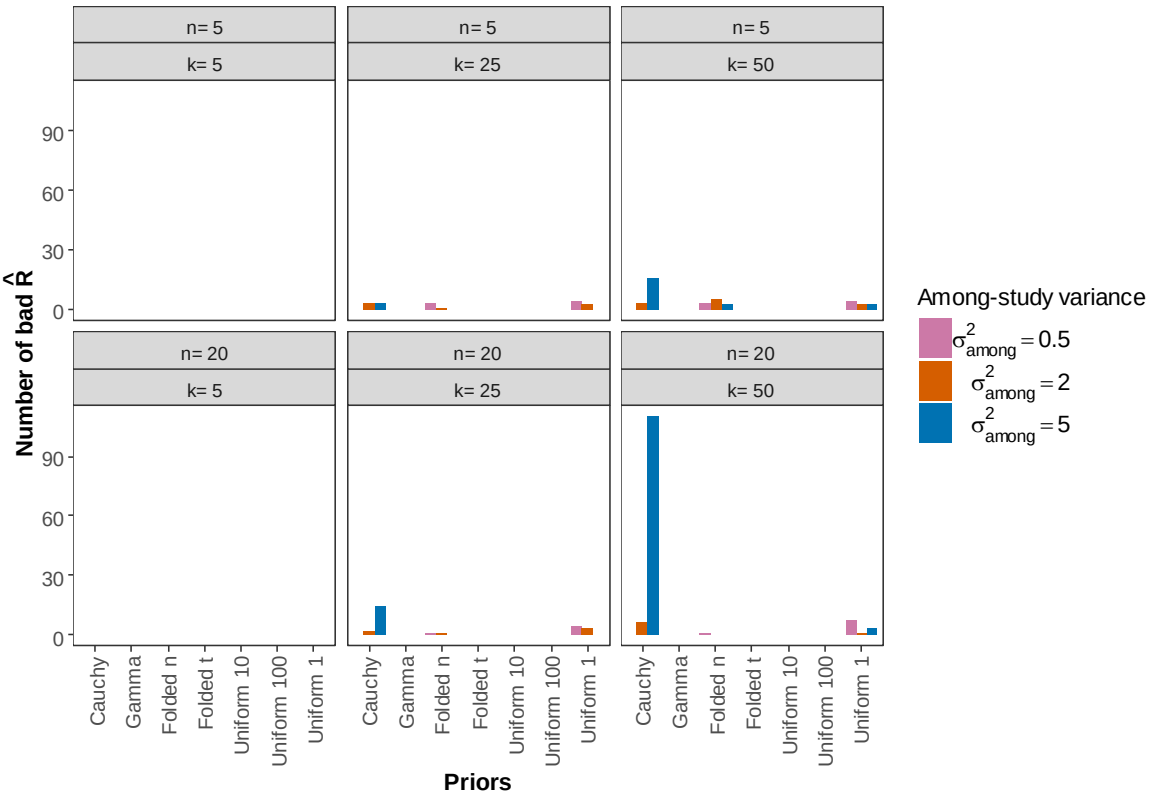


Figure S12. Number of replicates yielding bad $\hat{R}$ ($\hat{R} \geq 1.1$) for different combinations of priors, true among-study variance, mean number of replicates, and number of studies.

The folded-Cauchy, folded-normal, and Uniform(0, 1) priors had the highest number of bad $\hat{R}$ for the among-study variance (Figure S12), which indicates poor convergence to the posterior distribution. For those three priors, poor convergence was more common when the number of replicates was low, the true among-study variance was high, and the number of studies was high (Figure S12).

The folded-Cauchy and Uniform(0, 1) priors led to posterior samples that did not converge on the true (simulated) among-study variance (Figure S13). The performance of the Uniform (0, 10), Uniform (0, 100), Folded-t, and Gamma priors was similar for the combinations of simulation factors explored on our simulations (Figure S14).

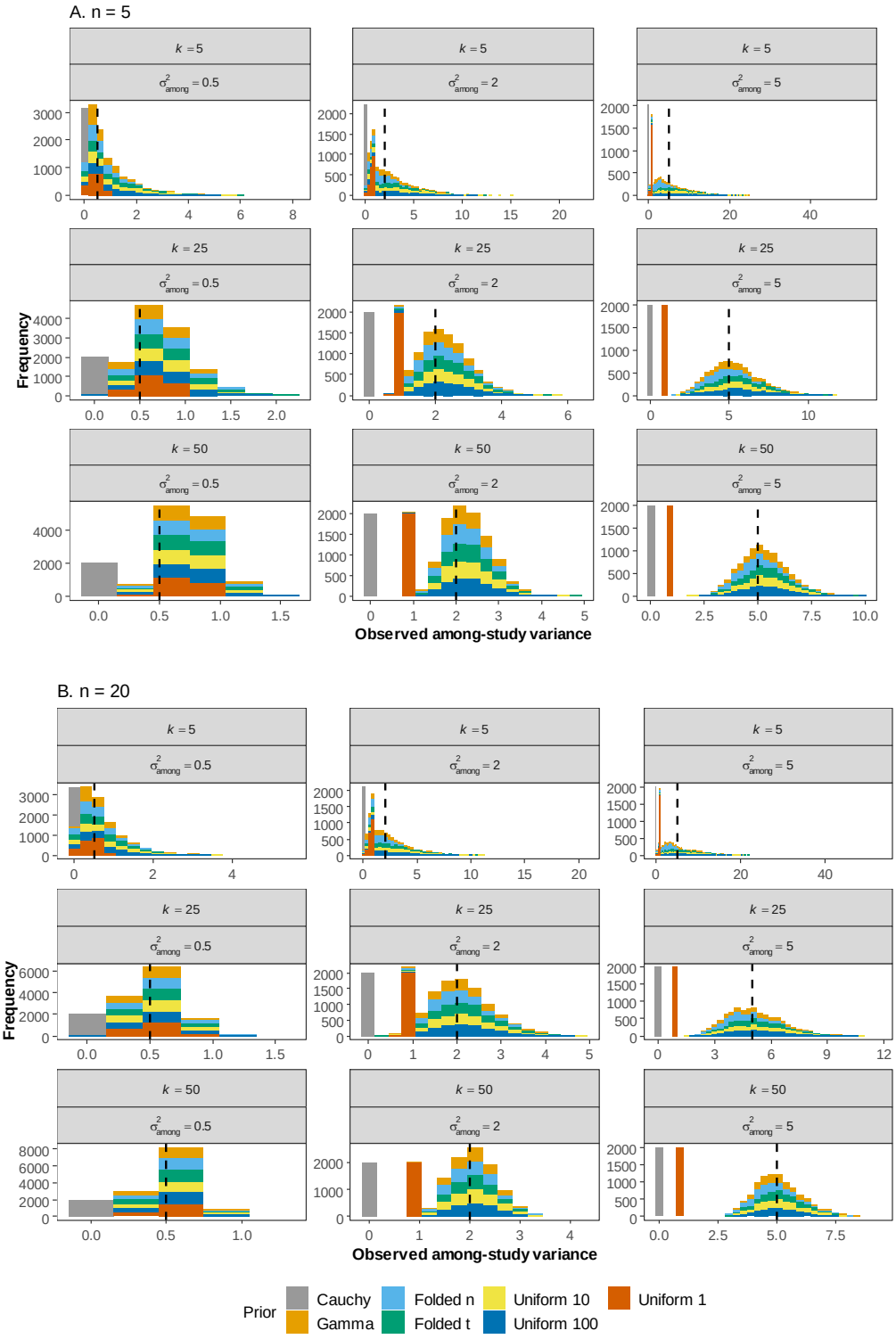


Figure S13. Median of the posterior distribution of the among-study variance for all the different priors tested, number of replicates, number of studies, and true among-study variance. A) $n = 5$; B) $n = 25$. The vertical dashed line in each panel indicates the true among-study variance.

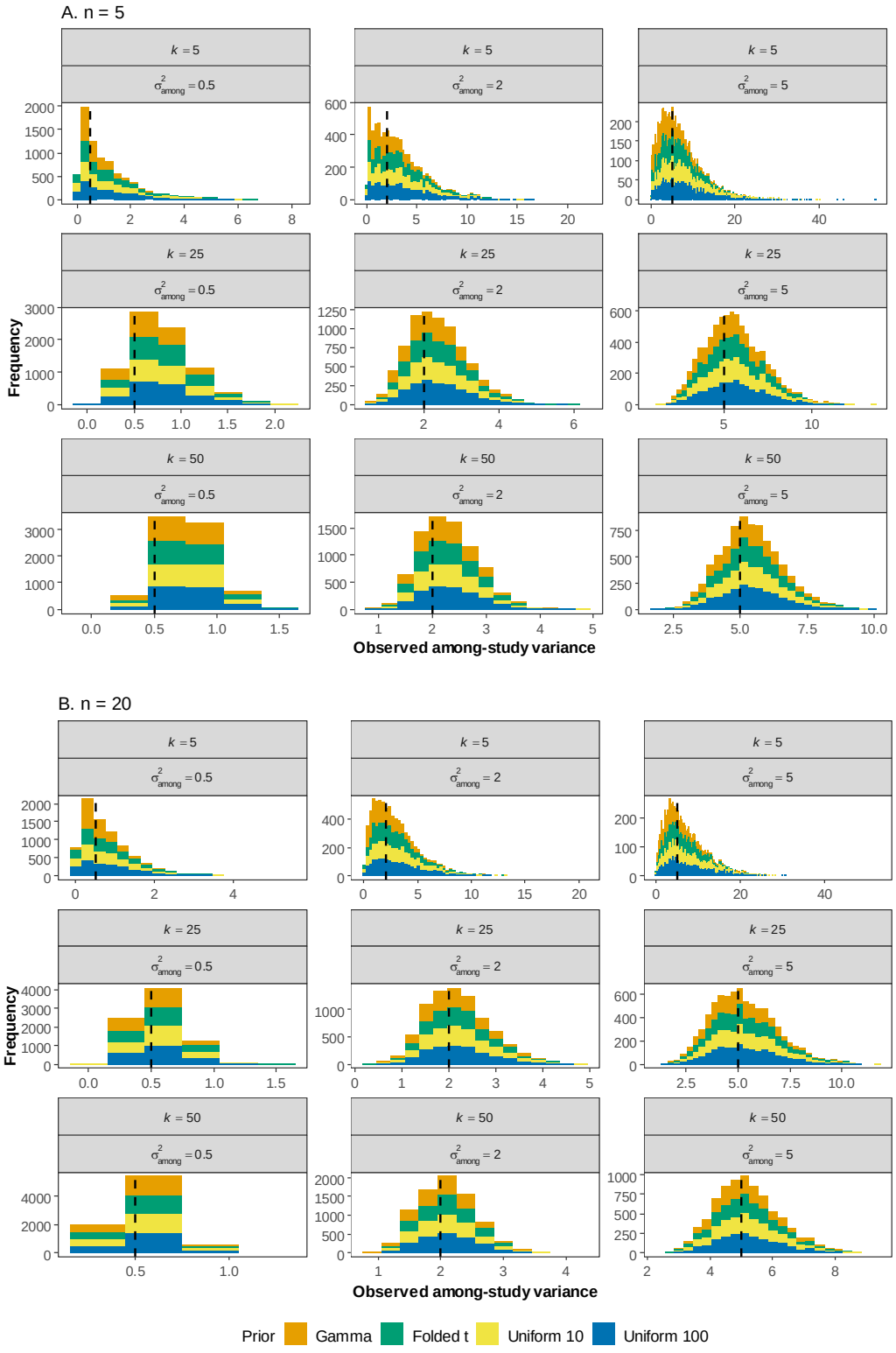
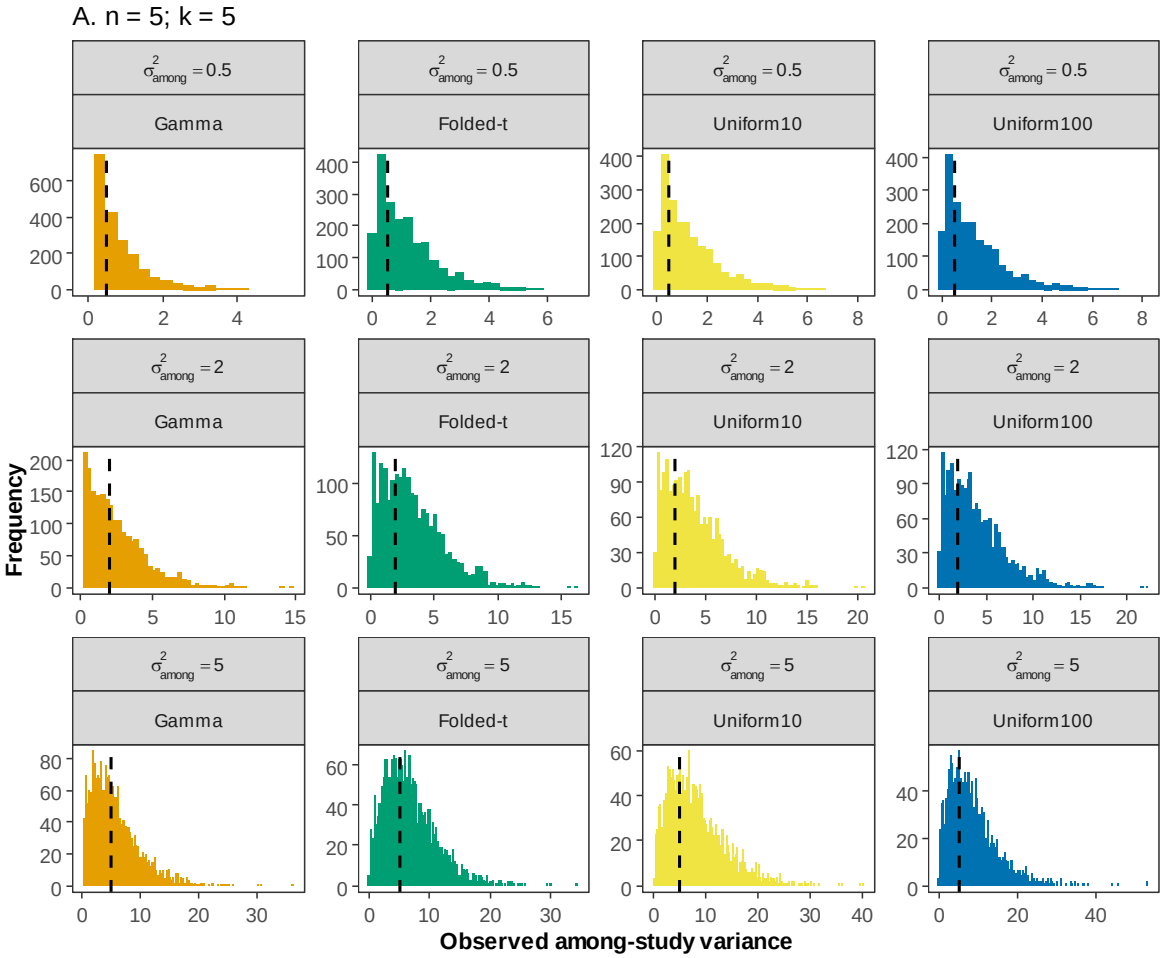


Figure S14. Median of the posterior distribution of the among-study variance for the four priors with the best performance (i.e., Uniform (0, 10), Uniform(0, 100), Gamma, Folded-t), number of replicates, number of studies, and true among-study variance. A) n = 5; B) n = 25. The vertical dashed line in each panel indicates the true among-study variance.

In a closer inspection of the best performing priors under the scenario of low number of studies we found that the folded-*t* and uniform priors converged better to the true among-study variance than the gamma prior (Figure S15). The worst performance of the gamma prior was under scenarios of low among-study variance (Figure S15). Our initial explorations (results not shown) showed that the Gamma prior did not perform well when the among-study variance was very low (i.e., σ^2_{among} = 0.001), but this case is unlikely to occur in ecological meta-analyses (Senior et al., 2016). We also tried a folded-t prior with parameter expansion to avoid the MCMC chains getting stuck at zero when σ^2_{among} = 0.001 (results not shown), but this did not perform better than the uniform prior, and the uniform prior led to faster sampling and convergence.



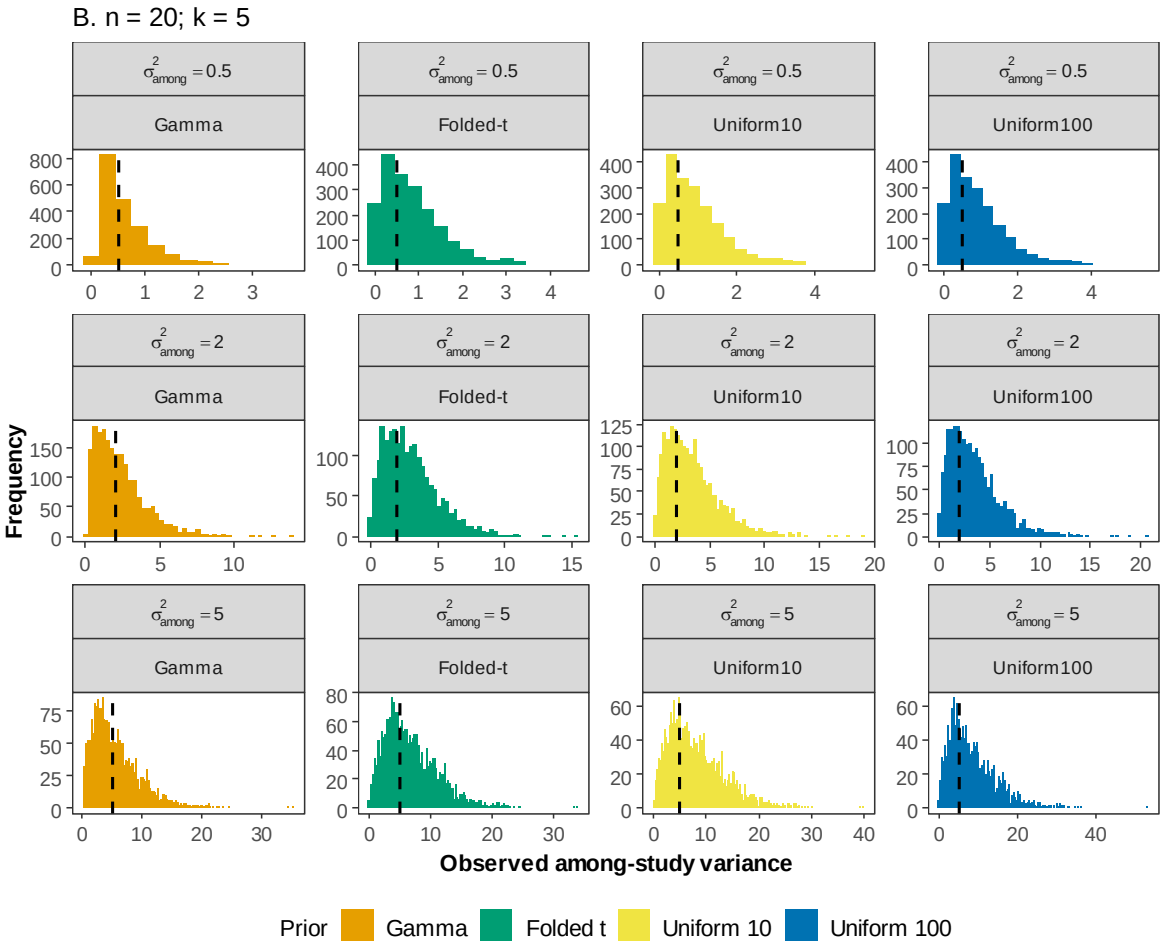


Figure S15. Median of the posterior distribution of the among-study variance for the four priors with the best performance (i.e., Uniform (0, 10), Uniform (0, 100), Gamma, Folded-t), when the number of studies was low (k = 5). A) n = 5; B) n = 25. The vertical dashed line in each panel indicates the true among-study variance.

For the combination of simulation factors used in our study, we opted for the Uniform(0, 10) prior, because it led to fast sampling and produced good convergence statistics (measured with $\hat{R}$).

References

Allan, E., Weisser, W., Weigelt, A., Roscher, C., Fischer, M., & Hillebrand, H. (2011). More diverse plant communities have higher functioning over time due to turnover

in complementary dominant species. *Proceedings of the National Academy of Sciences*, *108*(41), 17034–17039. https://doi.org/10.1073/pnas.1104015108

Anderson, J. T. (2016). Plant fitness in a rapidly changing world. *New Phytologist*, *210*(1), 81–87. https://doi.org/10.1111/nph.13693

Baig, S., Medlyn, B. E., Mercado, L. M., & Zaehle, S. (2015). Does the growth response of woody plants to elevated CO $_2$ increase with temperature? A model-oriented meta-analysis. *Global Change Biology*, *21*(12), 4303–4319. https://doi.org/10.1111/gcb.12962

Bennington, C. C., & McGraw, J. B. (1995). Natural Selection and Ecotypic Differentiation in Impatiens Pallida. *Ecological Monographs*, *65*(3), 303–324. https://doi.org/10.2307/2937062

Calfapietra, C., Gielen, B., Galema, A. N. J., Lukac, M., De Angelis, P., Moscatelli, M. C., Ceulemans, R., & Scarascia-Mugnozza, G. (2003). Free-air CO2 enrichment (FACE) enhances biomass production in a short-rotation poplar plantation. *Tree Physiology*, *23*(12), 805–814. https://doi.org/10.1093/treephys/23.12.805

Farnsworth, E., & Ellison, A. M. (1991). Patterns of herbivory in Belizean mangrove swamps. *Biotropica*, *23*(4b), 555–567.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–533.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gornish, E. S., Aanderud, Z. T., Sheley, R. L., Rinella, M. J., Svejcar, T., Englund, S. D., & James, J. J. (2015). Altered snowfall and soil disturbance influence the early life stage transitions and recruitment of a native and invasive grass in a cold desert. *Oecologia*, *177*(2), 595–606. https://doi.org/10.1007/s00442-014-3180-7

Hamner, B., & Frasco, M. (2018). Metrics: Evaluation Metrics for Machine Learning. R package version 0.1.4. https://CRAN.R-project.org/package=Metrics

Hollander, J., & Bourdeau, P. E. (2016). Evidence of weaker phenotypic plasticity by prey to novel cues from non-native predators. *Ecology and Evolution*, *6*(15), 5358–5365. https://doi.org/10.1002/ece3.2271

Jackson, M. C., Loewen, C. J. G., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: A meta-analysis. *Global Change Biology*, *22*(1), 180–189. https://doi.org/10.1111/gcb.13028

Johnson, M. S., & Black, R. (2008). Adaptive responses of independent traits to the same environmental gradient in the intertidal snail Bembicium vittatum. *Heredity*, *101*(1), 83–91. https://doi.org/10.1038/hdy.2008.33

Jokinen, I. E., Salo, H. M., Markkula, E., Rikalainen, K., Arts, M. T., & Browman, H. I. (2011). Additive effects of enhanced ambient ultraviolet B radiation and increased temperature on immune function, growth and physiological condition of juvenile (parr) Atlantic Salmon, Salmo salar. *Fish & Shellfish Immunology*, *30*(1), 102–108. https://doi.org/10.1016/j.fsi.2010.09.017

Koptur, S. (1984). Experimental evidence for defense of Inga (Mimosoideae) saplings by ants. *Ecology*, *65*(6), 1787–1793.

Lacerda, L. D. de, Jose, D. V., Rezende, C. E. de, Francisco, M. C. F., Wasserman, J. C., & Martins, J. C. (1986). Leaf chemical characteristics affecting herbivory in a new world mangrove forest. *Biotropica*, *18*(4), 350–355.

Lajeunesse, M.J. (2016). Facilitating systematic reviews, data extraction, and meta-analysis with the metagear package for R. *Methods in Ecology and Evolution*, *7*, 323-330.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., & Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, *24*(15), 2401–2428. https://doi.org/10.1002/sim.2112

Lavola, A., Nybakken, L., Rousi, M., Pusenius, J., Petrelius, M., Kellomäki, S., & Julkunen-Tiitto, R. (2013). Combination treatment of elevated UVB radiation, CO_2 and temperature has little effect on silver birch (*Betula pendula*) growth and phytochemistry. *Physiologia Plantarum*, *149*(4), 499–514. https://doi.org/10.1111/ppl.12051

Lemoine, N. P., Hoffman, A., Felton, A. J., Baur, L., Chaves, F., Gray, J., Yu, Q., & Smith, M. D. (2016). Underappreciated problems of low replication in ecological field studies. *Ecology*, *97*(10), 2554–2561. https://doi.org/10.1002/ecy.1506

Lent, J. van, Hergoualc'h, K., & Verchot, L. V. (2015). Reviews and syntheses: Soil N2O and NO emissions from land use and land-use change in the tropics and subtropics: A meta-analysis. *Biogeosciences*, *12*(23), 7299–7313. https://doi.org/10.5194/bg-12-7299-2015

Moher D., Liberati A., Tetzlaff J., et al., and the PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*. *151*:264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Muyssen, B. T. A., Messiaen, M., & Janssen, C. R. (2010). Combined cadmium and temperature acclimation in Daphnia magna: Physiological and sub-cellular effects. *Ecotoxicology and Environmental Safety*, *73*(5), 735–742. https://doi.org/10.1016/j.ecoenv.2009.12.018

Oduor, A. M. O., Leimu, R., & Kleunen, M. van. (2016). Invasive plant species are locally adapted just as frequently and at least as strongly as native plant species. *Journal of Ecology*, *104*(4), 957–968. https://doi.org/10.1111/1365-2745.12578

Parsons, K. E. (1997). Contrasting Patterns of Heritable Geographic Variation in Shell Morphology and Growth Potential in the Marine Gastropod Bembicium vittatum: Evidence from Field Experiments. *Evolution*, *51*(3), 784. https://doi.org/10.2307/2411154

Pfeifer-Meister, L., Bridgham, S. D., Little, C. J., Reynolds, L. L., Goklany, M. E., & Johnson, B. R. (2013). Pushing the limit: Experimental evidence of climate effects on plant range distributions. *Ecology*, *94*(10), 2131–2137. https://doi.org/10.1890/13-0284.1

Rodríguez-Castañeda, G. (2013). The world and its shades of green: A meta-analysis on trophic cascades across temperature and precipitation gradients: Trophic interactions across gradients. *Global Ecology and Biogeography*, *22*(1), 118–130. https://doi.org/10.1111/j.1466-8238.2012.00795.x

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S. A., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: Its magnitude and implications. *Ecology*, *97*(12), 3293–3299. https://doi.org/10.1002/ecy.1591

Verchot, L. V., Davidson, E. A., Cattânio, H., Ackerman, I. L., Erickson, H. E., & Keller, M. (1999). Land use change and biogeochemical controls of nitrogen oxide emissions from soils in eastern Amazonia. *Global Biogeochemical Cycles*, *13*(1), 31–46. https://doi.org/10.1029/1998GB900019

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, *26*(1), 37–52. https://doi.org/10.1002/sim.2514

Wasey, J.O. (2019). PRISMAstatement: Plot Flow Charts According to the "PRISMA" Statement. R package version 1.1.1. https://CRAN.R-project.org/package=PRISMAstatement

Way, D. A., & Sage, R. F. (2008). Elevated growth temperatures reduce the carbon gain of black spruce [Picea mariana (Mill.) B.S.P.]. *Global Change Biology*, *14*(3), 624–636. https://doi.org/10.1111/j.1365-2486.2007.01513.x

Werner, C., Zheng, X., Tang, J., Xie, B., Liu, C., Kiese, R., & Butterbach-Bahl, K. (2006). N2O, CH4 and CO2 emissions from seasonal tropical rainforests and a rubber plantation in Southwest China. *Plant and Soil*, *289*(1-2), 335–353. https://doi.org/10.1007/s11104-006-9143-y

Zhang, Y., Wang, Y., & Yu, S. (2014). Interspecific Neighbor Interactions Promote the Positive Diversity-Productivity Relationship in Experimental Grassland Communities. *PLoS ONE*, *9*(10), e111434. https://doi.org/10.1371/journal.pone.0111434

R Code used in “Comparing traditional and Bayesian approaches to ecological meta-analysis”

Journal: Methods in Ecology and Evolution

Authors: P. Pappalardo, K. Ogle, E.A. Hamman, J.R. Bence, B.A. Hungate, & C.W. Osenberg

Code availability

This R code is available as an rmarkdown file in the Dryad Data Repository associated with this publication: https://doi.org/10.5061/dryad.zw3r22863. The R functions presented here were compiled into an R file also available in the Dryad repository. We present here a text version for the readers that want to look at the code without downloading the data package.

Literature search in the Web of Science

Our search in the ISI Web of Science database in March 2017 returned 581 hits. To export the data, we used two tab delimited files (for Windows), since the maximum amount of data is 500 citations each time. The tab delimited file from WOS was tricky to read from R, so we opened both files in Excel and saved them as .csv files for follow up analysis.

Compiling citation data in R

First, we compiled WOS information from the two files and cleaned weird entries.

```
library(dplyr)

# load files and keep only columns of interest

wos1 <- read.csv("1_500.csv", header = T, as.is = T)
wos1 <- wos1[, c("PT", "AU", "TI", "SO", "VL", "IS", "BP",
    "EP", "DI", "PD", "PY", "AB", "TC", "Z9")]

wos2 <- read.csv("501_581.csv", header = T, as.is = T)
wos2 <- wos2[, c("PT", "AU", "TI", "SO", "VL", "IS", "BP",
    "EP", "DI", "PD", "PY", "AB", "TC", "Z9")]
```

```
# combine files

wos <- rbind(wos1, wos2)

# rename columns

names(wos) <- c("pub.type", "authors", "title", "pub.name",
    "vol", "issue", "b.page", "e.page", "doi", "pub.date",
    "pub.year", "abstract", "wos.citations", "alldb.citations")

# standardize journal names to uppercase

woscaps <- mutate_each(wos, funs(toupper))
wos$pub.name <- woscaps$pub.name

# identify weird journal names

name <- which(wos$pub.name == "ANNUAL REVIEW OF PHYTOPATHOLOGY, VOL 52")
name1 <- which(wos$pub.name == "ADVANCES IN ECOLOGICAL RESEARCH, VOL 50:
ECO-EVOLUTIONARY DYNAMICS")
name2 <- which(wos$pub.name == "ADVANCES IN ECOLOGICAL RESEARCH, VOL 46:
GLOBAL CHANGE IN MULTISPECIES SYSTEMS, PT 1")

# change to correct name

wos$pub.name[name] <- "ANNUAL REVIEW OF PHYTOPATHOLOGY"
wos$pub.name[name1] <- "ADVANCES IN ECOLOGICAL RESEARCH"
wos$pub.name[name2] <- "ADVANCES IN ECOLOGICAL RESEARCH"

# remove objects not in use

rm(wos1, wos2)

# write file with all citations

write.csv(wos, "WebOfScience_1980-2016.csv", row.names = F)
```

Abstract screening using *metagear*

After compiling all the citations, we screened abstracts using the viewer classification tool in the R package *metagear* (Lajeunesse, 2016). This package requires a file with the references, the search has to be independent, and ideally it will contain authors, title, doi, abstract. To begin, the function effort_initialize()adds the columns “study_id”, “reviewers” and “include”. The abstract_screener() function creates an interactive window that allows you to read the abstract and make a decision about including the paper (“yes”, “maybe”, “no”). The package adds a new column in the citation data with

this decision. It can also split the effort among different coauthors; if so, it is necessary to include a column with the reviewer/s name and assign reviewer's effort (see code below).

```
library(metagear)

# load file

wos <- read.csv("WebOfScience_1980-2016.csv", header = T)

# prepare the file for the screening effort

wos.scan <- effort_initialize(wos)

# save file with the IDs as a backup

write.csv(wos.scan, "wosWithIds.csv")

# randomly distribute screening effort to a team

theTeam <- c("Paula")

theRefs_unscreened <- effort_distribute(wos.scan, reviewers = theTeam,
    effort = c(100), save_split = T)

# start the abstract viewer to do first pass

abstract_screener("effort_Paula.csv", aReviewer = "Paula",
    abstractColumnName = "abstract", titleColumnName = "title")

# get the summary of your work

theRefs_screened <- effort_merge()
sum.scan <- effort_summary(theRefs_screened)
```

=== SCREENING EFFORT SUMMARY ===

319 candidate studies identified 234 studies excluded 28 challenging studies needing additional screening

581 TOTAL SCREENED

=== SCREENING DESIGN SUMMARY ===

```
  MAYBE  NO YES TOTAL   %
```

Paula 28 234 319 581 100 TOTAL 28 234 319 581 100

Get the pdfs for next screening stage

We used the PDFs_collect() function to try to get all the PDFs coded as “yes” and “maybe” to do the next step of the literature review. When we were unable to find the PDF using this tool we searched directly on the web using google scholar, or tried to obtain the pdf through our university library or by emailing the corresponding author.

```
# load revision file

refs <- read.csv("effort_Paula.csv")

# subset by inclusion decision

theRefs.included <- refs[which(refs$INCLUDE == "YES" | refs$INCLUDE ==
    "MAYBE"), ]

# try to get the pdf

PDFs_collect(theRefs.included, DOIcolumn = "doi", FileNamecolumn = "STUDY
_ID",
    directory = getwd())
```

As part of the Dryad repository we included the original .txt files from Web of Science. The final list of abstract screened and the information for each reference is provided in the main data file “Pappalardo_etal_LiteratureReview_Dataset.xlsx”.

Functions used in the simulation experiment and data analysis

The functions displayed here are available in the file “Functions_Pappalardo_etal.R” in the Dryad data repository https://doi.org/10.5061/dryad.zw3r22863.

All functions were written by Paula Pappalardo, with assistance from Kiona Ogle, Elizabeth Hamman, and Jim Bence. The functions to run a meta-analysis using bootstraping in metafor were edited from code provided in the metafor website (http://www.metafor-project.org/doku.php/tips:bootstrapping_with_ma).

Functions to simulate datasets:

```
makeLognormalRow <- function(mu, tau, sigma, mean.n, c.mean) {
    # Simulates a row of lognormal data representing study i

    # Args:

    # mu: numeric value for the overall effect size
    # tau: numeric value for the among-study variance
```

```r
    # sigma: numeric value for the among-replicates variation
    # mean.n: numeric value for number of replicates
    # c.mean: numeric value for the control mean

    # Returns:

    # Dataframe with a row of simulated data for a study in
    # the meta-analysis

    # Simulate number of replicates using a Poisson
    # distribution (rescale by 2 to avoid 0 and 1 replicates)
    pn <- rpois(1, lambda = mean.n - 2)
    n <- pn + 2

    # calculate a random error based on tau
    eta <- rnorm(1, mean = 0, sd = sqrt(tau))

    # simulate raw data for control and treatment
    c <- rlnorm(n, meanlog = log(c.mean) - (sigma^2)/2,
        sdlog = sigma)
    t <- rlnorm(n, meanlog = log(c.mean) - (sigma^2)/2 +
        mu + eta, sdlog = sigma)

    # calculate log ratio (lnrr) and its variance (var.lnrr)
    lnrr <- log(mean(t)/mean(c))
    var.lnrr <- sd(t)^2/(n * mean(t)^2) + sd(c)^2/(n * mean(c)^2)

    # calculate standard error if running a true Bayesian
    se.t <- sd(t)/sqrt(n)
    se.c <- sd(c)/sqrt(n)

    # put results in dataframe
    results <- data.frame(lnrr = lnrr, var.lnrr = var.lnrr,
        yt = mean(t), yc = mean(c), nt = n, nc = n, sd.t = sd(t),
        sd.c = sd(c), se.t = se.t, se.c = se.c)
    return(results)
}

makeLognormalDataset <- function(mu, tau, sigma, mean.n,
    c.mean, k) {
    # Create datasets for simulations

    # Args:

    # mu: numeric value for the overall effect size
    # tau: numeric value for the among-study variance
    # sigma: numeric value for the among-replicates variation
    # mean.n: numeric value for number of replicates
    # c.mean: numeric value for the control mean
```

```r
# k: number of studies

# Returns:

# list of meta-analytic datasets for all factor
# combinations

datasets <- list()  # empty list to hold results

# Loop through all combinations of parameters
for (m in mu) {
    for (i in sigma) {
        for (j in tau) {
            for (n in mean.n) {
              for (z in k) {
                # set a name to put things in a list
                name <- paste("sigma", i, "tau", j,
                  "mean.n", n, "k", z, "mu", m, "c.mean",
                  c.mean, sep = " ")

                # make empty list to put each row
                rows.list <- list()

                # creating a dataset with k= z
                for (x in 1:z) {
                  namek <- paste("k number", x)

                  # get dataset for this combination
                  datarow <- makeLognormalRow(sigma = i,
                    tau = j, mu = m, mean.n = n, c.mean = c.mean)

                  # put row in list
                  rows.list[[namek]] <- datarow
                }
                df.results <- do.call(rbind, rows.list)
                df.results$true.mu <- m
                df.results$true.tau2 <- j
                df.results$mean.n <- n
                df.results$k <- z
                df.results$sigma <- i
                df.results$c.mean <- c.mean
                row.names(df.results) <- NULL
                datasets[[name]] <- df.results
              }
            }
        }
    }
}
# return list with datasets
```

```
    return(datasets)
}
```

Functions to run meta-analysis with traditional methods:

```
runMetafor <- function(mydf, tau.method, ci) {
    # Run a weighted random-effects meta-analysis with metafor

    # Args:

    # tau.method: defines the method to estimate the among-study variance
    # ci: T to run Knapp-Hartung CI, F for default z-distribution CI

    # Returns:

    # Summary results from the meta-analysis

    library(metafor)

    # run meta-analysis
    res <- try(rma(lnrr, var.lnrr, data = mydf, knha = ci,
        method = tau.method, control = list(maxiter = 200)))

    # check if metafor gave an error
    error <- class(res)[1]

    # prepare dataframe in case of error
    if (error == "try-error") {

        # make dataframe to hold the data if there was an error
        res2 <- data.frame(error = 1)

        # return dataframe
        return(res2)

    } else {
        # get confidence intervals for tau
        res.tau <- confint(res)

        # extract CI's for tau
        tau.ci.lb <- res.tau$random["tau^2", "ci.lb"]
        tau.ci.ub <- res.tau$random["tau^2", "ci.ub"]

        # make dataframe in case there is no error
        res2 <- data.frame(obs.mu = res$b, mu.ci.lb = res$ci.lb,
            mu.ci.ub = res$ci.ub, se = res$se, obs.tau = res$tau2,
```

```
            tau.ci.lb, tau.ci.ub, true.mu = mydf$true.mu[1],
            sigma = mydf$sigma[1], true.tau = mydf$true.tau2[1],
            k = mydf$k[1], mean.n = mydf$mean.n[1], error = 0,
            i2 = res$I2)

        # calculate bias for overall effect and tau
        res2$bias.eff <- res2$obs.mu - res2$true.mu
        res2$bias.tau <- res2$obs.tau - res2$true.tau

        # calculate coverage by comparing with simulated values
        coverage.mu <- ifelse(res2$mu.ci.lb <= res2$true.mu &
            res2$mu.ci.ub >= res2$true.mu, 1, 0)
        coverage.tau <- ifelse(res2$tau.ci.lb <= res2$true.tau &
            res2$tau.ci.ub >= res2$true.tau, 1, 0)

        # add coverage to the final dataframe
        res2$cov.mu <- coverage.mu
        res2$cov.tau <- coverage.tau

        # calculate the width of the CI
        res2$mu.width <- res2$mu.ci.ub - res2$mu.ci.lb
        res2$tau.width <- res2$tau.ci.ub - res2$tau.ci.lb

        # return results
        print("meta-analysis with metafor done")
        return(res2)
    }
}

runMetafor_bootfunction <- function(mydf, indices) {
    # Function to bootstrap a meta-analysis with metafor

    # Args:

    # mydf: dataset to analyze
    # indices: vector of indices which define the bootstrap sample

    # Returns:

    # Summary results from the meta-analysis

    library(metafor)

    res <- try(rma(lnrr, var.lnrr, data = mydf, method = "REML",
        subset = indices), silent = TRUE)
    if (is.element("try-error", class(res))) {
        c(NA)
    } else {
        c(coef(res), res$tau2)
```

```
    }
}

bootMetafor <- function(mydf) {
    # Bootstrap the mean effect and associated CI from a meta-analysis

    # Args:

    # mydf: dataset to analyze

    # Returns: Summary results from the meta-analysis

    library(boot)
    library(metafor)

    # function for bootstrapping CI's
    res.boot <- try(boot(mydf, runMetafor_bootfunction,
        R = 5000))

    # use try in case function fails
    error.1 <- class(res.boot)
    if (error.1 == "try-error") {
        res.error <- data.frame(obs.mu = NA, mu.ci.lb = NA,
            mu.ci.ub = NA)
        return(res.error)

    } else {
        # run the boot.ci function to get non-parametric bootstrap CI's
        res.muci <- try(boot.ci(res.boot, index = 1))
        error.2 <- class(res.muci)

        if (error.2 == "try-error") {
            res.error <- data.frame(obs.mu = NA, mu.ci.lb = NA,
                mu.ci.ub = NA)
            return(res.error)

        } else {
            # select the bias corrected non parametric CIs
            mu.ci.lb <- res.muci$bca[4]
            mu.ci.ub <- res.muci$bca[5]
            obs.mu <- res.muci$t0[1]

            # make dataframe to hold answers in case there is no
            # error
            res1 <- data.frame(obs.mu, mu.ci.lb, mu.ci.ub,
                true.mu = mydf$true.mu[1], sigma = mydf$sigma[1],
                true.tau = mydf$true.tau2[1], k = mydf$k[1],
                mean.n = mydf$mean.n[1], error = 0, i2 = res.boot$t0[5])
```

```r
            # calculate bias for overall effect and tau
            res1$bias.eff <- res1$obs.mu - res1$true.mu

            # calculate coverage by comparing the metafor CIs with
            # the real answer

            coverage.mu <- ifelse(res1$mu.ci.lb <= res1$true.mu &
                res1$mu.ci.ub >= res1$true.mu, 1, 0)

            # add coverage to final dataframe
            res1$cov.mu <- coverage.mu

            # calculate the width of the CI
            res1$mu.width <- res1$mu.ci.ub - res1$mu.ci.lb

            print("weighted meta-analysis boostrap CI with Metafor")
            return(res1)
        }
    }
}

bootMetaforTau <- function(mydf) {
    # Bootstrap the among-study variance and associated CI

    # Args:

    # mydf: dataset to analyze

    # Returns: Summary results from the meta-analysis

    # function for bootstrapping CI's
    res.boot <- try(boot(mydf, runMetafor_bootfunction,
        R = 2000))

    # use try in case function fails
    error.1 <- class(res.boot)
    if (error.1 == "try-error") {
        res.error <- data.frame(obs.tau = NA, tau.ci.lb = NA,
            tau.ci.ub = NA)
        return(res.error)
    } else {

        # run the boot.ci function to get non-parametric bootstrap CI's
        res.tauci <- try(boot.ci(res.boot, index = 2))
        res.muci <- try(boot.ci(res.boot, index = 1))
        error.2 <- class(res.tauci)
        error.3 <- class(res.muci)

        if (error.2 == "try-error" | error.3 == "try-error") {
```

```
            res.error <- data.frame(obs.tau = NA, tau.ci.lb = NA,
                tau.ci.ub = NA)
            return(res.error)

        } else {
            tau.ci.lb <- res.tauci$bca[4]
            tau.ci.ub <- res.tauci$bca[5]
            obs.tau <- res.tauci$t0[1]

            mu.ci.lb <- res.muci$bca[4]
            mu.ci.ub <- res.muci$bca[5]
            obs.mu <- res.muci$t0[1]
        }

        # make dataframe to hold answers in case there is no
        # error
        res1 <- data.frame(obs.mu, mu.ci.lb, mu.ci.ub, obs.tau,
            tau.ci.lb, tau.ci.ub, true.mu = mydf$true.mu[1],
            sigma = mydf$sigma[1], true.tau = mydf$true.tau2[1],
            k = mydf$k[1], mean.n = mydf$mean.n[1], error = 0,
            i2 = res.boot$t0[5])

        # calculate bias for overall effect and tau
        res1$bias.eff <- res1$obs.mu - res1$true.mu
        res1$bias.tau <- res1$obs.tau - res1$true.tau

        # calculate coverage by comparing the metafor CIs with
        # the real answer
        coverage.mu <- ifelse(res1$mu.ci.lb <= res1$true.mu &
            res1$mu.ci.ub >= res1$true.mu, 1, 0)
        coverage.tau <- ifelse(res1$tau.ci.lb <= res1$true.tau &
            res1$tau.ci.ub >= res1$true.tau, 1, 0)

        # add coverage to final dataframe
        res1$cov.mu <- coverage.mu
        res1$cov.tau <- coverage.tau

        # calculate the width of the CI
        res1$mu.width <- res1$mu.ci.ub - res1$mu.ci.lb
        res1$tau.width <- res1$tau.ci.ub - res1$tau.ci.lb

        print("weighted meta-analysis boostrap CI for tau2 with Metafor")
        return(res1)
    }
}
```

Functions to calculate confidence intervals for data analysis:

```r
# load libraries we need

library(binom)

# Confidence intervals for coverage (binomial variable)

bi.95.l <- function(vec1) {
    vec <- vec1[!is.na(vec1)]
    s <- length(which(vec == 1))
    runs <- length(vec)
    res <- binom.confint(x = s, n = runs, conf.level = 0.95,
        method = c("wilson"))
    res.low <- res$mean - res$lower
    return(res.low)
}

bi.95.u <- function(vec1) {
    vec <- vec1[!is.na(vec1)]
    s <- length(which(vec == 1))
    runs <- length(vec)
    res <- binom.confint(x = s, n = runs, conf.level = 0.95,
        method = c("wilson"))
    res.up <- res$upper - res$mean
    return(res.up)
}

# Confidence intervals for bias

t.95CI <- function(vec) {
    x <- vec[!is.na(vec)]
    n <- length(x)
    t <- qt(0.975, df = n - 1)
    se <- sd(x)/sqrt(n)
    ci <- t * se
    return(ci)
}

z.95CI <- function(vec) {
    x <- vec[!is.na(vec)]
    n <- length(x)
    se <- sd(x)/sqrt(n)
    z <- qnorm(0.975) * se
    ci <- z * se
    return(ci)
}
```

Functions to run Hybrid Bayesian meta-analysis:

```r
# load libraries we need

library(rjags)
library(runjags)
library(coda)

make_JAGS_summary <- function(sims, sigma, true.mu, true.tau,
    k, mean.n) {
    # Summarize results from the JAGS function

    # Args:

    # sims: output of run.jags() function
    # true.tau: numeric value for the true among-study variance
    # true.mu: numeric value for the true overall effect
    # sigma: numeric value for the among-replicates variation
    # mean.n: numeric value for mean number of replicates
    # k: numeric value for the number of setudies

    # Returns:

    # Dataframe with the meta-analysis results

    # Get summary of the results

    summ <- sims$summaries
    summ2 <- sims$summary
    obs.mean.mu <- summ["mu.theta", "Mean"]
    obs.mean.tau <- summ["var.theta", "Mean"]
    obs.median.mu <- summ["mu.theta", "Median"]
    obs.median.tau <- summ["var.theta", "Median"]
    obs.sd.mu <- summ["mu.theta", "SD"]
    obs.sd.tau <- summ["var.theta", "SD"]
    n.burnin <- sims$burnin
    n.iter <- sims$sample
    n.chains <- summ2[["nchain"]]

    # Asses convergence of chains (it needs to be lower than 1.1)

    Rhat.mu <- summ["mu.theta", "psrf"]
    Rhat.tau2 <- summ["var.theta", "psrf"]

    # get samples
    mcmcsamp <- sims$mcmc
    mcmcmat <- do.call("rbind", mcmcsamp)
```

```r
    # Convert simulations to MCMC object
    mcmc.sims <- as.mcmc(mcmcmat)

    # Calculate 95% high density intervals
    hdi.sims <- HPDinterval(mcmc.sims, prob = 0.95)

    # HDI for the overall effect
    obs.mu.hdiL <- hdi.sims["mu.theta", "lower"]
    obs.mu.hdiU <- hdi.sims["mu.theta", "upper"]

    # width of the high density interval for the overall effect
    mu.width <- obs.mu.hdiU - obs.mu.hdiL

    # HDI for the among studies variance
    obs.tau.hdiL <- hdi.sims["var.theta", "lower"]
    obs.tau.hdiU <- hdi.sims["var.theta", "upper"]

    # width of the high density interval for the among-study variance
    tau.width <- obs.tau.hdiU - obs.tau.hdiL

    # calculate coverage for the overall effect and tau
    cov.mu <- ifelse(obs.mu.hdiL <= true.mu & obs.mu.hdiU >=
        true.mu, 1, 0)
    cov.tau <- ifelse(obs.tau.hdiL <= true.tau & obs.tau.hdiU >=
        true.tau, 1, 0)

    # add bias estimation
    bias.mean.eff <- obs.mean.mu - true.mu
    bias.mean.tau <- obs.mean.tau - true.tau
    bias.median.eff <- obs.median.mu - true.mu
    bias.median.tau <- obs.median.tau - true.tau

    # create dataframe with all the output we want
    bayes.results <- data.frame(true.mu, sigma, true.tau,
        k, mean.n, obs.mean.mu, obs.median.mu, obs.sd.mu,
        obs.mu.hdiL, obs.mu.hdiU, obs.mean.tau, obs.median.tau,
        obs.tau.hdiL, obs.tau.hdiU, obs.sd.tau, bias.mean.eff,
        bias.mean.tau, bias.median.eff, bias.median.tau,
        n.burnin, n.iter, n.chains, Rhat.mu, Rhat.tau2,
        cov.mu, cov.tau, mu.width, tau.width)

    # return data
    return(bayes.results)
}

#-----------------Hybrid Bayesian model with uniform prior --------------
modelString = "
model{
```

```
for (i in 1:N){
y[i] ~ dnorm (theta[i], tau.y[i])
theta[i] ~ dnorm (mu.theta, tau.theta)
}
# Setting priors
mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for Overall effect
tau.theta <- pow(sigma.theta, -2)     # define precision
sigma.theta ~ dunif (0, 10)              # prior for standard deviation
var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
"
# write model file:
writeLines(modelString, con = "modelHalfBayes_uniform_JAGS.txt")

runHalfBayes_uniform_JAGS <- function(mydf) {
    # runs Hybrid Bayesian meta-analysis with uniform prior

    # Args:
    # mydf: dataset with columns eff.size and var

    # Returns: One line dataframe with the meta-analysis results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)

    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain2 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain3 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = "modelHalfBayes_uniform_JAGS.txt", n.chains = 3,
        burnin = 1e+05, sample = 1e+05, monitor = c("mu.theta",
            "var.theta"), method = "rjags")
```

```r
    # Make dataframe with the results of the meta-analysis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with uniform prior done with JAGS"
)
    return(results)

    # remove things not in use
    rm(mydf, N, y, tau.y, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)
}


# ----------------------ALL HYBRID Bayesian MODELS----------------------

# ------------------Folded N

modelString = "
model{
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)  # prior for overall effect
  sigma.norm ~ dnorm (0, 1) # folded N prior for standard deviation
  sigma.theta <- abs(sigma.norm) # here I folded it
  tau.theta <- pow(sigma.theta, -2) # set precision for the likelihood
  var.theta <- pow(sigma.theta, 2) #  set variance to monitor
}
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_folded_N.txt")

# Function to run half bayesian analysis with folded N prior


runHalfBayes_folded_N <- function(mydf) {
    # runs Hybrid Bayesian meta-analysis with folded N prior

    # Args:

    # mydf: dataset with columns eff.size and var

    # Returns: One line dataframe with the meta-analysis results
```

```r
    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)

    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(theta = rnorm(N, muinit, sdinit), mu.theta = rnorm(1,
        muinit, sdinit), sigma.norm = rnorm(1, 0, sdinit))
    chain2 <- list(theta = rnorm(N, muinit, sdinit), mu.theta = rnorm(1,
        muinit, sdinit), sigma.norm = rnorm(1, 0, sdinit))
    chain3 <- list(theta = rnorm(N, muinit, sdinit), mu.theta = rnorm(1,
        muinit, sdinit), sigma.norm = rnorm(1, 0, sdinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = "modelHalfBayes_folded_N.txt", n.chains = 3,
        burnin = 1e+05, sample = 1e+05, monitor = c("mu.theta",
            "var.theta"), method = "rjags")

    # Make dataframe with the results of the meta-analysis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with folded N prior done with JAGS
")
    return(results)

    # remove things not in use
    rm(mydf, N, y, tau.y, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)
}

# -------------------Cauchy

modelString = "
model{
  for (i in 1:N){ # Likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
```

```
    theta[i] <- mu.theta + alpha*eps[i]
    eps[i] ~ dnorm(0, tau.eps)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for Overall effect
  # Cauchy prior for the standard deviation
  alpha ~ dnorm(0, 1)
  tau.eps~ dgamma(a, b)
  a <- 1/2
  b <- pow(AA, 2)/2
  AA <- 25
  sig.eps <- 1/sqrt(tau.eps)
  sigma.theta <- abs(alpha)/sig.eps
  var.theta <- pow(sigma.theta, 2) # set variance to monitor
}
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_Cauchy.txt")

# Function to run half bayesian analysis with uniform prior


runHalfBayes_cauchy <- function(mydf) {
    # runs Hybrid Bayesian meta-analysis with Cauchy prior

    # Args:

    # mydf: dataset with columns eff.size and var

    # Returns: One line dataframe with the meta-analysis results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)

    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(alpha = rnorm(1), tau.eps = runif(1),
        mu.theta = rnorm(1, muinit, sdinit))
    chain2 <- list(alpha = rnorm(1), tau.eps = runif(1),
        mu.theta = rnorm(1, muinit, sdinit))
    chain3 <- list(alpha = rnorm(1), tau.eps = runif(1),
```

```
        mu.theta = rnorm(1, muinit, sdinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = "modelHalfBayes_Cauchy.txt", monitor = c("mu.theta",
            "var.theta"), n.chains = 3, method = "rjags",
        burnin = 1e+05, sample = 1e+05)

    # Make dataframe with the results of the meta-analyis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half Bayesian meta-analysis with Cauchy prior done")
    return(results)

    # remove things not in use
    rm(mydf, N, y, tau.y, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)
}

# ------------------Uniform (0, 10)

modelString = "
model {
  for (i in 1:N){
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for Overall effect
  tau.theta <- pow(sigma.theta, -2)     # define precision
  sigma.theta ~ dunif (0, 10)              # prior for standard deviation
  var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_uniform.txt")

# ------------------Uniform (0, 1)

modelString = "
model {
  for (i in 1:N){
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
```

```
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for Overall effect
  tau.theta <- pow(sigma.theta, -2)     # define precision
  sigma.theta ~ dunif (0, 1)              # prior for standard deviation
  var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_uniform1.txt")

# ------------------Uniform (0, 100)
modelString = "
model {
  for (i in 1:N){
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)   # prior for Overall effect
  tau.theta <- pow(sigma.theta, -2)     # define precision
  sigma.theta ~ dunif (0, 100)              # prior for standard deviatio
n
  var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_uniform100.txt")

# Function to run half bayesian analysis with uniform prior


runHalfBayes_uniform <- function(mydf, mymodel) {
    # runs Hybrid Bayesian meta-analysis with Uniform prior

    # Args:

    # mydf: dataset with columns eff.size and var
    # mymodel: name of jags model to use

    # Returns: One line dataframe with the meta-analysis results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)

    # Specify data in a list form
```

```r
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain2 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain3 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = mymodel, monitor = c("mu.theta", "var.theta"),
        method = "rjags", n.chains = 3, burnin = 1e+05,
        sample = 1e+05)

    # Make dataframe with the results of the meta-analyis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with uniform prior done")
    return(results)

    # remove things not in use
    rm(mydf, N, y, var, muinit, sdinit, data, chain1, chain2,
        chain3, inits, sims, results)
}

runHalfBayes_uniform1 <- function(mydf, mymodel) {
    # runs Hybrid Bayesian meta-analysis with Uniform(0, 1)

    # Args:

    # mydf: dataset with columns eff.size and var
    # mymodel: name of jags model to use

    # Returns: One line dataframe with the meta-analysis
    # results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
```

```
    sdinit <- sd(mydf$lnrr)

    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0, 1))
    chain2 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0, 1))
    chain3 <- list(mu.theta = rnorm(1, muinit, sdinit),
        sigma.theta = runif(1, 0, 1))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = mymodel, monitor = c("mu.theta", "var.theta"),
        method = "rjags", n.chains = 3, burnin = 1e+05,
        sample = 1e+05)

    # Make dataframe with the results of the meta-analyis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with uniform prior (0, 1) done")
    return(results)

    # remove things not in use
    rm(mydf, N, y, tau.y, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)
}

# -------------------Gamma

modelString = "
model {
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)  # Prior for the overall effect
  tau.theta ~ dgamma(0.1, 0.1)    # Gamma prior for the precision
  sigma.theta <- 1/sqrt(tau.theta) # get the standard deviation
  var.theta <- pow(sigma.theta, 2)  # set variance to monitor
}
```

```
"
# some temporary filename:
writeLines(modelString, con = "modelHalfBayes_Gamma.txt")

# Function to run half bayesian analysis with gamma prior


runHalfBayes_gamma <- function(mydf) {
    # runs Hybrid Bayesian meta-analysis with Gamma prior

    # Args: mydf: dataset with columns eff.size and var

    # Returns: One line dataframe with the meta-analysis results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)
    tauinit <- 1/sdinit^2

    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(mu.theta = rnorm(1, muinit, sdinit),
        tau.theta = runif(1, 0.8 * tauinit, 1.2 * tauinit))
    chain2 <- list(mu.theta = rnorm(1, muinit, sdinit),
        tau.theta = runif(1, 0.8 * tauinit, 1.2 * tauinit))
    chain3 <- list(mu.theta = rnorm(1, muinit, sdinit),
        tau.theta = runif(1, 0.8 * tauinit, 1.2 * tauinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    gamma <- run.jags(data = datalist, inits = initslist,
        model = "modelHalfBayes_Gamma.txt", monitor = c("mu.theta",
            "var.theta"), n.chains = 3, burnin = 1e+05,
        sample = 1e+05, method = "rjags")

    # Make dataframe with the results of the meta-analysis
    results <- make_JAGS_summary(gamma, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with gamma prior done")
```

```
    return(results)

    # remove things not in use
    rm(mydf, N, y, tau.y, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)
}

# ------------------folded-t-

modelString = "
model{
  for (i in 1:N){ # likelihood
    y[i] ~ dnorm (theta[i], tau.y[i])
    theta[i] ~ dnorm (mu.theta, tau.theta)
  }
  # Setting priors
  mu.theta ~ dnorm (0.0, 1.0E-3)  # Overall effect prior
  # folded t prior
  A <- 5
  v <- 2
  B <- 1/(A*A)
  t.theta ~ dt (0, B, v)
  sigma.theta <- abs(t.theta)
  tau.theta <- pow(sigma.theta, -2)
  # set precision for the likelihood and variance to monitor
  var.theta <- pow(sigma.theta, 2)
}
"
# write model file:
writeLines(modelString, con = "modelHalfBayes_folded_t.txt")

# Function to run bayesian analysis with Folded t prior


runHalfBayes_folded_t <- function(mydf) {
    # runs Hybrid Bayesian meta-analysis with Folded t prior

    # Args:

    # mydf: dataset with columns eff.size and var

    # Returns: One line dataframe with the meta-analysis results

    # Organize data we need
    mydf <- mydf[complete.cases(mydf), ]
    N <- nrow(mydf)  # number of studies
    y <- mydf$lnrr  # the observed effect sizes
    tau.y <- 1/mydf$var.lnrr  # the variance of the effect sizes
    muinit <- mean(mydf$lnrr)
    sdinit <- sd(mydf$lnrr)
```

```
    # Specify data in a list form
    datalist <- list(N = N, y = y, tau.y = tau.y)

    # It is highly recommend to specify the initial values
    # for the chains
    chain1 <- list(mu.theta = rnorm(1, muinit, sdinit),
        t.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain2 <- list(mu.theta = rnorm(1, muinit, sdinit),
        t.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    chain3 <- list(mu.theta = rnorm(1, muinit, sdinit),
        t.theta = runif(1, 0.8 * sdinit, 1.2 * sdinit))
    initslist <- list(chain1, chain2, chain3)

    # Start the MCMC simulation:
    sims <- run.jags(data = datalist, inits = initslist,
        model = "modelHalfBayes_folded_t.txt", monitor = c("mu.theta",
            "var.theta"), n.chains = 3, burnin = 1e+05,
        sample = 1e+05, method = "rjags")

    # Make dataframe with the results of the meta-analysis
    results <- make_JAGS_summary(sims, sigma = mydf$sigma[1],
        true.mu = mydf$true.mu[1], true.tau = mydf$true.tau2[1],
        k = mydf$k[1], mean.n = mydf$mean.n[1])

    # Return dataframe and print 'done'
    print("half bayesian meta-analysis with folded t prior done")
    return(results)

    # remove things not in use
    rm(mydf, N, y, var, A, k, muinit, sdinit, data, chain1,
        chain2, chain3, inits, sims, results)

}
```

Extra function used for plots, not written by the authors, available from https://rpubs.com/sjackman/grid_arrange_shared_legend

```
# This function will allow for multiple plots that shared a legend

grid_arrange_shared_legend <- function(..., ncol = length(list(...)),
    nrow = 1, position = c("bottom", "right")) {

    plots <- list(...)
    position <- match.arg(position)
    g <- ggplotGrob(plots[[1]] + theme(legend.position = position))$grobs
```

```
    legend <- g[[which(sapply(g, function(x) x$name) ==
        "guide-box")]]
    lheight <- sum(legend$height)
    lwidth <- sum(legend$width)
    gl <- lapply(plots, function(x) x + theme(legend.position = "none"))
    gl <- c(gl, ncol = ncol, nrow = nrow)

    combined <- switch(position, bottom = arrangeGrob(do.call(arrangeGrob
,
        gl), legend, ncol = 1, heights = unit.c(unit(1,
        "npc") - lheight, lheight)), right = arrangeGrob(do.call(arrangeG
rob,
        gl), legend, ncol = 2, widths = unit.c(unit(1, "npc") -
        lwidth, lwidth)))

    grid.newpage()
    grid.draw(combined)

    # return gtable invisibly
    invisible(combined)

}
```

Simulation experiments

Simulating datasets

```
# load libraries for parallel computing
library(doParallel)
library(foreach)

# source functions we need
source("/home/pappalardop/bayes2/Functions_Pappalardo_etal.R")

# how many replicates
replicates <- 2000

# set up cluster
cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)

# Divide replicates among cores
x <- foreach(rep = 1:replicates, .combine = rbind) %dopar%
    {
        # Define parameters
```

```
        # AMONG-STUDY VARIANCE (tau2)
        tau <- c(0.1, 0.25, 0.5, 1, 2, 5)

        # OVERALL EFFECT
        mu <- c(0.5)

        # NUMBER OF STUDIES in the meta-analysis
        k <- c(5, 10, 15, 25, 35, 50)

        # NUMBER OF REPLICATES
        mean.n <- c(3, 5, 10, 15, 20, 30)

        # WITHIN-STUDY (sigma2)
        sigma <- c(1)

        # define control mean
        c.mean <- 1

        # create dataset
        datasets <- makeLognormalDataset(mu = mu, tau = tau,
            k = k, sigma = sigma, c.mean = c.mean, mean.n = mean.n)

        # save dataset
        save(datasets, file = paste("/home/pappalardop/bayes2/datasets/ln
rrDatasets_rep",
            rep, sep = ""))
    }

stopCluster(cl)
```

Running meta-analysis

The simulation experiment was run on a high performing cluster using parallel computing.

Meta-analysis using metafor

```
# load libraries for parallel computing
library(doParallel)
library(foreach)
library(plyr)
library(dplyr)
library(parallel)

# source functions we need
source("/home/pappalardop/bayes2/Functions_Pappalardo_etal.R")
```

```r
# how many replicates
replicates <- 2000

# set up cluster
cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)

# Divide replicates among cores
x <- foreach(rep = 1:replicates, .combine = rbind) %dopar%
    {

        # load libraries
        library(plyr)
        library(parallel)

        # load original dataset for this replicate
        load(paste("/home/pappalardop/bayes2/datasets/lnrrDatasets_rep",
            rep, sep = ""))

        # Run meta-analysis with default metafor options
        z.list <- list()
        z.list <- lapply(datasets, runMetafor, tau.method = "REML",
            ci = F)
        z.df <- do.call("rbind.fill", z.list)
        z.df$method <- "weighted_z"

        # Run meta-analysis with Knapp-Hartung option
        knha.list <- list()
        knha.list <- lapply(datasets, runMetafor, tau.method = "REML",
            ci = T)
        knha.df <- do.call("rbind.fill", knha.list)
        knha.df$method <- "weighted_knha"

        # combine dataframes
        traditional <- dplyr::bind_rows(z.df, knha.df)
        traditional$replicate <- rep

        # save summary files
        write.csv(traditional, file = paste("/home/pappalardop/bayes2/res
ults/metafor/metafor_NEW_",
            rep, ".csv", sep = ""), row.names = F)
    }

stopCluster(cl)
```

Meta-analysis using metafor and bootstraping

Bootstrap for the mean effect

```
# load libraries for parallel computing
library(doParallel)
library(foreach)
library(plyr)
library(dplyr)
library(parallel)

# source functions we need
source("Functions_Pappalardo_etal.R")

# how many replicates
replicates <- 2000

# set up cluster
cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)

# Divide replicates among cores
x <- foreach(rep = 1:replicates, .combine = rbind) %dopar%
    {

        # load libraries
        library(parallel)
        library(plyr)

        # load original dataset for this replicate
        load(paste("/home/pappalardop/bayes2/datasets/lnrrDatasets_rep",
            rep, sep = ""))

        # Metafor boot
        boot.list <- list()  # create list to hold results
        boot.list <- lapply(datasets, bootMetafor)  # run metafor boot me
ta-analysis
        boot.df <- do.call("rbind.fill", boot.list)  # put results in a d
ataframe
        boot.df$rep <- rep  # add replicate
        boot.df$method <- "weighted_boot"  # add method

        # save summary files
        write.csv(boot.df, paste("/home/pappalardop/bayes2/results/metafo
rboot/metaforboot",
            rep, ".csv", sep = ""), row.names = F)
```

```
    }

stopCluster(cl)
```

Bootstrap for the among-study variance

```
# load libraries for parallel computing
library(doParallel)
library(foreach)

# source functions we need
source("Functions_Pappalardo_etal.R")

# how many replicates
replicates <- 2000

# set up cluster
cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)

# Divide replicates among cores
x <- foreach(rep = 1:replicates, .combine = rbind) %dopar%
    {

        # load libraries
        library(metafor)
        library(boot)
        library(plyr)

        # load original dataset for this replicate
        load(paste("/home/pappalardop/bayes2/datasets/lnrrDatasets_rep",
            rep, sep = ""))

        # Metafor boot for tau2
        boottau.list <- list()  # create empty list to hold results
        boottau.list <- lapply(datasets, bootMetaforTau)  # run bootstrap
for tau2
        boottau.df <- do.call("rbind.fill", boottau.list)  # put results
in a dataframe
        boottau.df$rep <- rep  # add replicate number
        boottau.df$method <- "weighted_boot"  # add method

        # save summary files
        write.csv(boottau.df, file = paste("/home/pappalardop/bayes2/resu
lts/metaforboot/metafor_boot_tau",
            rep, ".csv", sep = ""), row.names = F)
    }
```

```
stopCluster(cl)
```

Hybrid Bayesian meta-analysis

```
# load libraries for parallel computing
library(doParallel)
library(foreach)

# source functions we need
source("Functions_Pappalardo_etal.R")

# how many replicates
replicates <- 2000

# set cluster and how many cores
cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)

# Divide replicates among cores
x <- foreach(rep = 1:replicates, .combine = rbind) %dopar%
    {
        # load libraries
        library(rjags)
        library(runjags)
        library(coda)
        library(plyr)

        # load original dataset for this replicate
        load(paste("/home/pappalardop/bayes2/datasets/lnrrDatasets_rep",
            rep, sep = ""))

        # Hybrid Bayesian meta-analysis with uniform prior
        halfbayes_uniform.list <- list()  # create list to hold results
        halfbayes_uniform.list <- lapply(datasets, runHalfBayes_uniform_J
AGS)  # run Hybrid Bayesian
        halfbayes_uniform.df <- do.call("rbind.fill", halfbayes_uniform.l
ist)  # put results in a dataframe
        halfbayes_uniform.df$rep <- rep  # add replicate
        halfbayes_uniform.df$method <- "halfbayes.uniform"  # add method

        # save summary files
        write.csv(halfbayes_uniform.df, paste("/home/pappalardop/bayes2/r
esults/bayesian/bayesian",
            rep, ".csv", sep = ""), row.names = F)

    }
```

```
stopCluster(cl)
```

Compile results from different replicates

```
# how many replicates

replicates <- 2000
```

Compile meta-analysis using metafor

```
# load results for the first replicate
mydata <- read.csv("/home/pappalardop/bayes2/results/metafor/metafor_NEW_
1.csv")

# Loop to get all result files
for (i in 2:replicates) {
    thisfile <- try(read.csv(paste("/home/pappalardop/bayes2/results/meta
for/metafor_NEW_",
        i, ".csv", sep = "")))
    error  # to check for errors
 <- class(thisfile)[1]
    print(paste("working in replicate", i, sep = ""))
    # in case of error
    if (error == "try-error") {
        # do nothing
    } else {
        mydata <- rbind(mydata, thisfile)
    }
}
# save summary file
write.csv(mydata, file = "/home/pappalardop/bayes2/results/summary_metafo
r.csv",
    row.names = FALSE)
```

Compile meta analysis using metafor and bootstraping

```
# METAFOR BOOT

# load results for the first replicate
mydata <- read.csv("/home/pappalardop/bayes2/results/metaforboot/metaforb
oot1.csv")

# Loop to get all result files
for (i in 2:replicates) {
```

```
    thisfile <- try(read.csv(paste("/home/pappalardop/bayes2/results/meta
forboot/metaforboot",
        i, ".csv", sep = "")))
    error  # to check for errors
 <- class(thisfile)[1]
    # in case of error
    if (error == "try-error") {
        # do nothing
    } else {
        mydata <- rbind(mydata, thisfile)
    }
}
# save summary file
write.csv(mydata, file = "/home/pappalardop/bayes2/results/summary_metafo
rboot.csv",
    row.names = FALSE)

# METAFOR BOOT TAU

# load results for the first replicate
mydata <- read.csv("/home/pappalardop/bayes2/results/metaforboot/metafor_
boot_tau1.csv")

# Loop to get all result files
for (i in 2:replicates) {
    thisfile <- try(read.csv(paste("/home/pappalardop/bayes2/results/meta
forboot/metafor_boot_tau",
        i, ".csv", sep = "")))
    error  # to check for errors
 <- class(thisfile)[1]
    # in case of error
    if (error == "try-error") {
        # do nothing
    } else {
        mydata <- rbind(mydata, thisfile)
    }
}
# save summary file
write.csv(mydata, file = "/home/pappalardop/bayes2/results/summary_metafo
rboot_tau.csv",
    row.names = FALSE)
```

Compile meta-analysis using Hybrid Bayesian approach

```
# load results for the first replicate
mydata <- read.csv("/home/pappalardop/bayes2/results/bayesian/bayesian1.c
sv")
```

```
# Loop to get all result files
for (i in 2:replicates) {
    thisfile <- try(read.csv(paste("/home/pappalardop/bayes2/results/baye
sian/bayesian",
        i, ".csv", sep = "")))
    error  # to check for errors
 <- class(thisfile)[1]
    # in case of error
    if (error == "try-error") {
        # do nothing
    } else {
        mydata <- rbind(mydata, thisfile)
    }
}

# save summary file
write.csv(mydata, file = "/home/pappalardop/bayes2/results/summary_bayesi
an.csv",
    row.names = FALSE)
```

Exploration of different priors for the Hybrid Bayesian model

We explored different priors for the among-study variance in the hybrid Bayesian model. To analyze the prior's influence on the posterior results, we chose a small (but representative) combination of parameters for the mean number of replicates (n), the number of studies (k), and the among-study variance (σ^2_{among}). We analyzed:

- mean number of replicates: 5, 20
- number of studies: 5, 25, 50
- among-study variance: 0.5, 2, 5

```
# load libraries for parallel computing

library(doParallel)
library(foreach)

# source functions we need

source("Functions_Pappalardo_etal.R")

# set cluster and how many cores

cl = makeCluster(4, type = "FORK")
registerDoParallel(cl)
```

```r
x <- foreach(rep = 1:2000, .combine = rbind) %dopar% {
    # load libraries
    library(rjags)
    library(runjags)
    library(coda)
    library(dplyr)

    # load original dataset for this replicate
    load(paste("/pool/genomics/pappalardop/bayes2/datasets/lnrrDatasets_r
ep",
        rep, sep = ""))

    # subset the target combination of factors

    keep <- c("sigma 1 tau 2 mean.n 5 k 5 mu 0.5 c.mean 1",
        "sigma 1 tau 2 mean.n 5 k 25 mu 0.5 c.mean 1", "sigma 1 tau 2 mea
n.n 5 k 50 mu 0.5 c.mean 1",
        "sigma 1 tau 2 mean.n 20 k 5 mu 0.5 c.mean 1", "sigma 1 tau 2 mea
n.n 20 k 25 mu 0.5 c.mean 1",
        "sigma 1 tau 2 mean.n 20 k 50 mu 0.5 c.mean 1",
        "sigma 1 tau 5 mean.n 5 k 5 mu 0.5 c.mean 1", "sigma 1 tau 5 mean
.n 5 k 25 mu 0.5 c.mean 1",
        "sigma 1 tau 5 mean.n 5 k 50 mu 0.5 c.mean 1", "sigma 1 tau 5 mea
n.n 20 k 5 mu 0.5 c.mean 1",
        "sigma 1 tau 5 mean.n 20 k 25 mu 0.5 c.mean 1",
        "sigma 1 tau 5 mean.n 20 k 50 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 5 k 5 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 5 k 25 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 5 k 50 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 20 k 5 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 20 k 25 mu 0.5 c.mean 1",
        "sigma 1 tau 0.5 mean.n 20 k 50 mu 0.5 c.mean 1")

    minidatasets <- datasets[names(datasets) %in% keep]

    # Half bayes with cauchy prior
    halfbayes_cauchy.list <- list()
    halfbayes_cauchy.list <- lapply(minidatasets, runHalfBayes_cauchy)
    halfbayes_cauchy.df <- do.call("rbind", halfbayes_cauchy.list)
    halfbayes_cauchy.df$prior <- "halfbayes.cauchy"  # add method

    # Half bayes with gamma prior
    halfbayes_gamma.list <- list()
    halfbayes_gamma.list <- lapply(minidatasets, runHalfBayes_gamma)
    halfbayes_gamma.df <- do.call("rbind", halfbayes_gamma.list)
    halfbayes_gamma.df$prior <- "halfbayes.gamma"

    # Half bayes with folded N prior
    halfbayes_foldedn.list <- list()
```

```
    halfbayes_foldedn.list <- lapply(minidatasets, runHalfBayes_folded_N)
    halfbayes_foldedn.df <- do.call("rbind", halfbayes_foldedn.list)
    halfbayes_foldedn.df$prior <- "halfbayes.foldedn"  # add method

    # Half bayes with folded N prior
    halfbayes_foldedt.list <- list()
    halfbayes_foldedt.list <- lapply(minidatasets, runHalfBayes_folded_t)
    halfbayes_foldedt.df <- do.call("rbind", halfbayes_foldedt.list)
    halfbayes_foldedt.df$prior <- "halfbayes.foldedt"  # add method

    # Half bayes with parameter expansion prior
    # halfbayes_paramExp.list <- list() # create list to
    # hold results halfbayes_paramExp.list <-
    # lapply(datasets, runHalfBayes_paramExp)
    # halfbayes_paramExp.df <- do.call('rbind',
    # halfbayes_paramExp.list) halfbayes_paramExp.df$method
    # <- 'halfbayes.paramExp' # add method

    # Half bayes with uniform prior create list to hold
    # results
    halfbayes_uniform1.list <- list()
    halfbayes_uniform.list <- list()
    halfbayes_uniform100.list <- list()

    # run half bayes meta-analysis with uniform prior
    halfbayes_uniform.list <- lapply(minidatasets, runHalfBayes_uniform_J
AGS)
    halfbayes_uniform1.list <- lapply(minidatasets, runHalfBayes_uniform1
,
        mymodel = "modelHalfBayes_uniform1.txt")
    halfbayes_uniform100.list <- lapply(minidatasets, runHalfBayes_unifor
m,
        mymodel = "modelHalfBayes_uniform100.txt")

    # put results in a dataframe
    halfbayes_uniform1.df <- do.call("rbind", halfbayes_uniform1.list)
    halfbayes_uniform.df <- do.call("rbind", halfbayes_uniform.list)
    halfbayes_uniform100.df <- do.call("rbind", halfbayes_uniform100.list
)

    # add method
    halfbayes_uniform1.df$prior <- "halfbayes.uniform1"
    halfbayes_uniform.df$prior <- "halfbayes.uniform10"
    halfbayes_uniform100.df$prior <- "halfbayes.uniform100"

    # combine all data
    priors <- rbind(halfbayes_cauchy.df, halfbayes_gamma.df,
        halfbayes_foldedn.df, halfbayes_foldedt.df, halfbayes_uniform.df,
        halfbayes_uniform100.df, halfbayes_uniform1.df)
```

```
    priors$rep <- rep

    # save summary files
    write.csv(priors, paste("/pool/genomics/pappalardop/bayes2/results/pr
iors/priors2_",
        rep, ".csv", sep = ""), row.names = F)

}

stopCluster(cl)
```

Summarize simulations results

We summarized the performance of traditional and Bayesian meta-analysis to estimate an overall effect for each combination of mu, tau, k and method, averaging between the 2000 replicates (results presented in the main manuscript, Figure 3):

```
library(dplyr)
library(Metrics)

# --------------Load Data------------------

# load metafor results

metafor <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysis_Pa
ula/ComparingMethods/MS/finalSims/Results_2000reps/summary_metafor.csv",
    header = T)

# load bootstraping results

metaforboot <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysi
s_Paula/ComparingMethods/MS/finalSims/Results_2000reps/summary_metaforboo
t.csv",
    header = T)

# load files with Bayesian results

bayesian <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysis_P
aula/ComparingMethods/MS/finalSims/Results_2000reps/summary_bayesian.csv"
,
    header = T)

# --------------Summarize data------------------

# Metafor
```

```r
met <- metafor %>% dplyr::filter(!is.na(obs.mu)) %>% dplyr::group_by(true
.mu,
    true.tau, sigma, k, mean.n, method) %>% dplyr::summarise(ov.effect =
mean(obs.mu),
    tau2 = mean(obs.tau), rmse.eff = rmse(obs.mu, true.mu),
    bias.effect = mean(bias.eff), ci.bias.eff = t.95CI(bias.eff),
    bias.tau2 = mean(bias.tau, na.rm = F), ci.bias.tau = t.95CI(bias.tau)
,
    coverage.mu = mean(cov.mu), coverage.mu.cil = bi.95.l(cov.mu),
    coverage.mu.ciu = bi.95.u(cov.mu), coverage.tau = mean(cov.tau),
    coverage.tau.cil = bi.95.l(cov.tau), coverage.tau.ciu = bi.95.u(cov.t
au),
    effect.width = mean(mu.width), tau2.width = mean(tau.width,
        na.rm = F), ci.mu.width = t.95CI(mu.width), ci.tau.width = t.95CI
(tau.width),
    mean.i2 = mean(i2), i2.ci = t.95CI(i2))

# Metafor with bootstraping

metboot <- metaforboot %>% dplyr::filter(!is.na(obs.mu)) %>%
    dplyr::group_by(true.mu, true.tau, sigma, k, mean.n,
        method) %>% dplyr::summarise(ov.effect = mean(obs.mu),
    rmse.eff = rmse(obs.mu, true.mu), bias.effect = mean(bias.eff),
    ci.bias.eff = t.95CI(bias.eff), coverage.mu = mean(cov.mu),
    coverage.mu.cil = bi.95.l(cov.mu), coverage.mu.ciu = bi.95.u(cov.mu),
    effect.width = mean(mu.width), ci.mu.width = t.95CI(mu.width))

# Hybrid Bayesian

hb <- bayesian %>% dplyr::filter(Rhat.mu < 1.1 | Rhat.tau2 <
    1.1) %>% dplyr::group_by(true.mu, true.tau, sigma, k,
    mean.n, method) %>% dplyr::summarise(ov.effect = mean(obs.median.mu),
    eff.cil = median(obs.mu.hdiL), eff.ciu = median(obs.mu.hdiU),
    tau2 = mean(obs.median.tau), rmse.eff = rmse(obs.median.mu,
        true.mu), rmse.tau = rmse(obs.median.tau, true.tau),
    bias.effect = mean(bias.median.eff), ci.bias.eff = t.95CI(bias.median
.eff),
    bias.tau2 = mean(bias.median.tau), ci.bias.tau = t.95CI(bias.median.t
au),
    coverage.mu = mean(cov.mu), coverage.mu.cil = bi.95.l(cov.mu),
    coverage.mu.ciu = bi.95.u(cov.mu), coverage.tau = mean(cov.tau),
    coverage.tau.cil = bi.95.l(cov.tau), coverage.tau.ciu = bi.95.u(cov.t
au),
    meanRhat.mu = mean(Rhat.mu), meanRhat.tau2 = mean(Rhat.tau2),
    maxRhat.mu = max(Rhat.mu), maxRhat.tau2 = max(Rhat.tau2),
    effect.width = mean(mu.width), ci.mu.width = t.95CI(mu.width),
    tau2.width = mean(tau.width), ci.tau.width = t.95CI(tau.width))
```

```
# remove things not in use

rm(metafor, bayesian, metboot)
```

We summarized the performance of traditional and Bayesian meta-analysis to estimate the among-study variance for each combination of mu, tau, k and method, averaging between the 2000 replicates (results presented in the Supporting Information, Figures S7-S10):

```
# --------------Load Data-----------------

# load Bayesian results

bayesian <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysis_P
aula/ComparingMethods/MS/finalSims/Results_2000reps/summary_bayesian.csv"
,
    header = T)

# load metafor results

metafor <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysis_Pa
ula/ComparingMethods/MS/finalSims/Results_2000reps/summary_metafor.csv",
    header = T)

# load metafor with bootstraping results

metaforboot <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysi
s_Paula/ComparingMethods/MS/finalSims/Results_2000reps/summary_metaforboo
t.csv",
    header = T)

metaforboottau <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnal
ysis_Paula/ComparingMethods/MS/finalSims/Results_2000reps/summary_metafor
boot_tau.csv",
    header = T)

# compile results from the different methods/files

metaforbootall <- metaforboottau %>% dplyr::select(obs.tau,
    tau.ci.lb, tau.ci.ub, bias.tau, cov.tau, tau.width,
    rep, k, mean.n, true.tau) %>% dplyr::right_join(metaforboot,
    by = c("true.tau", "k", "mean.n", "rep"))

# ----------Summarize data-----------------
```

```r
# Metafor

met <- metafor %>% dplyr::filter(!is.na(obs.tau)) %>% dplyr::group_by(tru
e.mu,
    true.tau, sigma, k, mean.n, method) %>% dplyr::summarise(ov.effect =
mean(obs.mu),
    tau2 = mean(obs.tau), rmse.eff = rmse(obs.mu, true.mu),
    bias.effect = mean(bias.eff), ci.bias.eff = t.95CI(bias.eff),
    bias.tau2 = mean(bias.tau, na.rm = F), ci.bias.tau = t.95CI(bias.tau)
,
    coverage.mu = mean(cov.mu), coverage.mu.cil = bi.95.l(cov.mu),
    coverage.mu.ciu = bi.95.u(cov.mu), coverage.tau = mean(cov.tau),
    coverage.tau.cil = bi.95.l(cov.tau), coverage.tau.ciu = bi.95.u(cov.t
au),
    effect.width = mean(mu.width), tau2.width = mean(tau.width,
        na.rm = F), ci.mu.width = t.95CI(mu.width), ci.tau.width = t.95CI
(tau.width),
    rmse.tau = rmse(tau2, true.tau))

# Metafor with bootstraping

metboot <- metaforbootall %>% dplyr::filter(!is.na(obs.tau)) %>%
    dplyr::group_by(true.mu, true.tau, sigma, k, mean.n,
        method) %>% dplyr::summarise(ov.effect = mean(obs.mu),
    tau2 = mean(obs.tau), rmse.eff = rmse(obs.mu, true.mu),
    bias.effect = mean(bias.eff), ci.bias.eff = t.95CI(bias.eff),
    bias.tau2 = mean(bias.tau, na.rm = F), ci.bias.tau = t.95CI(bias.tau)
,
    coverage.mu = mean(cov.mu), coverage.mu.cil = bi.95.l(cov.mu),
    coverage.mu.ciu = bi.95.u(cov.mu), coverage.tau = mean(cov.tau),
    coverage.tau.cil = bi.95.l(cov.tau), coverage.tau.ciu = bi.95.u(cov.t
au),
    effect.width = mean(mu.width), tau2.width = mean(tau.width,
        na.rm = F), ci.mu.width = t.95CI(mu.width), ci.tau.width = t.95CI
(tau.width),
    rmse.tau = rmse(tau2, true.tau))


# Hybrid Bayesian

hb <- bayesian %>% dplyr::filter(Rhat.mu < 1.1 | Rhat.tau2 <
    1.1) %>% dplyr::group_by(true.mu, true.tau, sigma, k,
    mean.n, method) %>% dplyr::summarise(ov.effect = mean(obs.median.mu),
    eff.cil = median(obs.mu.hdiL), eff.ciu = median(obs.mu.hdiU),
    tau2 = mean(obs.median.tau), rmse.eff = rmse(obs.median.mu,
        true.mu), rmse.tau = rmse(obs.median.tau, true.tau),
    bias.effect = mean(bias.median.eff), ci.bias.eff = t.95CI(bias.median
.eff),
    bias.tau2 = mean(bias.median.tau), ci.bias.tau = t.95CI(bias.median.t
```

```
au),
    coverage.mu = mean(cov.mu), coverage.mu.cil = bi.95.l(cov.mu),
    coverage.mu.ciu = bi.95.u(cov.mu), coverage.tau = mean(cov.tau),
    coverage.tau.cil = bi.95.l(cov.tau), coverage.tau.ciu = bi.95.u(cov.t
au),
    meanRhat.mu = mean(Rhat.mu), meanRhat.tau2 = mean(Rhat.tau2),
    maxRhat.mu = max(Rhat.mu), maxRhat.tau2 = max(Rhat.tau2),
    effect.width = mean(mu.width), ci.mu.width = t.95CI(mu.width),
    tau2.width = mean(tau.width), ci.tau.width = t.95CI(tau.width))


# remove objects not in use

rm(metafor, bayesian, metaforbootall, metaforboot, metaforboottau)
```

This is how we compiled and summarized the exploration of the different priors for the among-study variance in the Hybrid Bayesian model:

```
# load priors data

priors <- read.csv("c:/Users/Paula/Dropbox/Meta-analysis/MetaAnalysis_Pau
la/ComparingMethods/MS/finalSims/Results_2000reps/summary_priors_allreps.
csv",
    as.is = T)

# create factors for plots

priors$f.prior <- factor(priors$prior, levels = c("halfbayes.cauchy",
    "halfbayes.gamma", "halfbayes.foldedn", "halfbayes.foldedt",
    "halfbayes.uniform10", "halfbayes.uniform100", "halfbayes.uniform1"),
    labels = c("Cauchy", "Gamma", "Folded n", "Folded t",
        "Uniform 10", "Uniform 100", "Uniform 1"))
priors$f.tau2 <- factor(priors$true.tau, levels = c(0.5,
    2, 5), labels = c("tau2= 0.5", "tau2= 2", "tau2= 5"))
priors$f.k <- factor(priors$k, levels = c(5, 25, 50), labels = c("k= 5",
    "k= 25", "k= 50"))

# count total simulations in each combination

countsbygroup <- priors %>% dplyr::add_count(prior, true.tau,
    k, mean.n) %>% dplyr::rename(ntotal = n) %>% dplyr::distinct(prior,
    true.tau, k, mean.n, ntotal)

# count simulations with 'good' R hats

goodrhat <- priors %>% dplyr::filter(Rhat.tau2 < 1.1) %>%
```

```
    dplyr::add_count(prior, true.tau, k, mean.n) %>% dplyr::distinct(prio
r,
    true.tau, k, mean.n, n) %>% dplyr::rename(n.good = n)

# combine information for final summary
summaryrhat <- countsbygroup %>% dplyr::left_join(., goodrhat) %>%
    dplyr::mutate(n.bad = ntotal - n.good) %>% dplyr::mutate(prop.goodrha
t = n.good/ntotal) %>%
    dplyr::mutate(prop.badrhat = 1 - prop.rhat)
```

Summarize data from the literature review

The data compiled in the literature review is included as an excel file in the Dryad data repository.

Here the code we used to summarize the general information from the literature review, and information about the number of replicates and number of studies:

```
# load libraries

library(readxl)
library(tidyr)
library(Rmisc)
library(dplyr)
library(ggplot2)
library(Metrics)
library(modeest, quietly = T)

# load literature search data directly from excel file

rev <- as.data.frame(read_excel("C:/Users/Paula/Dropbox/Meta-analysis/Met
aAnalysis_Paula/ComparingMethods/MS/ns_ks_search/extractingData_Jan2019.x
lsx",
    sheet = "Revision", range = cell_cols("A:J")))

main <- as.data.frame(read_excel("C:/Users/Paula/Dropbox/Meta-analysis/Me
taAnalysis_Paula/ComparingMethods/MS/ns_ks_search/extractingData_Jan2019.
xlsx",
    sheet = "ClimChange", range = cell_cols("A:AJ")))

ks <- as.data.frame(read_excel("C:/Users/Paula/Dropbox/Meta-analysis/Meta
Analysis_Paula/ComparingMethods/MS/ns_ks_search/extractingData_Jan2019.xl
sx",
    sheet = "ClimChange_ks", range = cell_cols("A:G")))

ns <- as.data.frame(read_excel("C:/Users/Paula/Dropbox/Meta-analysis/Meta
Analysis_Paula/ComparingMethods/MS/ns_ks_search/extractingData_Jan2019.xl
```

```
sx",
    sheet = "ClimChange_ns", cell_cols("A:H"), col_types = c("numeric",
        "text", "skip", "skip", "numeric", "skip", "skip",
        "numeric")))


# ------------Number of replicates----------------

# calculate mean ni from treatment and control
ns$mean.rep <- rowMeans(ns[c("nt", "nc")], na.rm = T)

# add paper identifiers to ns and ks
ns <- main %>% dplyr::select(pdf.name, change.topic, topic.group,
    taxa, environment) %>% dplyr::right_join(ns, by = "pdf.name") %>%
    dplyr::select(paper.id, pdf.name, topic.group, taxa,
        environment, mean.rep) %>% dplyr::filter(!is.na(mean.rep))

# What proportion of ns are lower than 10

total.ns <- ns %>% nrow()
less10row <- ns %>% dplyr::filter(mean.rep < 10) %>% nrow()
perless <- less10row * 100/total.ns

# calculate median number of replicates
median.rep <- median(ns$mean.rep, na.rm = T)

# get the mode of the number of replicates
mode.n <- mfv(ns$mean.rep)[1]

# ------------Number of studies-------------------

ks <- main %>% dplyr::select(pdf.name, change.topic, topic.group,
    taxa, environment) %>% dplyr::right_join(ks, by = "pdf.name")

# subset ks by type of grouping

ks.analysis <- ks %>% filter(k.type == "analysis")
ks.overall <- ks %>% filter(k.type == "overall")
ks.category <- ks %>% filter(k.type == "category")

# what proportion of ks are lower than 20

total.ks <- ks.analysis %>% nrow()
less20row <- ks.analysis %>% filter(k < 20) %>% nrow
perless <- less20row * 100/total.ks

# what proportion of ks are lower than 40

total.ks <- ks.analysis %>% nrow()
```

```
less40row <- ks.analysis %>% filter(k < 40) %>% nrow
perless <- less40row * 100/total.ks

# get the modes for the number of studies

mode.ka <- mfv(ks.analysis$k)[1]
mode.kc <- mfv(ks.category$k)[1]

# get the median for the number of studies

median.ka <- median(ks.analysis$k)
median.kc <- median(ks.category$k)
median.ko <- median(ks.overall$k)

# ----------Revision related------------------

# check with journals are publishing more meta-analysis

byjournal <- rev %>% dplyr::filter(inclusion.criterion ==
    "yes") %>% dplyr::group_by(pub.name) %>% dplyr::count() %>%
    dplyr::arrange(desc(n))

# count papers reviewed for climate change

revised <- rev %>% dplyr::filter(!is.na(inclusion.criterion))

# papers included in the analysis

included <- rev %>% dplyr::filter(inclusion.criterion ==
    "yes")

# count papers providing information on ks and ns with
# paper id

ksinfo <- length(unique(ks$pdf.name))
nsinfo <- length(unique(ns$pdf.name))
```

Here the code used to summarize information about the type of confidence interval used to calculate a mean effect size:

```
library(plyr)
library(ggplot2)

#--------------CI's-----------------
```

```r
# count number of papers reporting each uncertainty interval

citable <- main %>% dplyr::count(ci.type) %>% dplyr::mutate(percent = n/s
um(n) *
    100)

# add a zero entry for KHSJ that was not reported

citable <- rbind(citable, c("KHSJ", 0, 0))

# re convert n as numeric

citable$n <- as.numeric(citable$n)

# add factor for plots

citable$f.ci <- factor(citable$ci.type, levels = c("not mentioned",
    "bootstrap", "z", "z and bootstrap", "se", "Bayesian",
    "Bayesian and bootstrap", "KHSJ"), labels = c("not mentioned",
    "bootstrap", "z-distribution", "z and bootstrap", "standard error",
    "Bayesian", "Bayesian and bootstrap", "KHSJ"))

# match type of CI with each k

kci <- ks %>% dplyr::left_join(main, by = "pdf.name") %>%
    dplyr::filter(k <= 40 & k.type == "analysis") %>% dplyr::group_by(ci.
type) %>%
    dplyr::count()

# how many low ks

under10 <- ks.analysis %>% dplyr::filter(k <= 10)
p.under10 <- (nrow(under10) * 100)/nrow(ks.analysis)

under40 <- ks.analysis %>% dplyr::filter(k <= 40)
p.under40 <- (nrow(under40) * 100)/nrow(ks.analysis)

# -----------Software-----------------

# check which software used the papers that didn't
# mention the type of CI

nms <- main %>% dplyr::filter(ci.type == "not mentioned")

metawin <- nms %>% dplyr::filter(software == "Metawin") %>%
    nrow()
cma <- nms %>% dplyr::filter(software == "CMA") %>% nrow()
metafor <- nms %>% dplyr::filter(software == "R, metafor") %>%
    nrow()
```

```r
used.defaults <- metawin + cma + metafor

# if assuming software defaults, count new z and t CIs

newz <- as.numeric(citable %>% filter(ci.type == "z") %>%
    magrittr::extract2("n")) + cma + metafor
newt <- metawin

# % percentage of papers that didn't report confidence interval

percent.nas <- as.numeric(citable %>% filter(ci.type ==
    "not mentioned") %>% magrittr::extract2("percent"))
```