METHODOLOGY

Quantitative evidence synthesis: a practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences

Shinichi Nakagawa[1*], Yefeng Yang[1], Erin L. Macartney[1], Rebecca Spake[2], and Malgorzata Lagisz[1].

[1] Evolution & Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

[2] School of Biological Sciences, Whiteknights Campus, University of Reading, Reading, RG6 6AS, UK

* Correspondence: S. Nakagawa & Yefeng Yang

e-mail: s.nakagawa@unsw.edu.au

e-mail: yefeng.yang1@unsw.edu.au

Short title: a practical guide for quantitative synthesis

Abstract

Meta-analysis is a quantitative way of synthesizing results from multiple studies to obtain reliable evidence of an intervention or phenomenon. Indeed, an increasing number of meta-analyses are conducted in environmental sciences, and resulting meta-analytic evidence is often used in environmental policies and decision-making. We conducted a survey of recent meta-analyses in environmental sciences and found poor standards of current meta-analytic practice and reporting. Only ~40% of the 73 reviewed meta-analyses reported heterogeneity (variation among effect sizes beyond sampling error), and publication bias was assessed in less than half. Furthermore, although almost all the meta-analyses had multiple effect sizes originating from the same studies, non-independence among effect sizes was considered in only half of the meta-analyses. To improve the implementation of meta-analysis in environmental sciences, we here outline practical guidance for conducting a meta-analysis in environmental sciences. We describe the key concepts of effect size statistics and meta-analysis, and detail procedures for fitting multilevel meta-analysis and meta-regression models and performing associated publication bias tests. We demonstrate a clear need for environmental scientists to embrace multilevel meta-analytic models, which explicitly model dependence among effect sizes, rather than the commonly used random-effects models. Further, we discuss how reporting and visual presentations of meta-analytic results can be much improved by following reporting guidelines such as PRISMA-EcoEvo (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Ecology and Evolutionary Biology). This paper, along with the accompanying online tutorial (link), serves as a practical guide on conducting a complete set of meta-analytic procedures (i.e., meta-analysis, heterogeneity quantification, meta-regression, publication bias tests and sensitivity analysis) and also as a gateway to more advanced, yet appropriate, methods.

KEYWORDS

Hierarchical models, robust variance estimation, spatial dependency, variance-covariance matrix, meta-analysis of variance, missing data, network meta-analysis, multivariate meta-analysis

Background

Evidence synthesis is an essential part of science. The method of systematic review provides the most trusted and unbiased way to achieve the synthesis of evidence [1-3]. Systematic reviews often include a quantitative summary of studies on the topic of interest, referred to as a meta-analysis (for discussion on the definitions of 'meta-analysis', see [4]). The term meta-analysis can also mean a set of statistical techniques for quantitative data synthesis. The methodologies of the meta-analysis were initially developed and applied in medical and social sciences. However, meta-analytic methods are now used in many other fields, including environmental sciences [5-7]. In environmental sciences, the outcomes of meta-analyses (within systematic reviews) have been used to inform environmental and related policies (see [8]). Therefore, the reliability of meta-analytic results in environmental sciences is important beyond mere academic interests; indeed, incorrect results could lead to ineffective or sometimes harmful environmental policies [8].

As in medical and social sciences, environmental scientists frequently use traditional meta-analytic models, namely fixed-effect and random-effects models [9, 10]. However, we contend that such models in their original formulation are no longer useful and are often incorrectly used, leading to unreliable estimates and their errors. This is mainly because the traditional models assume independence among effect sizes, but almost all primary research papers include more than one effect size, and this non-independence is often not considered (e.g., [11-13]). Furthermore, previous reviews of published meta-analyses in environmental sciences (hereafter, 'environmental meta-analyses') have demonstrated that less than half report or investigate heterogeneity (inconsistency) among effect sizes [14-16]. Many environmental meta-analyses also do not present any sensitivity analysis, for example, for publication bias (i.e., statistically significant effects being more likely to be published, making collated data unreliable; [17, 18]). These issues might have arisen for several reasons, for example, because of no clear conduct guideline for statistical part of meta-analyses in enviremetal sciences and rapid developments in meta-analytic methods. Taken together, the field

urgently requires a practical guide to implement correct meta-analyses and associated procedures (e.g., heterogeneity analysis, meta-regression, and publication bias tests; cf. [19]).

To assist environmental scientists conducting meta-analyses, the aims of this paper are five-fold. First, we provide an overview of the processes involved in a meta-analysis while introducing some key concepts. Second, after introducing the main types of effect size (statistics), we mathematically describe the two commonly used traditional meta-analytic models, demonstrate their futility, and introduce a practical, multilevel meta-analytic model for environmental sciences that appropriately handles non-independence among effect sizes. Third, we show how to quantify heterogeneity (i.e., consistencies among effect sizes and/or studies) using this model, and then and explain such heterogeneity using meta-regression. Fourth, we show how to test for publication bias in a meta-analysis and describe other common types of sensitivity analysis. Fifth, we cover other technical issues relevant to environmental sciences (e.g., scale and phylogenetic dependence) as well as some advanced meta-analytic techniques. In addition, these five aims (sections) are interspersed with two more sections, named 'Notes' on: 1) visualisation and interpretation and 2) reporting and archiving. Some of these sections are accompanied by results from a survey of 73 environmental meta-analyses published between 2019 and 2021; survey results depict current practices and highlight associated problems (for the method of the survey, see Supporting Information). Importantly, we provide easy-to-follow implementations of much of what is described below, using the *R* package, *metafor* [20] and other *R* packages at the webpage (https://itchyshin.github.io/Meta-analysis_tutorial/), which also connects the reader to the wealth of online information on meta-analysis (see also [21]).

Overview with key concepts

Statistically speaking, we have three general objectives when conducting a *meta-analysis* [12]: 1) estimating *an overall mean*, 2) quantifying consistency (*heterogeneity*) between studies, and 3)

explaining the heterogeneity (see Table 1 for the definitions of the terms in *italic*). A notable feature of a meta-analysis is that an overall mean is estimated by taking the *sampling variance* of each *effect size* into account: a study (effect size) with a low sampling variance (usually based on a larger sample size) is assigned more weight in estimating an overall mean than one with a high sampling variance (usually based on a smaller sample size). However, an overall mean estimate itself is often not informative because one can get the same overall mean estimates in different ways. For example, we may get an overall estimate of zero if all studies have zero effects with no heterogeneity. In contrast, we might also obtain a zero mean across studies that have highly variable effects (e.g., ranging from strongly positive to strongly negative), signifying high heterogeneity. Therefore, quantifying indicators of heterogeneity is an essential part of a meta-analysis, necessary for interpreting the overall mean appropriately. Once we observe non-zero heterogeneity among effect sizes, then, our job is to explain this variation by running *meta-regression* models, and, at the same time, quantify how much variation is accounted for (often quantified as R^2). In addition, it is important to conduct an extra set of analyses, often referred to as *publication bias tests*, which are a type of *sensitivity analysis* [11], to check the robustness of meta-analytic results.

Choosing an effect size statistic

In this section, we introduce different kinds of 'effect size statistics' or 'effect statistics'. In the literature, the term 'effect size' is typically used to refer to the magnitude or strength of an effect of interest or its biological interpretation (e.g., environmental significance). Effect sizes can be quantified using a range of statistics (for details, see [22]). In our survey of environmental meta-analyses, the two most commonly used effect size statistics are: the logarithm of response ratio, lnRR ([23]; also known as the ratio of means; [24]) and standardized mean difference, SMD (often referred to as Hedges' *g* or Cohen's *d* [25, 26]). These are followed by proportion (%) and Fisher's *z*-transformation of correlation, or *Zr*. These four effect statistics nearly fit into the three categories, which are named: 1) single-group statistics (a statistical summary form one group; e.g., proportion), 2) comparative statistics (comparing statistics between two groups e.g., SMD and lnRR), and 3)

association statistics (relationships between two variables; e.g., *Zr*). Table 2 summarizes effect statistics which are common or potentially useful for environmental scientists. Some of the readers may be surprised that 'mean' can be an effect size for a meta-analysis, although once one realises proportion is a type of mean, this makes sense (e.g., the mean concentration values of a certain pollutant can be meta-analysed across studies). Also, this example nicely illustrates that any statistics with sampling variance can become an 'effect size'. The main reason why SMD, lnRR, *Zr,* or proportion are popular effect statistics is that they are unitless, while a meta-analysis of mean, or mean difference, can only be conducted when all effect sizes have the same unit (e.g., cm, kg).

Table 2 also includes effect statistics that are likely to be unfamiliar to environmental scientists; these are effect sizes that characterise differences in the observed variability between samples, (i.e., lnSD, lnCV, lnVR and lnCVR; [27, 28]) rather than central tendencies (averages). These dispersion-based effect statistics can provide us with extra insights along with average-based effect statistics. Although the literature survey showed none of these were used in our sample, these effect sizes have been used in many fields, including agriculture (e.g., [29]), ecology (e.g., [30]), evolutionary biology (e.g., [31]), psychology (e.g., [32]), education (e.g., [33]), psychiatry (e.g., [34]), neurosciences (e.g., [35]), and more. Perhaps, it is not difficult to think of an environmental intervention that can affect not only mean but also variance of a group of individuals or plots. Indeed, per unit yield, organic agriculture has a lower temporal stability compared to conventional agriculture, although the former promotes greater biodiversity [29],. Also, environmental stressors such as pesticides and eutrophication are likely to increase variability in biological systems because stress accentuates individual differences in environmental responses (e.g., [36, 37]). Such ideas are yet to be tested meta-analytically (cf. [38, 39]).

Choosing a meta-analytic model

Two traditional models & a proposed practical model

The fixed-effect model, which should probably be more correctly referred to as the 'common-effect' model, can be written as [9, 10, 40]:

$$z_j = \beta_0 + m_j, \quad (1)$$

$$m_j \sim \mathrm{N}(0, v_j),$$

where the intercept, β_0 is the overall mean, z_j (the response/dependent variable) is the effect size from the jth study (j = 1, 2,…, N_{study}; in this model, N_{study} = the number of studies = the number of effect sizes), m_j is the effect of the jth sampling variance (v_j), which is normally distributed with the mean of 0 and the 'study-specific' sampling variance, v_j (see also Figure 1A). The overall mean is basically the same as a weighted average with the weights, $w_j = 1/v_j$. An important assumption of meta-analysis is that sampling variance is known (note that we can usually estimate sampling variance correctly and unbiasedly, using formulas; Table 2; but see section 'Scale dependence'). As you may see from this formulation, the fixed-effect model assumes that the only source of variation in effect sizes (z_j) is the effect due to sampling variance (whose main determinant is the sample size, n; Table 2).

Similarly, the random-effects model can be expressed as:

$$z_j = \beta_0 + u_j + m_j, \quad (2)$$

$$u_j \sim \mathrm{N}(0, \tau^2), \& \; m_j \sim \mathrm{N}(0, v_j),$$

where u_j is the jth study effect (or 'paper effect', as we here equate 'study' with 'paper'), which is normally distributed with the mean of 0 and the between-study variance, τ^2, and other notations are the same as in Equation 1 (Figure 1B). Here, the overall mean is the weighted average with weights $w_j = 1/(\tau^2 + v_j^2)$. The model assumes each study has its specific mean, $b_0 + u_j$, and

168 (in)consistencies among studies (effect sizes) are indicated by τ^2. When τ^2 is 0 (or not different
169 from 0), the random-effects model simplifies to the fixed-effect model (cf. Equations 1 & 2). Given
170 no studies in environmental sciences are conducted in the same manner or even at exactly the same
171 place and time, we should expect different studies to have different means. Therefore, in almost all
172 cases in the environmental sciences, the random-effects model is a more 'realistic' model [9, 10,
173 40]. Accordingly, most environmental meta-analyses (68.5%; 50 out of 73 studies) in our survey
174 used the random-effects model, while only 2.7% (2 of 73 studies) used the fixed-effect model.
175

176 **Multilevel meta-analytic models**

177 Although we have introduced the random-effects model as being more realistic than the fixed-effect
178 model (Equation 2), we argue that the random-effects model is rather limited and impractical for the
179 environmental sciences. This is because random-effects models, like fixed-effect models, assume all
180 effect sizes (z_j) to be independent. However, when multiple effect sizes are obtained from a study
181 (paper), these effect sizes are dependent (for more details, see the next section on non-
182 independence). Indeed, our survey showed that in almost all datasets used in environmental meta-
183 analyses, this type of non-independence among effect sizes occurred (97.3%; 71 out of 73 studies,
184 with two studies being unclear, so effectively 100%). Therefore, we propose the simplest and most
185 practical meta-analytic model for environmental sciences as [13, 40] (see also [41, 42]):

186 $$z_i = \beta_0 + u_{j[i]} + e_i + m_i, \qquad (3)$$

187 $$u_j \sim \mathrm{N}(0, \tau^2),\ e_i \sim \mathrm{N}(0, \sigma^2), \&\ m_i \sim \mathrm{N}(0, v_i)$$

188 where we explicitly recognize that N_{effect} (i = 1, 2,…, N_{effect}) > N_{study} (j = 1, 2,…, N_{study}) and,
189 therefore, we now have the study effect (between-study effect), $u_{j[i]}$ (for the jth study and ith effect
190 size) and effect-size level (within-study) effect, e_i (for the ith effect size), with the between-study
191 variance, τ^2, and with-study variance, σ^2, respectively, and other notations are the same as above.
192 We note that this model (Equation 3) is an extension of the random-effects model (Equation 2), and

193 we refer to it as the multilevel/hierarchical model (used in 7 out of 73 studies: 9.6%; note that
194 Equation 3 is also known as a three-level meta-analytic model; Figure 1C). Also, environmental
195 scientists who are familiar with (generalised) linear mixed-models may recognize u_j (the study
196 effect) as the effect of a random factor which is associated with a variance component, i.e., τ^2 [43];
197 also, e_i and m_i can be seen as parts of random factors, associated with σ^2 and v_i (the former is
198 comparable to the residuals, while the latter is something special for a meta-analysis).
199 It seems that many researchers are aware of the issue of non-independence so that they often use
200 average effect sizes per study (paper) or choose one effect size (at least 28.8%, 21 out of 73
201 environmental meta-analyses). However, as we discussed elsewhere [13, 40], such averaging or
202 selection of one effect size per study dramatically reduces our ability to investigate environmental
203 drivers of variation among effect sizes [13]. Therefore, we strongly support the use of the multilevel
204 model. Nevertheless, this proposed multilevel model, formulated as Equation 3 does not usually
205 deal with the issue of non-independence completely, which we elaborate on in the next section.

206 **Non-independence among effect sizes and among sampling errors**

207 When you have multiple effect sizes from a study (paper), there are two broad types and three cases
208 of non-independence (cf. [11, 12]): 1) effect sizes are calculated from different cohorts of
209 individuals (or groups of plots) within a study (Figure 2A, referred to as 'shared study identity'),
210 and 2) effects sizes are calculated from the same cohort of individuals (or group of plots; Figure 2B,
211 referred to as 'shared measurements') or *partially* from the same individuals and plots, more
212 concretely, sharing individuals and plots from the control group (Figure 2C, referred to as 'shared
213 control group'). The first type of non-independence induces dependence among effect sizes, but not
214 among sampling variances, and the second type leads to non-independence among sampling
215 variances. Many datasets, if not almost all, will have a combination of these three cases (or even are
216 more complex, see the section 'Complex non-independence'). Failing to deal with these non-
217 independences will inflate Type 1 error (note that the overall estimate, b_0 is unlikely to be biased,
218 but standard error of b_0, se(b_0), will be underestimated; note that this is also true for all other

219 regression coefficients, e.g., b_1; see Table 1). The multilevel model (as in Equation 3) only takes
220 care of cases of non-independence that are due to the shared study identity but neither shared
221 measurements nor shared control group.
222
223 There are two practical ways to deal with non-independence among sampling variances. The first
224 method is that we explicitly model such dependence using a variance-covariance (VCV) matrix
225 (used in 6 out of 73 studies: 8.2%). Imagine a simple scenario with a dataset of three effect sizes
226 from two studies where two effects sizes from the first study are calculated (partially) using the
227 same cohort of individuals (Figure 2B); in such a case, the sampling variance effect, m_i, as in
228 Equation 3, should be written as:

229 $$m_i \sim \mathrm{N}(0, \mathbf{M})$$

230 $$\mathbf{M} = \begin{bmatrix} v_{1[1]} & \rho\sqrt{v_{1[1]}v_{1[2]}} & 0 \\ \rho\sqrt{v_{1[2]}v_{1[1]}} & v_{1[2]} & 0 \\ 0 & 0 & v_{2[3]} \end{bmatrix}, \quad (4)$$

231 where **M** is the VCV matrix showing the sampling variances, $v_{1[1]}$ (study 1 & effect size 1), $v_{1[2]}$
232 (study 1 & effect size 2), and $v_{2[3]}$ (study 2 & effect size 3) in its diagonal, and sampling
233 covariance, $\rho\sqrt{v_{1[1]}v_{1[2]}} = \rho\sqrt{v_{1[2]}v_{1[1]}}$ in its off-diagonal elements, where ρ is a correlation
234 between two sampling variances due to shared samples (individuals/plots). Once this VCV matrix is
235 incorporated into the multilevel model (Equation 3), all the types of non-independence, as in Figure
236 2, are taken care of. Table 3 shows formulas for the sampling variance and covariance of the four
237 common effect sizes (SDM, lnRR, proportion and *Zr*). For comparative effect statistics (Table 2),
238 exact covariances can be calculated under the case of 'shared control group' (see [44, 45]). But this
239 is not feasible for most circumstances because we usually do not know what ρ should be. Some
240 have suggested to fix this value at 0.5 (e.g., [11]) or 0.8 (e.g., [46]); the latter is a more conservative
241 assumption. Or one can run both and use one for the main analysis and the other for sensitivity
242 analysis (for more, see the section 'Conducting sensitivity analysis & critical appraisal').

The second method overcomes this very issue of unknown ρ by approximating average dependence among sampling variance (and effect sizes) from the data and incorporating such dependence to estimate standard errors (only used in 1 out of 73 studies). This method is known as 'robust variance estimation', RVE, and the original estimator was proposed by Hedges and colleagues in 2010 [47]. Meta-analysis using RVE is relatively new, and this method has been applied to multilevel meta-analytic models only recently [48]. Note that the random-effects model (Equation 2) and RVE could correctly model both types of non-independence. However, we do not recommend the use of RVE with Equation 2 because, as we will later show, estimating σ^2 as well as τ^2 will constitute the important part of understanding and gaining more insights from one's data. We do not yet have a definite recommendation on which method to use to account for non-independence among sampling errors (using the VCV matrix or RVE). This is because no simulation work in the context of multilevel meta-analysis has been done so far, using multilevel meta-analyses [13, 48]. For now, one could use both VCV matrices and RVE in the same model [48] (see also [21]).

Quantifying & explaining heterogeneity

Measuring consistencies with heterogeneity

As mentioned earlier, quantifying heterogeneity among effect sizes is an essential component of any meta-analysis. Yet, our survey showed only 28 out of 73 environmental meta-analyses (38.4%) report at least one index of heterogeneity (e.g., τ^2, Q, and I^2). Conventionally, the presence of heterogeneity is tested by Cochrane's Q test. However, Q (often denoted as Q_T or Q_{total}), and its associated p value, are not particularly informative: the test diesn't tell about the extent of heterogeneity (e.g., [10]), only whether heterogeneity is zero or not (when $p < 0.05$). Therefore, for environmental scientists, we recommend two common ways of quantifying heterogeneity from a meta-analytic model: absolute heterogeneity measure (i.e., variance components, τ^2 and σ^2) and

relative heterogeneity measure (i.e., I^2; see also the section 'Notes on visualisation and interpretation' for another way of quantifying and visualising heterogeneity at the same time, using prediction intervals). We have already covered the absolute measure (Equations 2 & 3), so here we explain I^2, which ranges from 0 to 1. The heterogeneity measure, I^2, for the random-effect model (Equation 2) can be written as:

$$I^2 = \frac{\tau^2}{\tau^2 + \bar{v}}, \quad (5)$$

$$\bar{v} = \frac{\left(N_{effect} - 1\right)\sum_{j=1}^{k}(1/v_i)}{\left(\sum_{j=1}^{k}(1/v_i)\right)^2 - \sum_{j=1}^{k}(1/v_i)^2}, \quad (6)$$

where $\bar{v}$ is referred to as the typical sampling variance (originally this is called 'within-study' variance, as in Equation 2, and note that in this formulation, within-study effect and the effect of sampling error is confounded; see [49, 50]; see also [51]) and the other notations are as above. As you can see from Equation 5, we can interpret I^2 as relative variance due to differences between studies, or not due to sampling variance.

By seeing I^2 as a type of interclass correlation (also known as repeatability, [52]), we can generalize I^2 to multilevel models. In the case of Equation 3 ([40, 53]; see also [42]), we have:

$$I^2_{total} = \frac{\tau^2 + \sigma^2}{\tau^2 + \sigma^2 + \bar{v}}. \quad (7)$$

Because we can have two more I^2, Equation 7 is denoted as I^2_{total}; these other two are I^2_{study} and I^2_{effect}, respectively:

$$I^2_{study} = \frac{\tau^2}{\tau^2 + \sigma^2 + \bar{v}}, \quad (8)$$

$$I^2_{effect} = \frac{\sigma^2}{\tau^2 + \sigma^2 + \bar{v}}. \quad (9)$$

I^2_{total} represents variance due to differences both between and within studies or not due to differences in sampling variance, while I^2_{study} is variance due to differences between studies, and I^2_{effect} is variance due to differences within studies (Figure 3A). Once heterogeneity is quantified (note almost all data will have non-zero heterogeneity and an earlier meta-meta-analysis suggests in ecology, we have, on average, I^2 close to 90%, [53]), it is time to fit a meta-regression model to explain the heterogeneity. Notably, the magnitude of I^2_{study} (or τ^2) and I^2_{effect} (or σ^2) can already inform you which predictor variable (usually referred to as 'moderator') is likely to be important, which we explain in the next section.

Explaining variance with meta-regression

We can extend the multilevel model (Equation 3) to a meta-regression model with one moderator (also known as predictor, independent, explanatory variable, or fixed factor), as below:

$$z_i = \beta_0 + \beta_1 x_{1j[i]} + u_{j[i]} + e_i + m_i, \quad (10)$$

where β_1 is a slope of the moderator (x_1), $x_{1j[i]}$ denotes the value of x_1, corresponding to the *j*th study (and the *i*th effect sizes). Equation 10 (meta-regression) is comparable to the simplest regression with the intercept (β_0) and slope (β_1). Notably, $x_{1j[i]}$ differs between studies and, therefore, it will mainly explain the variance component, τ^2 (which relates to I^2_{study}). On the other hand, if noted like x_{1i}, this moderator would vary within studies or at the level of effect sizes, therefore, explaining σ^2 (relating to I^2_{effect}). Therefore, when τ^2 (I^2_{study}), or σ^2 (I^2_{effect}), is close to zero, there will be little point fitting a moderator(s) at the level of studies, or effect sizes, respectively.

As in multiple regression, we can have multiple (multi-moderator) meta-regression, which can be written as:

$$z_i = \beta_0 + \sum_{h=1}^{q} \beta_h x_{h[i]} + u_{j[i]} + e_i + m_i, \quad (11)$$

where $\sum_{h=1}^{q} \beta_h \, x_{h[i]}$ denotes the sum of all the moderator effects, with q being the number of slopes (staring with $h = 1$). We note that q is not necessarily the number of moderators. This is because when we have a categorical moderator, which is common, with more than two levels (e.g., method A, B & C), the fixed effect part of the formula is $\beta_0 + \beta_1 x_1 + \beta_2 x_2$, where x_1 and x_2 is what is known as 'dummy' variables, which code whether the *i*th effect size belongs to, for example, method B or C, with β_1 and β_2 being contrasts between A and B and between A and C, respectively (for more explanations of dummy variables, see our webpage; also see [54, 55]). Traditionally, researchers conduct separate meta-analyses per different groups (known as 'sub-group analysis'), but we prefer a meta-regression approach with a categorical variable, which is statistically more powerful [40]. Also, importantly, what can be used as a moderator(s) is very flexible, including, for example, individual/plot characteristics (e.g., age, location), environmental factors (e.g., temperature), methodological differences between studies (e.g., randomization), and bibliometric information (e.g., publication year; see more in the section 'Checking for publication bias and robustness'). Note that moderators should be decided *priori*, meaning one's meta-analysis plan and design should include a list of moderators to be considered.

As with meta-analysis, the *Q*-test (Q_m or $Q_{moderator}$) is often used to test the significance of a moderator(s). To complement this test, we can also quantify variance explained by a moderator(s) using R^2. We can define R^2 using Equation 11 as:

$$R^2 = \frac{f^2}{f^2 + \tau^2 + \sigma^2}, \qquad (12)$$

$$f^2 = \mathrm{Var}\left(\sum_{h=1}^{q} \beta_h \, x_{h[i]}\right), \qquad (13)$$

where R^2 is known as marginal R^2 (*sensu* [56, 57]; cf. [58]), f^2 is the variance due to a moderator(s), and $(f^2 + \tau^2 + \sigma^2)$ here equals to $(\tau^2 + \sigma^2)$ in Equation 7, as f^2 'absorbs' variance from τ^2 and/or σ^2. We can compare the similarities and differences in Figure 3B where we denote a

part of f^2 originating from τ^2 as f^2_{study} while σ^2 as f^2_{effect}. In a multiple meta-regression model, we often want to find a model with the 'best' or an adequate set of predictors (i.e., moderators). R^2 can potentially help such a model selection process. Yet, methods based on information criteria (such as AIC) may be preferable. Although model section based on the information criteria is beyond the scope of the paper, we refer the reader to relevant articles (e.g., [59, 60]), and we show an example of this procedure in our online tutorial (link).

Notes on visualisation and interpretation

Visualization and interpretation of results is an essential part of a meta-analysis [61, 62]. Traditionally, a forest plot is used to display the values and 95% of confidence intervals (CIs) for each effect size. and the overall effect and its 95% CI (denoted as a diamond, as shown in Figure 4A). More recently, adding a 95% prediction interval (PI) to the overall estimate has been strongly recommended because 95% PIs show a predicted range of values in which an effect size from a new study would fall, assuming there is no sampling error [63]. Here, we think that examining the formulas for 95% CIs and PIs for the overall mean (from Equation 3) is illuminating:

$$95\%\text{CI} = \beta_0 \pm t_{df[\alpha=0.05]} \cdot se(\beta_0), \quad (14)$$

$$95\%\text{PI} = \beta_0 \pm t_{df[\alpha=0.05]} \sqrt{se^2(\beta_0) + \tau^2 + \sigma^2}, \quad (15)$$

where $t_{df[\alpha=0.05]}$ denotes the *t* value with the degree of freedom, *df*, at 97.5 percentile (or $\alpha = 0.05$) and other notations are as above. In a meta-analysis, it has been conventional to use *z* value 1.96 instead of $t_{df[\alpha=0.05]}$, but simulation studies have shown the use of *t* value over *z* value reduces Type 1 errors under many scenarios and, therefore, is recommended (e.g., [13, 64]). Also, it is interesting to note that by plotting 95% PIs, we can visualize heterogeneity as Equation 15 includes τ^2 and σ^2.

A 'forest' plot can becomes quickly illegible as the number of studies (effect sizes) becomes large, so other methods of visualizing the distribution of effect sizes have been suggested. Some suggested to present a 'caterpillar' plot, which is a version of the forest plot, instead (Figure 4B; e.g., [65]). We here recommend an 'orchard' plot, as it is can present results across different groups (or a result of meta-regression with a categorical variable), as shown in Figure 4C [65]. For visualization of a continuous variable, we suggest what is called a 'bubble' plot, shown in Figure 4D. Visualization not only helps us interpret meta-analytic results, but can also help to identify something we may not see from statistical results, such as influential data points and outliers that could threaten the robustness of our results.

Checking for publication bias and robustness

Detecting and correcting for publication bias

Checking for and adjusting for any publication bias is necessary to ensure the validity of meta-analytic inferences [66]. However, our survey showed almost half of the environmental meta-analyses (46.6%; 34 out of 73 studies) neither tested for nor corrected for publication bias (cf. [14-16]). The most popular methods used were: 1) graphical tests using funnel plots (26 studies; 35.6%), 2) regression-based tests such as Egger regression (18 studies; 24.7%), 3) Fail-safe number tests (12 studies; 16.4%), and 4) trim-and-fill tests (10 studies; 13.7%). We recently showed that these methods are unsuitable for datasets with non-independent effect sizes, with the exception of funnel plots [67] (for an example of funnel plots, see Figure 5A). This is because these methods cannot deal with non-independence in the same way as the fixed-effect and random-effects models. Here, we only introduce a two-step method for multilevel models that can both detect and correct for publication bias [67] (originally proposed by [68, 69]), more specifically, the "small study effect" where an effect size value from a small-sample-sized study can be much larger in magnitude than a 'true' effect [70, 71]. This method is a simple extension of Egger's regression [72], which can be easily implemented by using Equation 10:

$$z_i = \beta_0 + \beta_1 \sqrt{\frac{1}{\tilde{n}_i}} + u_{j[i]} + e_i + m_i, \quad (16)$$

$$z_i = \beta_0 + \beta_1 \left(\frac{1}{\tilde{n}_i}\right) + u_{j[i]} + e_i + m_i, \quad (17)$$

Where $\tilde{n}_i$ is know as effective sample size; for *Zr* and proportion it is just n_i, and for SMD and lnRR, it is $n_{iC}n_{iT}/(n_{iC} + n_{iT})$, as in Table 2. When β_1 is significant, we conclude there exists a small-study effect (in terms of a funnel plot, this is equivalent to significant funnel asymmetry). Then, we fit Equation 17 and then, we look at the intercept β_0, which will be a bias-corrected overall estimate (note that β_0 in Equation 16 provides less accurate estimates when non-zero overall effects exist [68, 69]; Figure 5B). An intuitive explanation of why β_0 (Equation 17) is the bias-corrected estimate is that the intercept represents $1/\tilde{n}_i = 0$ (or $\tilde{n}_i = \infty$); in other words, β_0 is the estimate of the overall effect when we have a very large (infinite) sample size.

Conveniently, this proposed framework can be extended to test for another type of publication bias, known as time-lag bias, or the decline effect, where effect sizes tend to get closer to zero over time, as larger or statistically significant effects are published more quickly than smaller or non-statistically significant effects [73, 74]. Again, a decline effect can be statistically tested by adding year to Equation 3:

$$z_i = \beta_0 + \beta_1 c\left(year_{j[i]}\right) + u_{j[i]} + e_i + m_i, \quad (18)$$

where $c\left(year_{j[i]}\right)$ is the mean-centred publication year of a particular study (study *j* and effect size *i*); this centring makes the intercept β_0 meaningful, representing the overall effect estimate at the mean value of publication years (see [55]). When the slope is significantly different from 0, we deem that we have a decline effect (or time-lag bias; Figure 5C).

However, there may be some confounding moderators, which need to be modelled together. Indeed, Egger's regression (Equations 16 & 17) is known to detect the funnel asymmetry when there is little

heterogeneity; this means that we need to model $\sqrt{1/\tilde{n}_i}$ with other moderators that account for heterogeneity. Given this, we probably should use a multiple meta-regression model, as below:

$$z_i = \beta_0 + \beta_1 \sqrt{\frac{1}{\tilde{n}_i}} + \beta_2 c\left(year_{j[i]}\right) + \sum_{h=3}^{q} \beta_h \, x_{h[i]} + u_{j[i]} + e_i + m_i, \qquad (19)$$

where $\sum_{h=3}^{q} \beta_h \, x_{h[i]}$ is the sum of the other moderator effects apart from the small-study effect and decline effect, and other notations are as above (for more details see [67]). We need to carefully consider which moderators should go into Equation 19 (e.g., fitting all moderators or using an AIC-based model selection method; see [59, 60]). Of relevance, when running complex models, some model parameters cannot be estimated well, or they are not 'identifiable' [75]. This is especially so for variance components (random-effect part) rather than regression coeffects (fixed-effect part). Therefore, it is advisable to check whether model parameters are all identifiable, which can be checked using the *profile* function in *metafor* (for an example, see our tutorial webpage).

Conducting sensitivity analysis & critical appraisal

Sensitivity analysis explores the robustness of meta-analytic results by running a different set of analyses from the original analysis, and comparing the results (note that some consider publication bias tests a part of sensitivity analysis; [11]). For example, we might be interested in assessing how robust results are to the presence of influential studies, to the choice of method for addressing non-independence, or weighting effect sizes. Unfortunately, in our survey, only 37% of environmental meta-analyses (27 out of 73) conducted sensitivity analysis. There are two general and interrelated ways to conduct sensitivity analyses [60, 76, 77]. The first one is to take out influential studies (e.g., outliers) and re-run meta-analytic and meta-regression models. We can also systematically take each effect size out and run a series of meta-analytic models to see whether any resulting overall effect estimates are different from others; this method is known as 'leave-one-out', which is considered less subjective and thus recommended.

The second way of approaching sensitivity analysis is known as subset analysis, where a certain group of effect sizes (studies) will be excluded to re-run the models without this group of effect sizes. For example, one may want to run analysis without studies which did not randomize samples. Yet, as mentioned earlier, we recommend using meta-regression (Equation 13) with a categorical variable of randomization status ('randomized' or 'not randomized'), to statistically test for an influence of moderators. It is important to note that such tests for risk of bias (or study quality) can be considered as a way of quantitatively evaluating the importance of study features that were noted at the stage of critical appraisal, which is essential part of any systematic review (see [11, 78]). In other words, we can use meta-regression or subset analysis to quantitively conduct critical appraisal using (study-level) moderators that code, for example, blinding, randomization, and selective reporting. Despite the importance of critical appraisal ([78]), only 4 of 73 environmental meta-analyses (5.6%) in our survey assessed the risk of bias in each study included in a meta-analysis (i.e., evaluating a primary study in terms of the internal validity of study design and reporting). We emphasize that critically appraising each paper or checking them for risk of bias is an extremely important topic. Also, critical appraisal is not restricted to quantitative synthesis. Therefore, we do not cover any further in this paper (see more for [79, 80])

Notes on transparent reporting & open archiving

For environmental systematic reviews and maps, there are reporting guidelines called RepOrting standards for Systematic Evidence Syntheses in environmental research, ROSES [81] and synthesis assessment checklist, the Collaboration for Environmental Evidence Synthesis Appraisal Tool (CEESAT; [82]). However, these guidelines are somewhat limited in terms of reporting quantitative synthesis because they cover only a few core items. These two guidelines are complemented by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Ecology and Evolutionary Biology (PRISMA-EcoEvo; [83]; cf. [84, 85]), which provides an extended set of reporting items covering what we have described above. Items 20-24 from PRISMA-EcoEvo are most relevant: these items outline what should be reported in the Methods section: i) sample sizes

and study characteristics, ii) meta-analysis, iii) heterogeneity, iv) meta-regression and v) outcomes of publication bias and sensitivity analysis (see Table 4). Our survey, as well as earlier surveys, suggest there is a large room for improvement in the current practice ([14-16]). Incidentally, the orchard plot is well aligned with Item 20, as this plot type shows both the number of effect sizes and studies for different groups (Figure 4C). Further, our survey of environmental meta-analyses highlighted the poor standards of data openness (with 24 studies sharing data: 32.9%) and code sharing (7 studies: 29.2%). Environmental scientists must archive their data as well as their analysis code in accordance with the FAIR principles (Findable, Accessible, Interoperable, and Reusable; [86]) using dedicated depositories such as Dryad, FigShare, Open Science Framework (OSF), Zenodo or others (cf. [87, 88]), preferably not on the publisher's webpage (as paywall may block access). However, archiving itself is not enough; data requires metadata (detailed descriptions) and the code needs to also be FAIR [89, 90].

Other relevant and advanced issues

Scale dependence

The issue of scale dependence is a unique yet widespread problem in environmental sciences (see [7, 91]); our literature survey indicated three quarters of the environmental meta-analyses (56 out of 73 studies) have inferences that are potentially vulnerable to scale-dependence [92]. For example, studies that set out to compare group means in biodiversity measures, such as species richness, can vary as a function of the scale (size) of the sampling unit. When the unit of replication is a plot (not an individual animal or plant), the aerial size of a plot (e.g., 100 cm^2 or 1 km^2) will affect both the precision and accuracy of effect size estimates (e.g., lnRR and SMD). In general, a study with larger plots might have more accurately estimated species richness differences, but less precisely than a study with smaller plots and greater replication. Lower replication means that our sampling variance estimates are likely to be misestimated, and the study with larger plots will generally have less weight than the study with smaller plots, due to a higher sampling variance. Inaccurate variance

estimates in little-replicated ecological studies are known to cause an accumulating bias in precision-weighted meta-analysis, requiring correction [93]. To assess the potential for scale-dependence, it is recommended that analysts test for possible for covariation among plot size, replication, variances, and effect sizes [91]. If detected, analysts should use an effect size statistic that is less sensitive to scale dependence (lnRR), and could use the size of a plot as a moderator in meta-regression and consider, or we could run an unweighted model ([7]; note that only 12%, 9 out of 73 studies, accounted sampling area in some way).

Missing data

In many fields, meta-analytic data almost always encompass missing values see [94-96]. Broadly, we have two types of missing data in meta-analyses [97, 98]: 1) missing data in standard deviations or sample sizes, associated with means, preventing effect size calculations (Table 2), and 2) missing data in moderators. There are several solutions for both types. The best, and first to try, should be contacting the authors. If this fails, we can potentially 'impute' missing data. Single imputation methods using the strong correlation between standard deviation and mean values (known as mean-variance relationship) are available, although single imputation can lead to Type I error [94, 95] (see also [93]) because we do not model the uncertainty of imputation itself. Contrastingly, multiple imputation, which creates multiple versions of imputed datasets, incorporates such uncertainty. Indeed, multiple imputation is a preferred and proven solution for missing data in effect sizes and moderators [97, 98]. Yet, correct implementation can be challenging (see [98]). What we require now is an automated pipeline of merging meta-analysis and multiple imputation, which accounts for imputation uncertainty, although it may be challenging for complex meta-analytic models. Fortunately, however, for lnRR, there is a series of new methods that can perform better than the conventional method and which can deal with missing SDs [99]; note that these methods do not deal with missing moderators. Therefore, where applicable, we recommend these new methods, until an easy-to-implement multiple imputation workflow arrives.

Complex non-independence

Above, we have only dealt with the model that include study identities as a clustering/grouping (random) factor. However, many datasets are more complex, with potentially more clustering variables in addition to the study (paper) identity. It is certainly possible that an environmental meta-analysis contains data from multiple species. Such a situation creates an interesting dependence among effect sizes from different species, known as phylogenetic relatedness, where closely related species are more likely to be similar in effect sizes compared to distantly related ones (e.g., mice *vs.* rats and mice *vs.* sparrows). Our multilevel model framework is flexible and can accommodate phylogenetic relatedness. A phylogenetic multilevel meta-analytic model can be written as [40, 100, 101]:

$$z_i = \beta_0 + a_{k[i]} + s_{k[i]} + u_{j[i]} + e_i + m_i, \quad (20)$$

$$a_k \sim \mathrm{N}(0, \omega^2 \mathrm{A}),\ \ s_k \sim \mathrm{N}(0, \gamma^2),\ \ u_j \sim \mathrm{N}(0, \tau^2),\ \ e_i \sim \mathrm{N}(0, \sigma^2), \&\ \ m_i \sim \mathrm{N}(0, v_i),$$

where $a_{k[i]}$ is the phylogenetic (species) effect for the *k*th species (effect size *i*; N_{effect} (i = 1, 2,…, N_{effect}) > N_{study} (j = 1, 2,…, N_{study}) > $N_{species}$ (k = 1, 2,…, $N_{species}$)), normally distributed with $\omega^2\mathbf{A}$ where is the phylogenetic variance and **A** is a correlation matrix coding how close each species are to each other and ω^2 is the phylogenetic variance, $s_{k[i]}$ is the non-phylogenetic (species) effect for the *k*th species (effect size *i*), normally distributed with the variance of γ^2 (the non-phylogenetic variance), and other notations are as above. It is important to realize that **A** explicitly models relatedness among species, and we do need to provide this correlation matrix, using a distance relationship usually derived from a molecular-based phylogenetic tree (for more details, see [40, 100, 101]). Some may think that the non-phylogenetic term ($s_{k[i]}$) is unnecessary or redundant because $s_{k[i]}$ and the phylogenetic term ($a_{k[i]}$) are both modelling variance at the species level. However, a simulation recently demonstrated that failing to have the non-phylogenetic term ($s_{k[i]}$) will often inflate the phylogenetic variance ω^2, leading to an incorrect conclusion that there is a strong phylogenetic signal (as shown in [101]). The non-phylogenetic variance (γ^2) arises from, for

example, ecological similarities among species (herbivores *vs.* carnivores or arboreal *vs.* ground-living) not phylogeny [40].

Like phylogenetic relatedness, effect sizes arising from closer geographical locations are likely to be more correlated [102]. Statistically, spatial correlation can be also modelled in a manner analogous to phylogenetic relatedness (i.e., rather than a phylogenetic correlation matrix, **A**, we fit a spatial correlation matrix). For example, Maire and colleagues [103] used a meta-analytic model with spatial autocorrelation to investigate the temporal trends of fish communities in the network of rivers in France. We note that a similar argument can be made for temporal correlation, but in many cases, temporal correlation could be dealt with, albeit less accurately, as a special case of 'shared measurements', as in Figure 2. An important idea to take away is that one can model different, if not all, types of non-independence as a random factor(s) in a multilevel model.

Advanced techniques

Here we touch upon five advanced meta-analytic techniques with potential utility for environmental sciences, providing relevant references so that interested readers can obtain more information on these advanced topics. The first one is the meta-analysis of magnitudes, or absolute values (effect sizes), where researchers may be interested in deviations from 0, rather than the directionality of the effect [104]. For example, Cohen and colleagues [105] investigated absolute values of phenological responses, as they were concerned with the magnitudes of changes in phenology rather than directionality.

The second method is the meta-analysis of interaction where our focus is on synthesizing the interaction effect of, usually, 2 ×2 factorial design (e.g., the effect of two simultaneous environmental stressors; [44, 106, 107]). Recently, Siviter and colleagues [108] showed that agrochemicals interact synergistically (i.e., non-additively) to increase the mortality of bees; that is,

two agrochemicals together caused more mortality than the sum of mortalities of each chemical (i.e., additive effect).

Third, network meta-analysis has been heavily used in medical sciences; network meta-analysis usually compares different treatments in relation to placebo and ranks these treatments in terms of effectiveness [109]. The very first 'environmental' network meta-analysis, as far as we know, investigated the effectives of ecosystem services among different land types [110].

Fourth, a multivariate meta-analysis is where one can model two or more different types of effect sizes with the estimation of pair-wise correlations between different effect sizes. The benefit of such an approach is known as the 'borrowing of strength', where the error of fixed effects (moderators; e.g., b_0 and b_1) can be reduced when different types of effect sizes are correlated (i.e., $se(b_0)$ and $se(b_1)$ can be smaller [111]) For example, it is possible for lnRR (differences in mean) and lnVR (differences in SDs) to be modelled together (cf. [112]).

Fifth, as with network meta-analysis, there has been a surge in the use of 'individual participants data', called 'IPD meta-analysis', in medical sciences [113, 114]. The idea of IPD meta-analysis is simple – rather than using summary statistics reported in papers (sample means and variances), we directly use raw data from all studies. We can either model raw data using one complex multilevel (hierarchical) model (one-step method) or calculate statistics for each study and use a meta-analysis (two-step method; note that both methods will usually give the same results). Study-level random effects can be incorporated to allow the response variable of interest to vary among studies, and overall effects correspond to fixed, population-level estimates. The use of IPD or 'full-data analyses' has also surged in ecology, aided by open-science policies that encourage the archival of raw data alongside articles, and initiatives that synthesise raw data (e.g., PREDICTS [115], BioTime [116]). In health disciplines, such meta-analyses are considered the 'gold standard' [117],

584 owing to their potential for resolving issues regarding study-specific designs and confounding
585 variation, and it is unclear whether and how they might resolve issues such as scale dependence in
586 environmental meta-analyses [91, 118].

587 ## Conclusions

588 In this article, we have attempted to describe the most practical ways to conduct quantitative
589 synthesis, including meta-analysis, meta-regression, and publication bias tests. In addition, we have
590 shown that there is much to be improved in terms of meta-analytic practice and reporting via a
591 survey of 73 recent environmental meta-analyses. Such improvements are urgently required,
592 especially given the potentially influence that environmental meta-analyses can have for policies
593 and decision making [8]. So often, meta-analysts have called for better reporting of primary
594 research (e.g., [119, 120]), and now this is the time to raise the standards of reporting in meta-
595 analyses. We hope our contribution will helps to catalyse a turning point for better practice in
596 quantitative synthesis in environmental sciences. We remind the reader most of what is described is
597 implemented in the *R* environment at our tutorial webpage and researchers can readily use the
598 proposed models and techniques (link). Finally, meta-analytic techniques are always developing
599 and improving. It is certainly possible that in the future, our proposed models and related methods
600 will become dated, just as the traditional fixed-effect and random-effects models already are.
601 Therefore, we must endeavour to be open-minded to new ways of doing quantitative research
602 synthesis in environmental sciences.

603 ## Acknowledgements

604 SN, ELM, and ML were supported by the ARC (Australian Research Council) Discovery grant
605 (DP200100367), and SN, YY, and ML by the ARC Discovery grant (DP210100812). YY was also
606 supported by the National Natural Science Foundation of China (32102597).

Author contributions

SN was commissioned to write this article, so that he assembled a team of co-authors. SN discussed the idea with YY, ELM, RS and ML, and all of them contributed to the design of this review. ML led the survey working with YY and ELM, while YY led the creation of the accompanying webpage working with RS. SN supervised all aspects of this work and wrote the first draft, which was commented on, edited, and therefore, significantly improved by the other co-authors.

Conflict of interest

The author reported no conflict of interest

ORCID

Shinichi Nakagawa: 0000-0002-7765-5182

Yefeng Yang: 0000-0002-8610-4016

Rebecca Spake: 0000-0003-4671-2225

Erin Macartney: 0000-0003-3866-143X

Malgorzata Lagisz: 0000-0002-3993-6127

Figure legends

Fig. 1 Visualisation of the three statistical models of meta-analysis: A) a fixed-effect model (1-level), B) a random-effects model (2-level), and C) a multilevel model (3-level; see the text for what symbols mean).

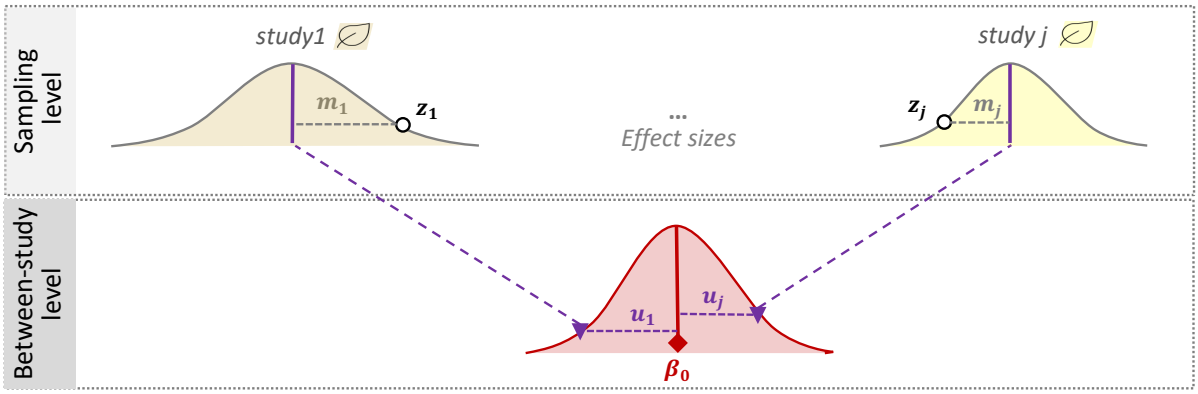
Fig. 2 Visualisation of the three types of non-independence among effect sizes: 1) due to shared study identities (effect sizes from the same study may be similar in values), 2) due to shared measurements (effect sizes come from the same group of individuals/plots but are based on

631 different types of measurements), and 3) due to shared control (effect sizes are calculated using the
632 same control group and multiple treatment groups; see the text for more details).
633
634 **Fig. 3** Visualisation of variation (heterogeneity) partitioned into different variance components: A)
635 quantifying different types of I^2 from a multilevel model (3-level; see Fig. 1C) and B) variance
636 explained, R^2, by moderators. Note that different levels of variances would be explained, depending
637 on which level a moderator belongs to (study level and effect-size level).
638
639 **Fig. 4** Different types of plots useful for a meta-analysis using data from Midolo et al. [121]: A) a
640 typical forest plot with the overall mean shown as a diamond at the bottom (20 effect sizes from 20
641 studies are used), B) a caterpillar plot (100 effect sizes from 24 studies are used), C) an orchard plot
642 of categorical moderator with seven levels (all effect sizes are used), and D) a bubble plot of a
643 continuous moderator. Note that the first two only show confidence intervals, while the latter two
644 also show prediction intervals (see the text for more details).
645
646 **Fig. 5** Different types of plots for publication bias tests: A) a funnel plot using model residuals,
647 showing a funnel (white) that shows the region of statistical non-significance (30 effect sizes from
648 30 studies are used), B) a bubble plot visualising a multilevel meta-regression that tests for the
649 small study effect (note that the slope was non-significant: β = 0.120, 95%CI = [-0.095, 0.334]; all
650 effect sizes are used), and C) a bubble plot visualising a multilevel meta-regression that tests for the
651 decline effect (the slope was non-significant: β = 0.003, 95%CI = [-0.002, 0.008]).
652
653

Fig 1

A Fixed-effect model: $z_j = \beta_0 + m_j$

B Random-effects model: $z_j = \beta_0 + u_j + m_j$



C Multilevel model: $z_i = \beta_0 + u_{j[i]} + m_i + e_i$

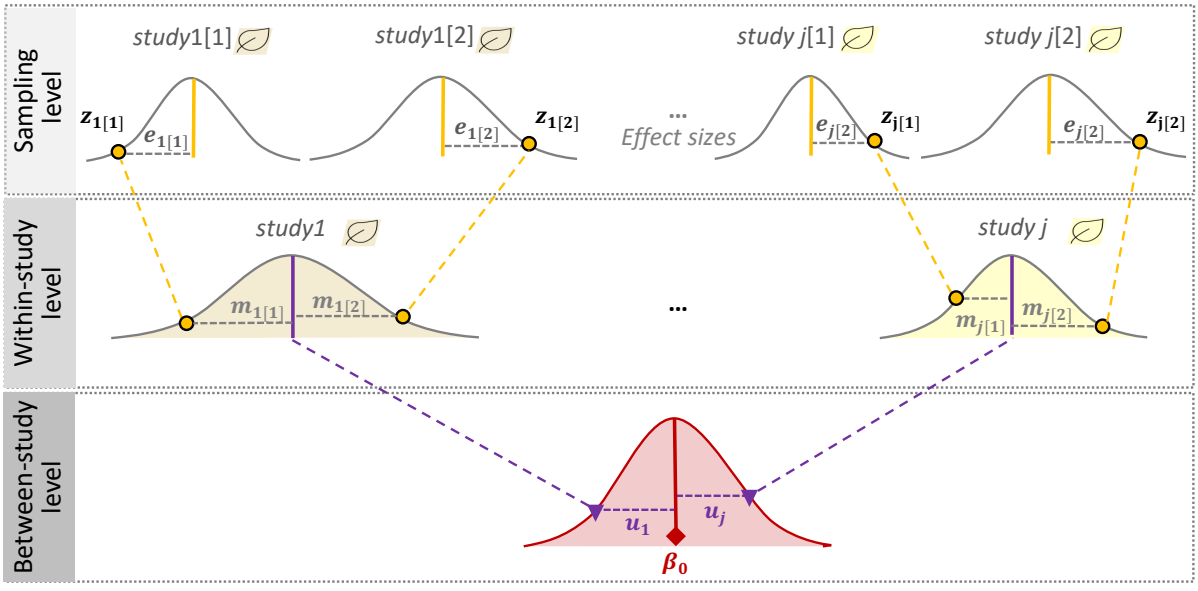


Fig 2

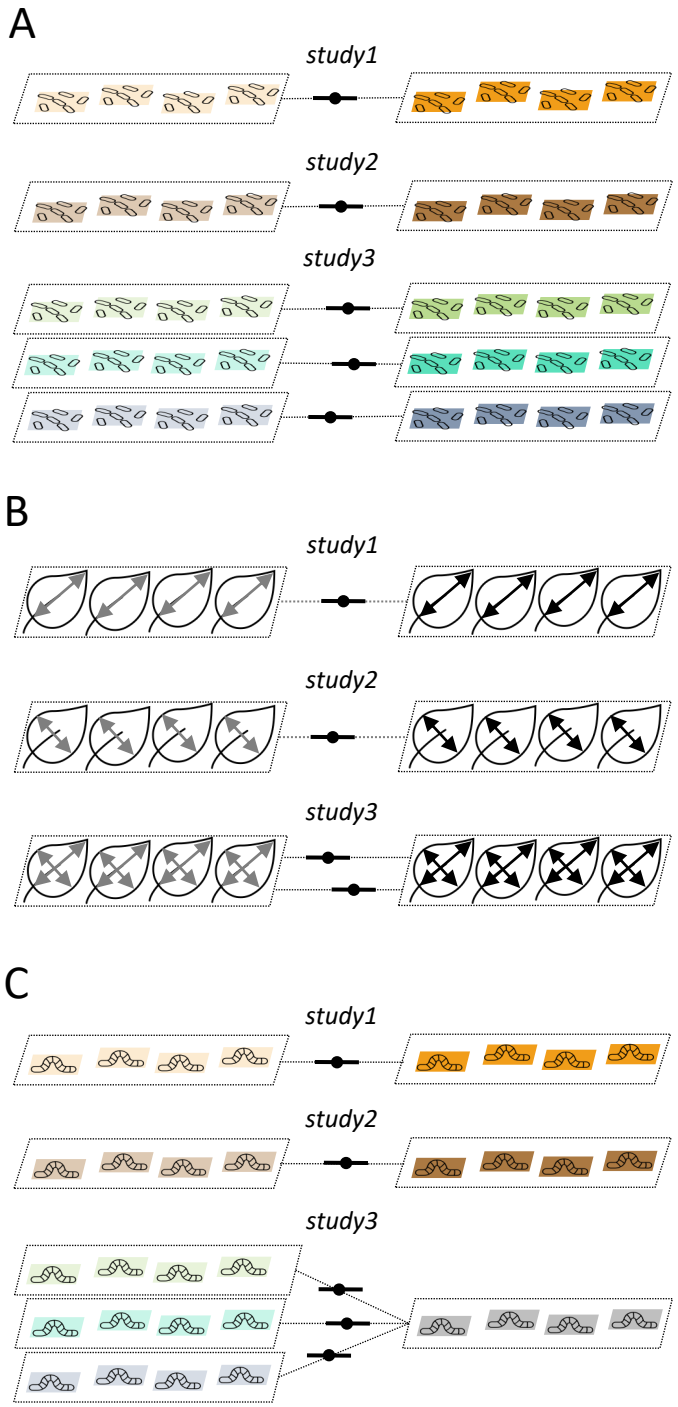
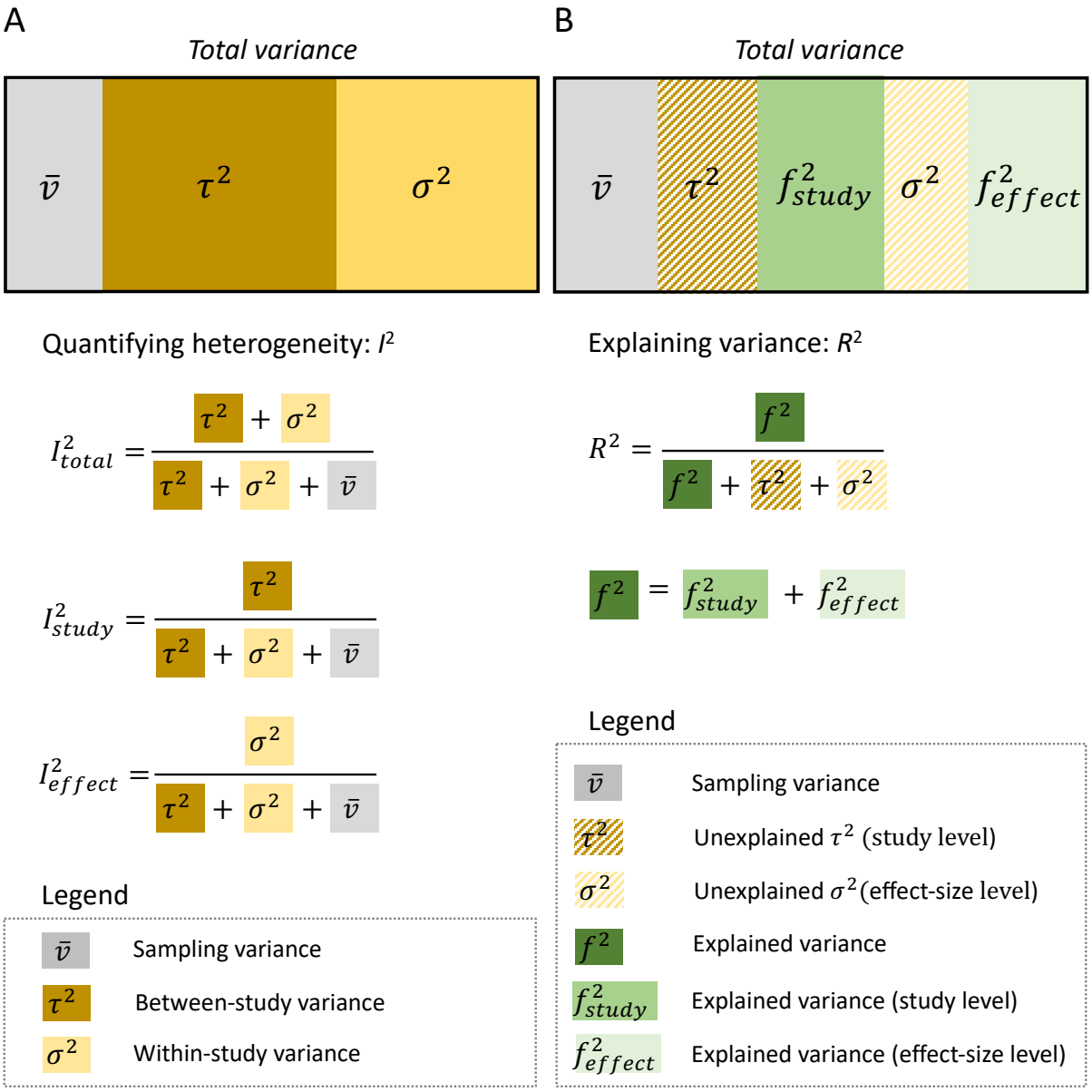
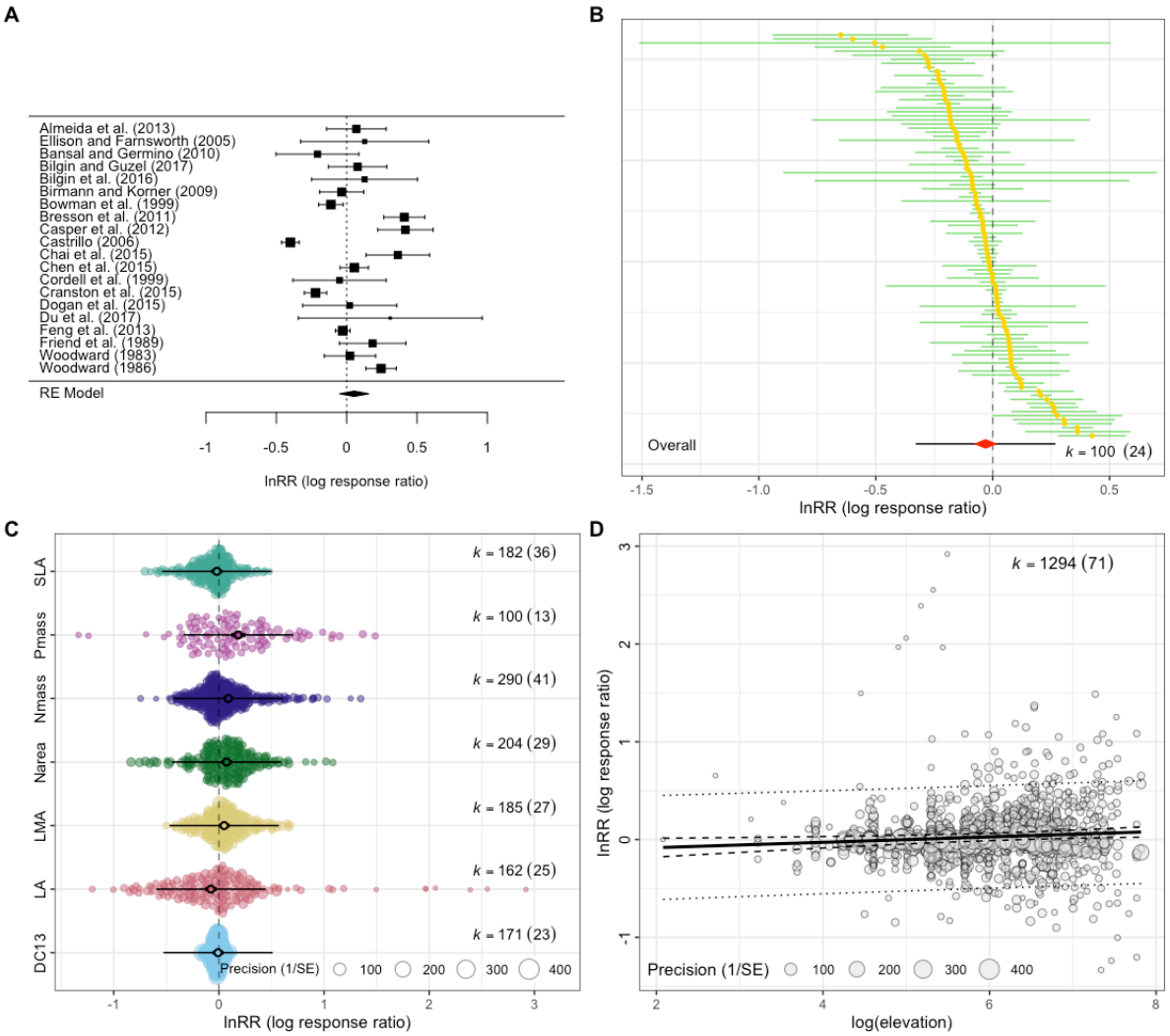

A
study1
study2
study3
B
study1
study2
study3
C
study1
study2
study3

Fig 3

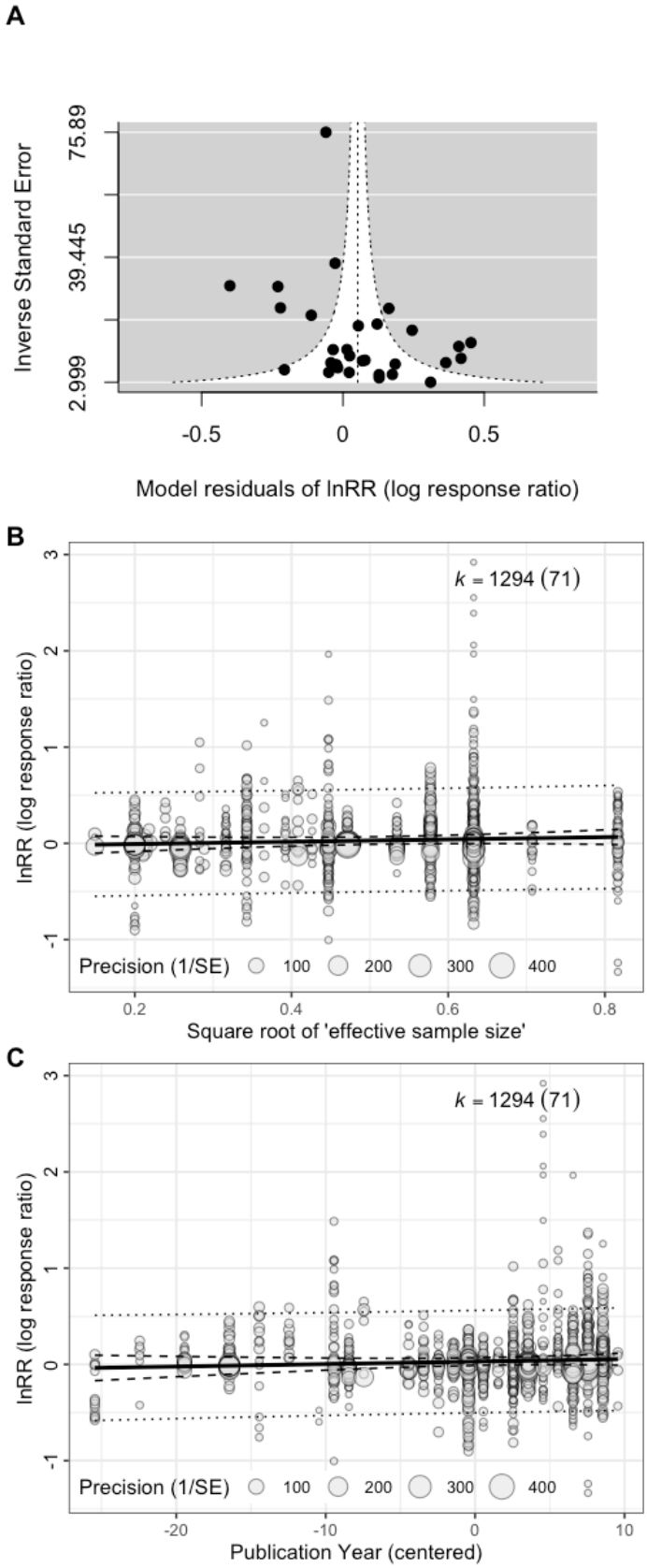

A
Total variance
$\bar{v}$
τ^2
σ^2
B
Total variance
$\bar{v}$
τ^2
f^2_{study}
σ^2
f^2_{effect}
Quantifying heterogeneity: I^2
$I^2_{total} = \frac{\tau^2 + \sigma^2}{\tau^2 + \sigma^2 + \bar{v}}$
$I^2_{study} = \frac{\tau^2}{\tau^2 + \sigma^2 + \bar{v}}$
$I^2_{effect} = \frac{\sigma^2}{\tau^2 + \sigma^2 + \bar{v}}$
Legend
$\bar{v}$ Sampling variance
τ^2 Between-study variance
σ^2 Within-study variance
Explaining variance: R^2
$R^2 = \frac{f^2}{f^2 + \tau^2 + \sigma^2}$
$f^2 = f^2_{study} + f^2_{effect}$
Legend
$\bar{v}$ Sampling variance
τ^2 Unexplained τ^2 (study level)
σ^2 Unexplained σ^2 (effect-size level)
f^2 Explained variance
f^2_{study} Explained variance (study level)
f^2_{effect} Explained variance (effect-size level)

667 **Fig 4**
668



669
670
671

672 **Fig 5**
673



674
675
676
677

678 # Tables

679 **Table 1 Definitions of key concepts and associated statistical parameters, which are used in formulas in the main text**

Term	Definition (with associated parameters, if any)
Effect size	A measurement of effect (usually state of a single group, comparison between groups, or association, see Table 2). In a meta-analytic model, it becomes the response variable (denoted as z_i in the formulas).
Sampling variance	A measure of uncertainty in effect size (denoted as v_i). Its inverse is often called 'weight' (the square-root of weight is 'precision', and the square root of sampling variance is 'sampling standard error').
Meta-analysis	A statistical method to aggregate effect sizes from studies on the same or similar topics, by assigning different weights basing on sampling variance of effect sizes. Strictly speaking, in a formal (weighted) meta-analysis, sampling variance needs to be incorporated and it is assumed to be known (Table 2).
Overall mean (effect)	An average effect size based on a meta-analytic model (denoted as β_0 and its standard errors se(β_0)).
Heterogeneity	An indicator of consistency among effect sizes, or an extent of variation around the overall effect (β_0); heterogeneity can be quantified by absolute measures, such as τ^2, or relative measures, such as I^2.

Meta-regression	A regression model which extends a meta-analytic model with a moderator(s), aiming to explain heterogeneity (quantified as R^2) and quantifying the effect of a moderator (denoted as, for example, β_1).
Publication bias tests	A set of statistical methodologies to detect and correct for publication bias, where a subset of results (positive findings) is more likely to be published and present in the meta-analytic dataset than otherwise.
Sensitivity analysis	A set of statistical analyses that checks the robustness of one's main analysis; if sensitivity analysis shows different results (qualitatively and/or quantitively), then we must doubt the robustness of the main findings.

682 **Table 2 Selected list of effect size statistics and their sampling variances, belonging to three types: 1) single-group effect, 2) comparative effect**

683 **and 3) association effect**

Type	Statistic	Estimate	Sampling variance	Reference
Single group	Mean	$\bar{x}_i$	s_i^2/n_i	[122]
Single group	Proportion	$p_i = \frac{y_i}{n_i}$	$\frac{p_i(1-p_i)}{n_i} = \frac{y_i(n_i-y_i)}{n_i^3}$	[122]
Single group	Log standard deviation (lnSD)	$\ln s_i$	$\frac{1}{2(n_i-1)}$	[27]
Single group	Log coefficient of variation (lnCV)	$\ln\left(\frac{s_i}{\bar{x}_i}\right)$	$\frac{s_i^2}{n_i\bar{x}_i^2} + \frac{1}{2(n_i-1)}$	[27]
Comparative	Mean difference (MD)	$\bar{x}_{iT} - \bar{x}_{iC}$	$\frac{s_{iC}^2}{n_{iC}} + \frac{s_{iT}^2}{n_{iT}}$	[122]
Comparative	Standardised mean difference (SMD)	$d_i = \frac{\bar{x}_{iT} - \bar{x}_{iC}}{\sqrt{\frac{(n_{iC}-1)s_{iC}^2 + (n_{iT}-1)s_{iT}^2}{n_{iC}+n_{iT}-2}}}$	$\frac{1}{n_{iC}} + \frac{1}{n_{iT}} + \frac{d_i^2}{2(n_{iC}+n_{iT})}$	[25]
Comparative	Risk (proportion) difference (RD)	$\frac{y_{iT}}{n_{iT}} - \frac{y_{iC}}{n_{iC}}$	$\frac{y_{iT}(n_{iT}-y_{iT})}{n_{iT}^3} + \frac{y_{iC}(n_{iC}-y_{iC})}{n_{iC}^3}$	[122]

Comparative	Log odds ratio (lnOR)	$\ln\left(\frac{y_{iT}}{n_{iT}-y_{iT}}\right)-\ln\left(\frac{y_{iC}}{n_{iC}-y_{iC}}\right)$	$\frac{1}{y_{iT}}+\frac{1}{n_{iT}-y_{iT}}+\frac{1}{y_{iC}}+\frac{1}{n_{iC}-y_{iC}}$	[122]
Comparative	Log response ratio (lnRR)	$\ln\left(\frac{\bar{x}_{iT}}{\bar{x}_{iC}}\right)$	$\frac{s_{iC}^2}{n_{iC}\bar{x}_{iC}^2}+\frac{s_{iT}^2}{n_{iT}\bar{x}_{iT}^2}$	[123]
Comparative	Log variability ratio (lnVR)	$\ln\left(\frac{s_{iT}}{s_{iC}}\right)$	$\frac{1}{2(n_{iC}-1)}+\frac{1}{2(n_{iT}-1)}$	[27]
Comparative	Log coefficient of variation ratio (lnCVR)	$\ln\left(\frac{s_{iT}}{\bar{x}_{iT}}\right)-\ln\left(\frac{s_{iC}}{\bar{x}_{iC}}\right)$	$\frac{s_{iC}^2}{n_{iC}\bar{x}_{iC}^2}+\frac{1}{2(n_{iC}-1)}+\frac{s_{iT}^2}{n_{iT}\bar{x}_{iT}^2}+\frac{1}{2(n_{iT}-1)}$	[27]
Association	Fisher's z-transformation of correlation, *r* (*Zr*)	$\frac{1}{2}\ln\left(\frac{1+r_i}{1-r_i}\right)$	$\frac{1}{n_i-3}$	[122]

For the column 3rd and 4th, notations represent: $\bar{x}$ (mean), *s* (standard deviation), *n* (sampling size), *y* (the number of events), the subscript *T* (treatment group), the subscript *C* (control group) and the subscript *i* (the *i*th effect size or study). Note that better estimators may be found in the relevant references; for example, SMD can be best estimated by multiplying by $\left(1-\frac{3}{4(n_{iC}+n_{iT}-2)-1}\right)$, and see also [93].

688 **Table 3 Examples of dependence between two sampling variances (v_1 and v_2) and their covariance for four common effect size statistics**

Statistic	Situation	Variances	Covariance
Proportion	Shared measurement	$v_1 = \frac{y_1(n_1 - y_1)}{n_1^3}$ $v_2 = \frac{y_2(n_2 - y_2)}{n_2^3}$	$\rho\sqrt{\frac{y_1(n_1 - y_1)}{n_1^3}\frac{y_2(n_2 - y_2)}{n_2^3}}$
Zr	Shared measurement	$v_1 = \frac{1}{2}\ln\left(\frac{1 + r_1}{1 - r_1}\right)$ $v_2 = \frac{1}{2}\ln\left(\frac{1 + r_2}{1 - r_2}\right)$	$\rho\sqrt{\frac{1}{4}\ln\left(\frac{1 + r_1}{1 - r_1}\right)\ln\left(\frac{1 + r_2}{1 - r_2}\right)}$
lnRR	Shared measurement	$v_1 = \frac{s_{1C}^2}{n_{1C}\bar{x}_{1C}^2} + \frac{s_{1T}^2}{n_{1T}\bar{x}_{1T}^2}$ $v_2 = \frac{s_{2C}^2}{n_{2C} \cdot \bar{x}_{2C}^2} + \frac{s_{2T}^2}{n_{2T} \cdot \bar{x}_{2T}^2}$	$\rho\sqrt{\left(\frac{s_{1C}^2}{n_{1C}\bar{x}_{1C}^2} + \frac{s_{1T}^2}{n_{1T}\bar{x}_{1T}^2}\right)\left(\frac{s_{2C}^2}{n_{2C} \cdot \bar{x}_{2C}^2} + \frac{s_{2T}^2}{n_{2T} \cdot \bar{x}_{2T}^2}\right)}$
	Shared control	$v_1 = \frac{s_{1C}^2}{n_{1C}\bar{x}_{1C}^2} + \frac{s_{1T}^2}{n_{1T}\bar{x}_{1T}^2}$	$\frac{s_{1C}^2}{n_{1C} \cdot \bar{x}_{1C}^2}$

		$v_2 = \frac{s_{1C}^2}{n_{1C} \cdot \bar{x}_{1C}^2} + \frac{s_{2T}^2}{n_{2T} \cdot \bar{x}_{2T}^2}$		
SMD	Shared measurement	$v_1 = \frac{1}{n_{1C}} + \frac{1}{n_{1T}} + \frac{d_1^2}{2(n_{1C} + n_{1T})}$ $v_2 = \frac{1}{n_{2C}} + \frac{1}{n_{2T}} + \frac{d_1^2}{2(n_{2C} + n_{2T})}$	$\rho\sqrt{\left(\frac{1}{n_{1C}} + \frac{1}{n_{1T}} + \frac{d_1^2}{2(n_{1C} + n_{1T})}\right)\left(\frac{1}{n_{2C}} + \frac{1}{n_{2T}} + \frac{d_2^2}{2(n_{2C} + n_{2T})}\right)}$	
	Shared control	$v_1 = \frac{1}{n_{1C}} + \frac{1}{n_{1T}} + \frac{d_1^2}{2(n_{1C} + n_{1T} + n_{2T})}$ $v_2 = \frac{1}{n_{1C}} + \frac{1}{n_{2T}} + \frac{d_2^2}{2(n_{1C} + n_{1T} + n_{2T})}$	$\frac{1}{n_{1C}} + \frac{d_1 d_2}{2(n_{1C} + n_{1T} + n_{2T})}$	

For the 2nd column, see Figure 2. For the 3rd and 4th column, notations represent: the subscript 1*C* and 2*C* (control group for 1st and 2nd effect size, respectively, but for shared control, 1*C* is used for both effect sizes, but 1*C* and 2*C* are the same cohort or set of plots), the subscript *1T* and 2*T*(treatment group for the 1st and 2nd effect size, respectively; for shared groups, 1T and 2T represents different groups of individuals/plots whereas for shared measurements, 1T and 2T are the same set of individuals/plots), and the other notations are as in Table 1 and the main text (see also [44, 45]).

693 **Table 4 Items relevant to reporting results for a meta-analysis from the Preferred Reporting Items for Systematic reviews and Meta-Analysis**

694 **for Ecology and Evolutionary Biology (PRISMA-EcoEvo; [83])**

Item	Description
20: Sample sizes and study characteristics	"Report the number of studies and effect size for data included in meta-analyses and subsets of data included in meta-regressions. Provide a summary of kye characteristics for reported outcomes (either in text or figures; e.g., one quarter of effect sizes reported for vertebrates and the rest invertebrates) and their limitations (e.g., collinearity and overlaps between moderators), including characteristics related individual study quality (risk of bias)."
21: Meta-analysis	"Provide a quantitative synthesis of results across studies, including estimates for the main effect size, with confidence/credible intervals."
22: Heterogeneity	"Report indicators of heterogeneity in the estimated effect (e.g. I^2, tau^2 and other variance components)."
23: Meta-regression	"Provide estimates of meta-regression slopes (i.e. regression coefficients) for all variables that were assessed for their contribution to heterogeneity. Include confidence/credible intervals, and report interactions if they were included. Describe outcomes from model selection, if done (e.g. R^2 and AIC)."
24: Outcomes of publication bias and sensitivity analysis	"Provide results for the assessments of the risks of bias (e.g. Egger's regression, funnel plots) and robustness of the review's results (e.g. subgroup analyses, meta-regression of study quality, results from alternative methods of analysis, and temporal trends)"

695

REFERENCES

1. Higgins JPTe, Thomas Je, Chandler Je, Cumpston Me, Li Te, Page MJe, Welch VAe: **Cochrane handbook for systematic reviews of interventions**, Second edition. edn. Chichester, UK: Wikey; 2019.
2. Cooper HM, Hedges LV, Valentine JC: **The handbook of research synthesis and meta-analysis**, 3rd ed. edn. New York: Russell Sage Foundation; 2019.
3. Schmid CHe, Stijnen Te, White Ie: **Handbook of meta-analysis**, 1st edn. Boca Ranton: CRC; 2021.
4. Vetter D, Rucker G, Storch I: **Meta-analysis: A need for well-defined usage in ecology and conservation biology**. *Ecosphere* 2013, **4**(6).
5. Koricheva J, Gurevitch J, Mengersen K (eds.): **Handbook of meta-analysis in ecology and evolution**. Princeton: Princeton Univesity Press; 2017.
6. Gurevitch J, Koricheva J, Nakagawa S, Stewart G: **Meta-analysis and the science of research synthesis**. *Nature* 2018, **555**(7695):175-182.
7. Spake R, Doncaster CP: **Use of meta-analysis in forest biodiversity research: key challenges and considerations**. *Forest Ecol Manag* 2017, **400**:429-437.
8. Bilotta GS, Milner AM, Boyd I: **On the use of systematic reviews to inform environmental policies**. *Environ Sci Policy* 2014, **42**:67-77.
9. Hedges LV, Vevea JL: **Fixed- and random-effects models in meta-analysis**. *Psychol Methods* 1998, **3**(4):486-504.
10. Borenstein M, Hedges LV, Higgins JPT, Rothstein H: **Introduction to meta-analysis**, Second edition. edn. Chichester: Wiley; 2021.
11. Noble DWA, Lagisz M, O'dea RE, Nakagawa S: **Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses**. *Mol Ecol* 2017, **26**(9):2410-2425.
12. Nakagawa S, Noble DWA, Senior AM, Lagisz M: **Meta-evaluation of meta-analysis: ten appraisal questions for biologists**. *Bmc Biol* 2017, **15**.
13. Nakagawa S, Senior AM, Viechtbauer W, Noble DWA: **An assessment of statistical methods for nonindependent data in ecological meta-analyses: Comment**. *Ecology* 2022, **103**(1):e03490.
14. Romanelli JP, Meli P, Naves RP, Alves MC, Rodrigues RR: **Reliability of evidence-review methods in restoration ecology**. *Conserv Biol* 2021, **35**(1):142-154.
15. Koricheva J, Gurevitch J: **Uses and misuses of meta-analysis in plant ecology**. *J Ecol* 2014, **102**(4):828-844.
16. O'Leary BC, Kvist K, Bayliss HR, Derroire G, Healey JR, Hughes K, Kleinschroth F, Sciberras M, Woodcock P, Pullin AS: **The reliability of evidence review methodology in environmental science and conservation**. *Environ Sci Policy* 2016, **64**:75-82.
17. Rosenthal R: **The "file drawer problem" and tolerance for null results**. *Psychol Bull* 1979, **86**(3):638-641.
18. Nakagawa S, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, Sánchez-Tójar A, Yang Y, O'Dea RE: **Methods for testing publication bias in ecological and evolutionary meta-analyses**. *Methods in Ecology and Evolution* 2022, **13**(1):4-21.
19. Cheung MWL: **A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes**. *Neuropsychol Rev* 2019, **29**(4):387-396.
20. Viechtbauer W: **Conducting Meta-Analyses in R with the metafor Package**. *J Stat Softw* 2010, **36**(3):1-48.
21. Yang Y, Macleod M, Pan J, Lagisz M, Nakagawa S: **Advanced Methods and Implementations for the Meta-analyses of Animal Models: Current Practices and**

Future Recommendations. *Neuroscience & Biobehavioral Reviews* 2022, **DOI: 10.1016/j.neubiorev.2022.105016**:105016.
22. Nakagawa S, Cuthill IC: **Effect size, confidence interval and statistical significance: a practical guide for biologists**. *Biol Rev* 2007, **82**(4):591-605.
23. Hedges LV, Gurevitch J, Curtis PS: **The meta-analysis of response ratios in experimental ecology**. *Ecology* 1999, **80**(4):1150-1156.
24. Friedrich JO, Adhikari NKJ, Beyene J: **The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study**. *Bmc Med Res Methodol* 2008, **8**.
25. Hedges L, Olkin I: **Statistical methods for meta-analysis**. New York: Academic Press; 1985.
26. Cohen J: **Statistical power analysis for the beahvioral sciences**, 2nd edn. Hillsdale, New Jersey: Lawrence Erlbaum; 1988.
27. Senior AM, Viechtbauer W, Nakagawa S: **Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio**. *Res Synth Methods* 2020, **11**(4):553-567.
28. Nakagawa S, Poulin R, Mengersen K, Reinhold K, Engqvist L, Lagisz M, Senior AM: **Meta-analysis of variation: ecological and evolutionary applications and beyond**. *Methods in Ecology and Evolution* 2015, **6**(2):143-152.
29. Knapp S, van der Heijden MGA: **A global meta-analysis of yield stability in organic and conservation agriculture**. *Nat Commun* 2018, **9**.
30. Porturas LD, Anneberg TJ, Cure AE, Wang SP, Althoff DM, Segraves KA: **A meta-analysis of whole genome duplication and theeffects on flowering traits in plants**. *Am J Bot* 2019, **106**(3):469-476.
31. Janicke T, Morrow EH: **Operational sex ratio predicts the opportunity and direction of sexual selection across animals**. *Ecol Lett* 2018, **21**(3):384-391.
32. Chamberlain R, Brunswick N, Siev J, McManus IC: **Meta-analytic findings reveal lower means but higher variances in visuospatial ability in dyslexia**. *Brit J Psychol* 2018, **109**(4):897-916.
33. O'Dea RE, Lagisz M, Jennions MD, Nakagawa S: **Gender differences in individual variation in academic grades fail to fit expected patterns for STEM**. *Nat Commun* 2018, **9**.
34. Brugger SP, Angelescu I, Abi-Dargham A, Mizrahi R, Shahrezaei V, Howes OD: **Heterogeneity of Striatal Dopamine Function in Schizophrenia: Meta-analysis of Variance**. *Biol Psychiat* 2020, **87**(3):215-224.
35. Usui T, Macleod MR, McCann SK, Senior AM, Nakagawa S: **Meta-analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research**. *Plos Biol* 2021, **19**(5).
36. Hoffmann AA, Merila J: **Heritable variation and evolution under favourable and unfavourable conditions**. *Trends Ecol Evol* 1999, **14**(3):96-101.
37. Wood CW, Brodie ED, 3rd: **Environmental effects on the structure of the G-matrix**. *Evolution* 2015, **69**(11):2927-2940.
38. Hillebrand H, Donohue I, Harpole WS, Hodapp D, Kucera M, Lewandowska AM, Merder J, Montoya JM, Freund JA: **Thresholds for ecological responses to global change do not emerge from empirical data**. *Nat Ecol Evol* 2020, **4**(11):1502-+.
39. Yang YF, Hillebrand H, Lagisz M, Cleasby I, Nakagawa S: **Low statistical power and overestimated anthropogenic impacts, exacerbated by publication bias, dominate field studies in global change biology**. *Global Change Biol* 2022, **28**(3):969-989.
40. Nakagawa S, Santos ESA: **Methodological issues and advances in biological meta-analysis**. *Evol Ecol* 2012, **26**(5):1253-1274.
41. Bishop J, Nakagawa S: **Quantifying crop pollinator dependence and its heterogeneity using multi-level meta-analysis**. *J Appl Ecol* 2021, **58**(5):1030-1042.

42. Cheung MWL: **Modeling Dependent Effect Sizes With Three-Level Meta-Analyses: A Structural Equation Modeling Approach**. *Psychol Methods* 2014, **19**(2):211-229.
43. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS: **Generalized linear mixed models: a practical guide for ecology and evolution**. *Trends Ecol Evol* 2009, **24**(3):127-135.
44. Lajeunesse MJ: **On the meta-analysis of response ratios for studies with correlated and multi-group designs**. *Ecology* 2011, **92**(11):2049-2055.
45. Gleser LJ, Olkin I: **Stochastically dependent effect sizes**. In: *The handbook of research synthesis and meta-analysis* Edited by Cooper H, Hedges LV, Valentine JC. New York: Russell Sage Foundation; 2009.
46. Tipton E, Pustejovsky JE: **Small-Sample Adjustments for Tests of Moderators and Model Fit Using Robust Variance Estimation in Meta-Regression**. *J Educ Behav Stat* 2015, **40**(6):604-634.
47. Hedges LV, Tipton E, Johnson MC: **Robust variance estimation in meta-regression with dependent effect size estimates (vol 1, pg 39, 2010)**. *Res Synth Methods* 2010, **1**(2):164-165.
48. Pustejovsky JE, Tipton E: **Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models**. *Prev Sci* 2021.
49. Higgins JPT, Thompson SG: **Quantifying heterogeneity in a meta-analysis**. *Stat Med* 2002, **21**(11):1539-1558.
50. Higgins JPT, Thompson SG, Deeks JJ, Altman DG: **Measuring inconsistency in meta-analyses**. *Brit Med J* 2003, **327**(7414):557-560.
51. Xiong CJ, Miller JP, Morris JC: **Measuring Study-Specific Heterogeneity in Meta-Analysis: Application to an Antecedent Biomarker Study of Alzheimer's Disease**. *Stat Biopharm Res* 2010, **2**(3):300-309.
52. Nakagawa S, Schielzeth H: **Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists**. *Biol Rev* 2010, **85**(4):935-956.
53. Senior AM, Grueber CE, Kamiya T, Lagisz M, O'Dwyer K, Santos ESA, Nakagawa S: **Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications**. *Ecology* 2016, **97**(12):3293-3299.
54. Gelman A, Hill J: **Data analysis using regression and multilevel/hierarchical models**. Cambridge: Cambridge University Press; 2007.
55. Schielzeth H: **Simple means to improve the interpretability of regression coefficients**. *Methods in Ecology and Evolution* 2010, **1**(2):103-113.
56. Nakagawa S, Schielzeth H: **A general and simple method for obtaining R2 from generalized linear mixed-effects models**. *Methods in Ecology and Evolution* 2013, **4**(2):133-142.
57. Nakagawa S, Johnson PCD, Schielzeth H: **The coefficient of determination R-2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded**. *J R Soc Interface* 2017, **14**(134).
58. Aloe AM, Becker BJ, Pigott TD: **An alternative to R-2 for assessing linear models of effect size**. *Res Synth Methods* 2010, **1**(3-4):272-283.
59. Cinar O, Umbanhowar J, Hoeksema JD, Viechtbauer W: **Using information-theoretic approaches for model selection in meta-analysis**. *Res Synth Methods* 2021.
60. Viechtbauer W: **Model checking in meta-analysis**. In: *Handbook of meta-analysis.* Edited by Schmid CH, Stijnen T, White IR. Boca Ranton: CRC; 2021: 219-254.
61. Anzures-Cabrera J, Higgins JPT: **Graphical displays for meta-analysis: An overview with suggestions for practice**. *Res Synth Methods* 2010, **1**(1):66-80.
62. Kossmeier M, Tran US, Voracek M: **Charting the landscape of graphical displays for meta-analysis and systematic reviews: a comprehensive review, taxonomy, and feature analysis**. *Bmc Med Res Methodol* 2020, **20**(1).

63. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ: **Plea for routinely presenting prediction intervals in meta-analysis**. *Bmj Open* 2016, **6**(7).
64. Moeyaert M, Ugille M, Beretvas SN, Ferron J, Bunuan R, Van den Noortgate W: **Methods for dealing with multiple outcomes in meta-analysis**

a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *Int J Soc Res Method* 2017, **20**(6):559-572.
65. Nakagawa S, Lagisz M, O'Dea RE, Rutkowska J, Yang YF, Noble DWA, Senior AM: **The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond**. *Res Synth Methods* 2021, **12**(1):4-12.
66. Rothstein H, Sutton AJ, Borenstein M: **Publication bias in meta-analysis : prevention, assessment and adjustments**. Chichester, England ; Hoboken, NJ: Wiley; 2005.
67. Nakagawa S, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, Sanchez-Tojar A, Yang YF, O'Dea RE: **Methods for testing publication bias in ecological and evolutionary meta-analyses**. *Methods in Ecology and Evolution* 2022, **13**(1):4-21.
68. Stanley TD, Doucouliagos H: **Meta-regression analysis in economics and business**. New York: Routledge; 2012.
69. Stanley TD, Doucouliagos H: **Meta-regression approximations to reduce publication selection bias**. *Res Synth Methods* 2014, **5**(1):60-78.
70. Sterne JAC, Becker BJ, Egger M: **The funnel plot**. In: *Publication bias in meta-analysis : prevention, assessment and adjustments.* Edited by Rothstein H, Sutton AJ, Borenstein M. Chichester: Wiley; 2005: 75-98.
71. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH *et al*: **Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials**. *Bmj-British Medical Journal* 2011, **343**.
72. Egger M, Smith GD, Schneider M, Minder C: **Bias in meta-analysis detected by a simple, graphical test**. *Brit Med J* 1997, **315**(7109):629-634.
73. Jennions MD, Moller AP: **Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution**. *P Roy Soc B-Biol Sci* 2002, **269**(1486):43-48.
74. Koricheva J, Kulinskaya E: **Temporal instability of evidence base: a threat to policy making?** *Trends Ecol Evol* 2019, **34**(10):895-902.
75. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U, Timmer J: **Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**. *Bioinformatics* 2009, **25**(15):1923-1929.
76. Matsushima Y, Noma H, Yamada T, Furukawa TA: **Influence diagnostics and outlier detection for meta-analysis of diagnostic test accuracy**. *Res Synth Methods* 2020, **11**(2):237-247.
77. Viechtbauer W, Cheung MWL: **Outlier and influence diagnostics for meta-analysis**. *Res Synth Methods* 2010, **1**(2):112-125.
78. Haddaway NR, Macura B: **The role of reporting standards in producing robust literature reviews comment**. *Nat Clim Change* 2018, **8**(6):444-447.
79. Frampton G, Whaley P, Bennett M, Bilotta G, Dorne JLCM, Eales J, James K, Kohl C, Land M, Livoreil B *et al*: **Principles and framework for assessing the risk of bias for studies included in comparative quantitative environmental systematic reviews**. *Environ Evid* 2022, **11**(1).
80. Stanhope J, Weinstein P: **Critical appraisal in ecology: What tools are available, and what is being used in systematic reviews?** *Res Synth Methods* 2022.
81. Haddaway NR, Macura B, Whaley P, Pullin AS: **ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps**. *Environ Evid* 2018, **7**(1).

82. Woodcock P, Pullin AS, Kaiser MJ: **Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: A methodology**. *Biol Conserv* 2014, **176**:54-62.
83. O'Dea RE, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, Gurevitch J, Page MJ, Stewart G, Moher D *et al*: **Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension**. *Biol Rev* 2021, **96**(5):1695-1722.
84. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P: **Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement**. *Plos Med* 2009, **6**(7).
85. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE *et al*: **The PRISMA 2020 statement: An updated guideline for reporting systematic reviews**. *Plos Med* 2021, **18**(3).
86. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, Santos LBD, Bourne PE *et al*: **Comment: The FAIR Guiding Principles for scientific data management and stewardship**. *Sci Data* 2016, **3**.
87. Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen-Windhouwer S, Manghi P: **Navigating the unfolding open data landscape in ecology and evolution**. *Nat Ecol Evol* 2018, **2**(3):420-426.
88. Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, Kokko H, Jennions MD, Kruuk LEB: **Troubleshooting Public Data Archiving: Suggestions to Increase Participation**. *Plos Biol* 2014, **12**(1).
89. Roche DG, Kruuk LEB, Lanfear R, Binning SA: **Public Data Archiving in Ecology and Evolution: How Well Are We Doing?** *Plos Biol* 2015, **13**(11).
90. Culina A, van den Berg I, Evans S, Sanchez-Tojar A: **Low availability of code in ecology: A call for urgent action**. *Plos Biol* 2020, **18**(7).
91. Spake R, Mori AS, Beckmann M, Martin PA, Christie AP, Duguid MC, Doncaster CP: **Implications of scale dependence for cross-study syntheses of biodiversity differences**. *Ecol Lett* 2021, **24**(2):374-390.
92. Osenberg CW, Sarnelle O, Cooper SD: **Effect size in ecological experiments: The application of biological models in meta-analysis**. *Am Nat* 1997, **150**(6):798-812.
93. Doncaster CP, Spake R: **Correction for bias in meta-analysis of little-replicated studies**. *Methods in Ecology and Evolution* 2018, **9**(3):634-644.
94. Noble DWA, Nakagawa S: **Planned missing data designs and methods: Options for strengthening inference, increasing research efficiency and improving animal welfare in ecological and evolutionary research**. *Evol Appl* 2021, **14**(8):1958-1968.
95. Nakagawa S, Freckleton RP: **Missing inaction: the dangers of ignoring missing data**. *Trends Ecol Evol* 2008, **23**(11):592-596.
96. Mavridis D, Chaimani A, Efthimiou O, Leucht S, Salanti G: **Addressing missing outcome data in meta-analysis**. *Evid-Based Ment Heal* 2014, **17**(3):85-+.
97. Ellington EH, Bastille-Rousseau G, Austin C, Landolt KN, Pond BA, Rees EE, Robar N, Murray DL: **Using multiple imputation to estimate missing data in meta-regression**. *Methods in Ecology and Evolution* 2015, **6**(2):153-163.
98. Kambach S, Bruelheide H, Gerstner K, Gurevitch J, Beckmann M, Seppelt R: **Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis**. *Ecol Evol* 2020, **10**(20):11699-11712.
99. Nakagawa S, Noble DW, Lagisz M, Spake R, Viechtbauer W, M SA: **A robust and readily implementable method for the meta-analysis of response ratios with and without missing standard deviations**. *Ecol Lett* 2022, **DOI: 10.1111/ele.14144**.
100. Hadfield JD, Nakagawa S: **General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters**. *J Evolution Biol* 2010, **23**(3):494-508.

101. Cinar O, Nakagawa S, Viechtbauer W: **Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny**. *Methods in Ecology and Evolution* 2021.
102. Ives AR, Zhu J: **Statistics for correlated data: Phylogenies, space, and time**. *Ecol Appl* 2006, **16**(1):20-32.
103. Maire A, Thierry E, Viechtbauer W, Daufresne M: **Poleward shift in large-river fish communities detected with a novel meta-analysis framework**. *Freshwater Biol* 2019, **64**(6):1143-1156.
104. Morrissey MB: **Meta-analysis of magnitudes, differences and variation in evolutionary parameters**. *J Evolution Biol* 2016, **29**(10):1882-1904.
105. Cohen JM, Lajeunesse MJ, Rohr JR: **A global synthesis of animal phenological responses to climate change**. *Nat Clim Change* 2018, **8**(3):224-+.
106. Gurevitch J, Morrison JA, Hedges LV: **The interaction between competition and predation: A meta-analysis of field experiments**. *Am Nat* 2000, **155**(4):435-453.
107. Macartney EL, Lagisz M, Nakagawa S: **The relative benefits of environmental enrichment on learning and memory are greater when stressed: a meta-analysis of interactions in rodents**. *Neurosci Biobehav R* 2022, **in press**.
108. Siviter H, Bailes EJ, Martin CD, Oliver TR, Koricheva J, Leadbeater E, Brown MJF: **Agrochemicals interact synergistically to increase bee mortality**. *Nature* 2021, **596**(7872).
109. Salanti G, Schmid CH: **Research Synthesis Methods special issue on network meta-analysis: introduction from the editors**. *Res Synth Methods* 2012, **3**(2):69-70.
110. Gomez-Creutzberg C, Lagisz M, Nakagawa S, Brockerhoff EG, Tylianakis JM: **Consistent trade-offs in ecosystem services between land covers with different production intensities**. *Biol Rev* 2021, **96**(5):1989-2008.
111. Jackson D, White IR, Price M, Copas J, Riley RD: **Borrowing of strength and study weights in multivariate and network meta-analysis**. *Stat Methods Med Res* 2017, **26**(6):2853-2868.
112. Sanchez-Tojar A, Moran NP, O'Dea RE, Reinhold K, Nakagawa S: **Illustrating the importance of meta-analysing variances alongside means in ecology and evolution**. *J Evolution Biol* 2020, **33**(9):1216-1223.
113. Riley RD, Lambert PC, Abo-Zaid G: **Meta-analysis of individual participant data: rationale, conduct, and reporting**. *Bmj-British Medical Journal* 2010, **340**.
114. Riley RDe, Tierney JFe, Stewart LAe: **Individual Participant Data Meta-Analysis : A Handbook for Healthcare Research**, 1st edn.
115. Hudson LN, Newbold T, Contu S, Hill SLL, Lysenko I, De Palma A, Phillips HRP, Alhusseini TI, Bedford FE, Bennett DJ *et al*: **The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project**. *Ecol Evol* 2017, **7**(1):145-188.
116. Dornelas M, Antao LH, Moyes F, Bates AE, Magurran AE, Adam D, Akhmetzhanova AA, Appeltans W, Arcos JM, Arnold H *et al*: **BioTIME: A database of biodiversity time series for the Anthropocene**. *Global Ecol Biogeogr* 2018, **27**(7):760-786.
117. Mengersen K, Gurevitch J, Schmid CH: **Meta-analysis of primary data**. *Handbook of meta-analysis in ecology and evolution* 2013:300-312.
118. Spake R, O'Dea RE, Nakagawa S, Doncaster CP, Ryo M, Callaghan CT, Bullock JM: **Improving quantitative synthesis to achieve generality in ecology**. *Nat Ecol Evol* 2022, **6**(12):1818-1828.
119. Gerstner K, Moreno-Mateos D, Gurevitch J, Beckmann M, Kambach S, Jones HP, Seppelt R: **Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting**. *Methods in Ecology and Evolution* 2017, **8**(6):777-784.
120. Haddaway NR: **A call for better reporting of conservation research data for use in meta-analyses**. *Conserv Biol* 2015, **29**(4):1242-1245.

121. Midolo G, De Frenne P, Holzel N, Wellstein C: **Global patterns of intraspecific leaf trait responses to elevation**. *Global Change Biol* 2019, **25**(7):2485-2498.

122. White IR, Schmid CH, Stijnen T: **Choice of effect measure and issues in extracting outcome data**. In: *Handbook of meta-analysis.* Edited by Schmid CH, Stijnen T, White IR. Boca Ranton: CRC; 2021.

123. Lajeunesse MJ: **Bias and correction for the log response ratio in ecological meta-analysis**. *Ecology* 2015, **96**(8):2056-2063.