Preliminary reading on meta-analysis

|  |  |  |
| --- | --- | --- |
| Reference | Date | Information |
| (Gurevitch et al., 2018) | 31/3/22 | **Meta-analysis and the science of research synthesis**  Meta-analysis is the quantitative, scientific synthesis of research results.  Since the term and modern approaches to research synthesis were first introduced in the 1970s.  Resolve seemingly contradictory research outcomes.  Systematic reviews aim to provide a robust overview of the efficacy of an intervention, or of a problem or field of research. They can be combined with quantitative meta-analyses to assess the magnitude of the outcome across relevant primary studies and to analyse the causes of variation among study outcomes (effect sizes).  Narrative reviews remain useful for exploring the development of particular ideas (as we do here) and for advancing conceptual frameworks, but they cannot accurately summarize results across studies.  If the systematic review reveals sufficient and appropriate quantitative data from the studies that are being summarized, then a meta-analysis can be conducted.   * Meta-analyses are often part of a systematic review. Meta-analysis is statistical procedure   Systematic reviews aim to be transparent, reproducible and updatable, and to address well-defined questions.  In a meta-analysis, one or more outcomes in the form of effect sizes are extracted from each study. Effect sizes are designed to put the outcomes of the different studies being combined on the same scale. The effect sizes are then entered into a statistical model with the goal of assessing overall effects and heterogeneity in outcomes.  The publication of methodologically flawed meta-analyses indicates that peer reviewers, editors and authors are not fully aware of or are indifferent to the large body of well-developed meta-analytic methodology, and that reviewers might feel unqualified to address statistical issues  These models are based on an assumption of either a common effect (‘fixed effect’) or random effects (Fig. 1b) 16. The common-effect (or fixed-effect) model assumes that variation in effect sizes among studies is due to within-study (sampling) variance and that all studies share a common ‘true’ effect. The random-effects model assumes that, in addition to sampling variance, the true effects from different studies also differ from one another, representing a random sample of a population of outcomes.  Random-effects models include an extra variance component to account for between-study variance (heterogeneity) in addition to within-study variance. Common-effect models are based on the assumption that the results apply only to a given group of studies. Random-effects models apply more generally.  Quantifying heterogeneity is universally important.  In carrying out a meta-analysis, the central tendency (the mean) and its confidence limits are evaluated, as well as the heterogeneity in the effect across studies. To identify the magnitude and sources of variation in effect size among studies (Fig. 1c), earlier studies relied on simple heterogeneity tests16, whereas more recent work often uses meta-regressions17. Heterogeneity tests and meta-regressions both use weighting based on the precision of the estimate of the effect: larger studies with higher precision are weighted more heavily than smaller and/or more variable studies18  metafor (free and comprehensive R package for meta-analysis) released in 2010.  Fundamental goal is to reach broad generalizations across larger numbers of study outcomes (dozens to hundreds) to provide a more comprehensive picture than can be attained from an individual primary study.  Meta-analysis was first adopted by ecologists and evolutionary biologists some 25 years ago.  Critics have claimed that the potential for publication bias in the literature (that is, the under-reporting of non-significant results or disconfirming evidence21) invalidates the use of meta-analysis.  Some ecologists have claimed that ecological studies are too heterogeneous to be combined statistically in a meaningful way9 and that ecology is best served by accumulating a catalogue of case studies53.  Meta-analyses and systematic reviews can highlight areas in which evidence is deficient, but they cannot overcome these deficiencies.  An individual primary study may now be seen as a contribution towards the accumulation of evidence rather than revealing the conclusive answer to a scientific problem.  The term meta-analysis should be applied only to studies that use well-established statistical procedures, such as appropriate effect-size calculation, weighting and heterogeneity analysis57, and statistical models that take into account the distinct hierarchical structure of meta-analytic data, or to studies that develop rigorously justified methodological advances of these methods.  Unfortunately, the term is often misapplied to any study that uses data from several primary publications, regardless of the rigour of the methodology.  Statistically flawed procedures such as vote-counting, which provide only limited information about study outcomes, can be very misleading and have long been discredited, are still used in published papers6,50. Vote-counting is a deceptively plausible and appealingly convenient procedure whereby the generality of findings in a group of studies is assessed by counting up the number of significant and non-significant results in individual studies (or by elaborations on this approach).  In an unweighted analysis, within- and between-study variation cannot be readily separated, and so common- and random-effects models cannot be used and heterogeneity may be difficult to assess properly. Unweighted meta-analysis also increases the influence of small studies29. However, the variances needed for a weighted meta-analysis are frequently unavailable owing to poor data reporting in the primary studies.  Pre-registration (called ‘registration’ in some fields) of planned studies can reduce selective reporting of outcomes; publication of ‘registered reports’ in which the methods and proposed analyses for a study are peer-reviewed and published before the research is conducted can reduce publication bias.  Advances in meta-analytic techniques are being driven by methodological developments. Advances include: the use of machine learning and artificial intelligence (AI) to screen studies for inclusion in systematic reviews and meta-analyses67. |
| (Arnqvist and Wooster, 1995) | 13/4/22 | **Meta-analysis - synthesizing research findings in ecology and evolution**  The growing number of empirical studies performed in ecology and evolution creates a need for quantitative summaries of research domains to generate higher order conclusions about general trends and patterns.  Meta-analysis (the area of statistics that is designed for summarizing and analyzing multiple independent studies).  Studies very rarely show identical results, but instead typically differ both in the magnitude of effects and in the occurrence of significant results.  No two studies in a set of studies are equally ‘reliable’. In MA, this critical fact is accounted for by giving estimates from different studies different weights, primarily based on their sample size.  Data or test statistics from these studies are transformed into a ‘common currency’, called ‘effect size’. Common measures of effect size are the standardized difference between means of experimental and control groups or the Pearson product moment correlation coefficient. These effect sizes are combined into a common estimate of the magnitude of the effect. The significance level of this overall effect size is computed.  It is less subjective than narrative reviews, since it is based on a formal, predetermined set of statistical procedures rather than individual interpretations of the data.  Perhaps the most universal problem is the potential bias that will result when the studies included in the MA are not representative of all studies conducted. This may result from biases either in publication rates or in selection/retrieval of studies.  More problems can be at least partly ameliorated in MA by detailed consideration of how studies are selected/retrieved, tests of robustness of conclusions, careful consideration of potential interdependences across studies and assessments of different between subgroups of studies by focused tests of homogeneity.  Meta-analysis will prove most useful in areas of ecology and evolution where (1) there is a moderate to large amount of empirical work available, (2) the results are variable across studies, (3) the expected magnitude of the effect is relatively weak, and/or (4) the sample sizes of individual studies are limited for some reason. |
| (Viechtbauer, 2010) | 13/4/22 | **Conducting Meta-Analyses in R with the metafor Package**  **Abstract**  The package includes functions for fitting the meta-analytic fixed- and random-effects models and allows for the inclusion of moderators variables.  **Intro**  Researchers trying to aggregate and synthesize the literature on a particular topic are increasingly conducting meta-analyses (Olkin 1995).  In a meta-analysis, the relevant results from each study are quantified in such a way that the resulting values can be further aggregated and compared.  we may be able to express the results from a randomized clinical trial examining the effectiveness of a medication in terms of an odds ratio, indicating how much higher/lower the odds of a particular outcome (e.g., remission) were in the treatment compared to the control group.  The set of odds ratios from several studies examining the same medication then forms the data which is used for further analyses.  Again, the idea is that the relevant results of each study are expressed in terms of an outcome measure putting the results on a common scale. Depending on the types of studies and the information provided therein, a variety of different outcome measures can be used for a meta-analysis, including the odds ratio, relative risk, risk difference, the correlation coefficient, and the (standardized) mean difference (e.g., Borenstein 2009; Fleiss and Berlin 2009).  In essence, the various meta-analytic models are just special cases of the general linear (mixed effects) model with heteroscedastic sampling variances that are assumed to be known.  Before beginning with a meta-analysis, one must first obtain a set of effect size estimates with their corresponding sampling variances. The metafor package also provides the escalc() function, which can be used to calculate various effect size or outcome measures (and the corresponding sampling variances).  Cell entries with a zero can be problematic especially for the relative risk and the odds ratio. Adding a small constant to the cells of the 2 × 2 tables is a common solution to this problem.  The various meta-analytic models can be fitted with the rma.uni() function (with alias rma()).  One simply needs to supply the observed outcomes via the yi argument and the corresponding sampling variances via the vi argument (or the standard errors, the square root of the sampling variances, via the sei argument).  One or more moderators can be included in the model via the mods argument.  A graphical overview of the results so far can be obtained by creating a forest plot (Lewis and Clarke 2001) with the forest() function.  The predict() function provides the fitted values in addition to standard errors and confidence interval bounds.  One can conduct full versus reduced model comparisons via likelihood ratio tests with the anova() function. |
| (Nakagawa et al., 2017) | 13/4/22 | **Meta-evaluation of meta-analysis: ten appraisal questions for biologists**  Documentation on keyword strings and inclusion criteria is often also very poor, making replication of search outcomes difficult or impossible.  The meta-analytic questions and hypotheses addressed will generally determine the types of effect size statistics the authors use [29–32], as we explain below. Three broad groups of effect size statistics are based on are: 1) the difference between the means of two groups (for example, control versus treatment); 2) the relationship, or correlation, between two variables; and 3) the incidence of two outcomes (for example, dead or alive) in two groups (often represented in a 2 by 2 contingency table).  Corresponding common effect size statistics are: 1) standardized mean difference (SMD; often referred to as d, Cohen’s d, Hedges’ d or Hedges’ g) and the natural logarithm (log) of the response ratio (denoted as either lnR or lnRR [33]); 2) Fisher’s z-transformed correlation coefficient (often denoted as Zr); and 3) the natural logarithm of the odds ratio (lnOR) and relative risk (lnRR; not to be confused with the response ratio).  Failing to account for non-independence among effect sizes (or data points) can lead to erroneous conclusions [14, 41–44]—typically, an invalid conclusion of statistical significance (type I error).  There are three main kinds of meta-analytic models, which differ in their assumptions about the data being analyzed, but for all three the common and primary goal is to estimate an overall effect (but see Q5). These models are: i) fixed-effect models (also referred to as common-effect models [31]); ii) random-effects models [50]; and iii) multilevel (hierarchical) models.  The overall effect reported by a meta-analysis cannot be properly interpreted without an analysis of the heterogeneity, or inconsistency, among effect sizes.  Moderators are equivalent to explanatory (independent) variables or predictors in a normal linear model [8, 49, 62]. Models that examine the effects of moderators are referred to as meta-regressions.  Meta-analyses should focus on biological importance (which is reflected in estimated effects and their uncertainties) rather than on p values and statistical significance.  Meta-analysts have to assume that research is published regardless of statistical significance, and that authors have not selectively reported results (that is, that there is no publication bias and no reporting bias) [74–76]. This is unlikely. Therefore, meta-analysts should check for publication bias using statistical and graphical tools. The reader should know that the commonly used methods for assessing publication bias are funnel plots (Fig. 6a, b), radial (Galbraith) plots (Fig. 6c), and Egger’s (regression) tests [57, 77, 78]; these methods visually or statistically (Egger’s test) help to detect funnel asymmetry, which can be caused by publication bias [79].  It is reasonable to expect the authors to discuss what conventional wisdoms the meta-analysis has confirmed or refuted and what new insights the meta-analysis has revealed [8, 19, 71, 100]. New insights from meta-analyses are known as ‘review-generated evidence’ (as opposed to ‘study-generated evidence’).  medical and social scientists are aware that updating meta-analyses is extremely important, especially given that time-lag bias is a common phenomenon [87–89]. Although updating is still rare in biological meta-analyses [8]. |
| (Van Klink et al., 2020)  Specific e.g. of MA used in studying insect trends | 27/4/22 | **Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances**  **Abstract**  We compiled data from 166 long-term surveys of insect assemblages across 1676 sites to investigate trends in insect abundances over time. 41 countries  Overall, we found considerable variation in trends even among adjacent sites but an average decline of terrestrial insect abundance by ~9% per decade and an increase of freshwater insect abundance by ~11% per decade.  Both patterns were largely driven by strong trends in North America and some European regions. We found some associations with potential drivers (e.g., land-use drivers), and trends in protected areas tended to be weaker.  **Main body**  Despite the attention from the media, policy-makers, and scientists, it remains unclear whether such declines are widespread across realms and among geographic regions.  We used the amassed data to evaluate changes in total insect abundance and biomass, as well as the geographic distribution of such changes.  Among these, 130 datasets reported only changes in insect abundances (i.e., number of individuals) in an assemblage, 13 datasets reported only the biomass of all insects in an assemblage, and 23 datasets reported both metrics. The data spanned from 1925 to 2018.  We analyzed the data using a hierarchical Bayesian model accounting for variation at the study, study area, and site level.  The mean trend estimates of insect abundance and biomass were similar (Fig. 2A) but differed in strength of evidence because of the lower data availability for biomass (table S2).  The positive trends in the fresh-water realm may partially counter the negative terrestrial trends, because a model combining both realms showed no evidence for a directional trend (Fig. 2A). However, because fresh water represents only 2.4% of the earth’s terrestrial surface (15, 16), such a combined model is likely to be a poor representation of trends in total insect numbers at any spatial scale.  We found that the trends in protected areas were weaker than those in unprotected areas (Fig. 4), although there was still a moderate negative trend in terrestrial protected areas. This difference suggests a possible association between insect trends and land-use change.  We calculated the relative change in temperature and precipitation over the sampling period at local and regional scales for each site (14) to test for a potential role of climate change, but found no evidence for any associations at either scale.  As with most data compilations of this kind, our data sources were not representatively spread across the world.  Protected areas were overrepresented in our dataset (34% of the sites) relative to the percentage of the terrestrial surface currently under protection (15%) (31). This means that locations where human land use is most intensive, and thus where the strongest effects on insect trends might be expected, were underrepresented.  We found an average increase in fresh-water insect abundances that might, at least partially, reflect improvements in water quality. This, in combination with our finding that trends were weaker in protected areas, suggests that appropriate habitat protection and restoration may be effective strategies for mitigating changes in insect assemblages.  **Methods (in supplementary materials)**  We searched the scientific literature and public data repositories for time-series of freshwater and terrestrial insect and arachnid assemblages, spanning at least 10 years between the first and last sampling date.  We used autoregressive mixed effects models to assess trends in insect assemblage size (measured as abundance or biomass), and to test whether these temporal trends differed among realms (freshwater or terrestrial), unit (abundance or biomass), continents, capture strata, climatic zones, and with different global change drivers.  Criticised by (Jähnig et al., 2021)   * First, total abundance and biomass alone are poor indicators of the status of freshwater insect assemblages, and the observed differences may well have been driven by the replacement of sensitive species with tolerant ones. * Second, many of the datasets poorly represent global trends and reflect responses to local conditions or non-random site selection |
| (Dirzo et al., 2014)  Specific e.g. of MA used in studying insect trends | 27/4/22 | **Includes a meta‐analysis of effects of disturbance on Lepidoptera species richness and abundance**  Likewise, among pairs of disturbed and undisturbed sites globally, Lepidopteran species richness is on average 7.6 times higher in undisturbed than disturbed sites, and total abundance is 1.6 times greater (Fig. 1D) (19).  **Methods (in supplementary materials)**  We reviewed the published literature for studies of Lepidoptera species richness in disturbed and undisturbed habitats.  In both disturbance and richness analyses, we calculated effect sizes (Hedge’s g) using Comprehensive Meta‐analysis software, ver 2.0.  To test for publication bias, we examined funnel plots and conducted Duval and Tweedie's trim and fill analysis.  In total we calculated 52 diversity effect sizes from 15 studies (126‐140) that met our criteria for inclusion in the meta‐analysis (Fig S3). The meta‐analysis revealed an overall negative effect (Z = 5.91, P < 0001) with a large mean effect estimate of 1.02 (95% CI 0.68 to 0.1.36). |
| (Nakagawa and Santos, 2012) | 27/4/22 | **Methodological issues and advances in biological meta-analysis**  **Abstract**  Meta-analysis has changed the way researchers conduct literature reviews not only in medical and social sciences (where meta-analytic techniques were originally developed) but also in biological sciences.  The main reason for such differences is that biological meta-analysis often integrates complex data composed of multiple strata with, for example, different measurements and a variety of species.  **Intro**  In a more statistical sense, meta-analysis combines common effect-size statistics (e.g., Hedges’ d or the correlation coefficient, r) extracted from relevant studies by accounting for the sample sizes of the studies (i.e. sampling errors).  **Effect-size statistics and standard meta-analytic models**  Three categories (comprising six types in total) of effect-size statistics are commonly used in meta-analysis (Borenstein et al. 2009); these three types are based on (1) two different means, (2) relationships/correlations (r) between two variables, (3) 2-by-2 contingency tables (i.e. binary data).   * The first category incorporates two types: (a) standardized mean differences (e.g., d or g) and (b) response ratios (Hedges et al. 1999) * The third category includes three types: (a) odds ratio, (b) relative risk, and (c) risk difference.   These six effect-size statistics have two properties in common, which are required for ‘traditional’ meta-analysis: (1) they, or their transformations, are normally distributed (e.g., Fisher’s z transformation of r, Zr, or the natural logarithm of odds ratio, ln(OR)), and (2) their, or their transformations’, sampling variances are estimable from formulas.  Standardized mean difference was the most popular category of effect-size statistics reported (61.5%; e.g., Hedges’ d and the response ratio). In fact, the response ratio was the most commonly used effect-size statistic overall (32.7%), although the use of this effect size statistic is almost exclusive to the field of ecology and evolution (Borenstein et al. 2009).  The nature of meta-analysis usually dictates which effect-size statistic is used (Nakagawa and Cuthill 2007; Borenstein et al. 2009; Cooper et al. 2009), but this is not always the case nor straightforward (e.g., Osenberg et al. 1999; Lajeunesse and Forbes 2003).  There are two types of standard meta-analytic models: (1) fixed-effect models, and (2) random-effects models.  A fixed-effect meta-analysis can be written as: zj = u + mj, where zj is an effect size for the jth study (j = 1,…, Nstudy; Nstudy is the number of studies), u is the meta-analytic mean (or intercept), mj is a sampling (measurement) error effect for the jth study, m is a 1 by Nstudy vector of mj, which is normally distributed.  A random-effects meta-analysis uses the more reasonable assumption that each study has a ‘true’ effect size different from each other. It can be written as: zj = u + uj + mj where uj is the study specific effect of the jth study, u is a 1 by Nstudy vector of uj, which is normally distributed.  Although the random-effects meta-analytic model described above is useful and widely used, there is one serious limitation. This model is designed for meta-analyses where the number of effect sizes equals the number of studies. Biological meta-analyses will frequently include studies that often provide more than one effect-size estimate.   * How can there be more than one effect size? Quite a few studies generated multiple effect sizes, from: i) assessing one or more functions by multiple insect species, or ii) repeated measures of one or more functions by the same species   We could categorise the non-independence problem in biological meta-analysis into two types: (1) non-independence of observations from related sources, and (2) non-independence from phylogenetic relatedness of species when multi-species are used.  Effect-size estimates can originate from the same study groups, populations or species. Also, one study may measure multiple traits as proxies for, say, fitness (e.g., mating success, breeding success and survival). In the past, researchers often acknowledged this type of non-independence, but they, nonetheless, carried out meta-analysis with non-independent datasets using the standard fixed- or random-effects models. Some meta-analyses opted to include only one effect-size estimate per study, losing statistical power and potentially forgoing important information.  Therefore, to address this issue most effectively, we should use multilevel/hierarchical models, which explicitly model correlations within the different levels (for an example of a Bayesian hierarchical meta-analysis). This multilevel meta-analytic model can be implemented in, for example, the metafor package in R.  **Heterogeneity**  Next to finding a general trend or calculating meta-analytic means, the most important function of meta-analysis is to investigate inconsistency (or consistency) across studies, or quantifying heterogeneity in the data. The reliability of a general trend (meta-analytic mean) depends on the degree of consistency among studies (heterogeneity).  Currently, the two test statistics most commonly used to quantify heterogeneity are Cochran’ Q (Hedges and Olkin 1985) and I2 (Higgins and Thompson 2002; Higgins et al. 2003). All major meta-analytic programmes provide Q, which probably explains its widespread application in traditional meta-analyses. More recently, I 2 is increasingly being used and newer computer programmes such as metafor.  The statistic I2 shifts the focus from statistical significance to actual variance and is therefore an improvement over Q as a meta-analytic heterogeneity measurement. I2 = 25, 50 and 75% are considered as low, moderate and high heterogeneity, respectively (Higgins et al. 2003).  Q and I 2 are designed only for quantifying the degree of between-study variance. Therefore, they are unsuitable for multilevel meta-analysis. Authors propose a simple solution to this problem by extending the idea on which the formulation of I 2 is based.  Once heterogeneity is identified, the next step in meta-analysis is to incorporate moderators (referred to as explanatory or independent variables in the linear model framework), which may explain the observed heterogeneity. In other words, we move onto meta-regression (also called mixed-effects meta-analysis).  Because heterogeneity is almost always expected, we suggest that meta-regression should be the default meta-analytic model for biological meta-analysis.  One can use information criteria such as Akaike information criterion (AIC; e.g., Jones et al. 2009; Knowles et al. 2009; Weir et al. 2011) or deviance information criterion (DIC; e.g., Horva´thova´ et al. 2012) to select ‘better’ fitting meta-regression models.  **Publication bias**  Publication bias affects any sort of literature synthesis, including meta-analysis because studies with statistically significant results are more likely to be published than otherwise (Rosenthal 1979).  The ability to detect publication bias in a given field is a key strength of meta-analysis, because identification of publication bias will challenge the validity of common views in that area, and will spur further investigations.  There are two types of statistical procedures for dealing with publication bias in meta-analysis: (1) methods for identifying the existence of publication bias, and (2) methods for assessing the impact of publications bias (Sutton 2009). The former includes: (a) the funnel plot (and other visualisation methods such as the normal quantile plot), and (b) regression/correlation-based tests; while the latter includes: (i) the fail-safe (file-drawer) number (N), (ii) the trim and fill method, and (iii) selection model approaches (Sutton 2009).  Funnel plot - identify publication bias, which is manifested as funnel asymmetry. Funnel plots are not statistical tests so that one cannot statistically judge whether observed funnel shapes are more asymmetric than symmetric. There are two commonly used statistical methods to quantify funnel asymmetry: (1) The rank correlation method proposed by Begg and Mazumdar (1994), and (2) the linear regression method proposed by Egger et al. (1997) (often referred to as Egger’s regression or Egger test). Detection of publication bias largely relies on sample size. Publication bias is only one of the potential causes of funnel asymmetry, it could also just be down to chance.  The trim and fill method provides a conceptually easy and visually appealing way to adjust for the impact of missing studies in meta-analyses. It restores funnel symmetry by using existing data points to impute missing studies. However, again this method has been designed for meta-analysis where independence of data points can be assumed. |
| (Côté and Reynolds, 2012) | 27/4/22 | **Meta-analysis at the intersection of evolutionary ecology and conservation**  The adoption of meta-analysis in ecological fields coincided with, and undoubtedly fuelled, a growing scientific appetite for pulling results together to make geographically or taxonomically broad-scale comparisons. In the field of biodiversity conservation, many of the greatest hits in high-flying journals have been global syntheses, particularly in recent years.  Given the ever increasing number of studies linking conservation and evolution, and the fact that meta-analysis has infiltrated both evolutionary and applied ecology, it is timely to ask how important meta-analytical techniques have been in contributing towards syntheses at the intersection of these two disciplines.  For this paper, we view meta-analysis strictly as the combining of effect sizes (i.e., scale-independent measures of the magnitude of an effect of interest, which can be meaningfully compared across studies), weighted to reflect their statistical robustness (often by the inverse of the study variance), into a global effect size.  Our initial search turned up more than 1,500 papers that appeared to be meta-analyses related to conservation, and a similar number related to evolutionary ecology. However, when we combined these topics and scrutinized the papers, we were surprised to find only 23 formal meta-analyses linking conservation to evolution.  Meta-analysts typically had to drop over half of the papers they wanted to summarize because of lack of basic information such as variances and sample sizes. This is a sad commentary on the standards of reporting in primary research in evolutionary ecology and conservation. |
| (Gurevitch and Hedges, 1999) | 27/4/22 | **Statistical issues in ecological meta-analyses**  **Abstract**  Special statistical methods are usually needed for meta-analysis, both because effect-size indexes are typically highly heteroscedastic and because it is desirable to be able to distinguish between-study variance from within-study sampling-error variance.  **Intro**  Issues that have been discussed at great length in other fields in which meta-analysis is widely used include approaches to searching the literature, methods for dealing with studies of mixed quality, and publication bias.  Statistical analyses of effect sizes can be constructed to answer a great many questions. For example, how large is the effect overall? Is it positive or negative, and is it reliably different than zero? Are the results consistent across studies? If the results are not in agreement among studies, are there differences in the magnitude of the effect among meaningful categories of studies (e.g., does the effect differ among systems, trophic levels, etc.)?  Once the data are collected, the process of carrying out a meta-analysis typically involves choosing an appropriate metric of effect size (Osenberg et al. 1999), calculating grand-mean effect size across studies and means for different categories of explanatory variables (or slopes, where the explanatory variable is continuous), determining the confidence intervals around the means or slopes, and then carrying out statistical tests to determine the consistency of the effects within and among categories of studies.  Unfortunately, vote counting has very poor properties as a statistical procedure. The results of vote counts are seriously biased, the method has low statistical power, and most importantly it fails to provide critical information on the overall results of the body of studies.  One of the most useful solutions when faced with the heteroscedasticity typically found in meta-analysis data is to weight effect sizes for statistical analysis by the inverse of the sampling variance of the effect size.  Weighting does have the desirable property (in addition to the increase in precision) of counting large studies more heavily than small ones, which often seems reasonable in summarizing overall results.  Such methods generally fall into one of three categories: fixed-, mixed-, or random-effects procedures. In fixed-effect models, it is assumed that all studies with similar-enough characteristics share a common, ‘‘true’’ effect size, and estimates differ from one another by sampling error only.  In a random-effects model, the true effect size is expected to differ among studies, and the goal of the analysis is to quantify the variation in the effect parameters. Random-effects models are intuitively more appealing in ecology than fixed-effects models because we would often expect the true effect to vary among studies.  Mixed models in meta-analysis as a way to combine the advantages of random- and fixed effects models, much as in the analysis of primary data. Mixed models are appropriate for analyzing differences between groups of experiments when the groups are not expected to be internally homogeneous. In essence the effects of the groupings of experiments on effect parameters are fixed effects while the variations among the effect parameters of experiments within groupings are taken to be random effects.  Four issues that are of particular concern to ecologists are: incomplete data reporting, the lack of independence among effect-size estimates, publication bias, and research bias. We believe strongly that by far the most serious of these problems, and the one that is in principle easiest for ecologists to do something about, is poor data reporting.  Ecological experiments commonly fail to report sample sizes and variances, for instance, which makes it impossible to include those studies in a meta-analysis that uses the standard (weighted) parametric statistical tests designed for meta-analysis.  Both types of non-independence can lead to underestimates of the standard error of the mean effect and therefore liberal evaluations of the statistical significance of effects.  If several different measurements are made on each replicate in a study (e.g., measures at several points in time or of slightly different outcomes) and different effect sizes are computed from each, the different effect sizes may be correlated because the data on which they are based are correlated.  One alternative is to conduct a different meta-analysis for each kind of effect measure. For example, a set of studies might report on the outcome of competition in terms of both effects on growth and effects on survival, and one could then do a meta-analysis on growth effects and a separate meta-analysis on survival effects. Other (more complex) approaches can sometimes be employed that permit the use of all of the data via multivariate methods that explicitly model the dependence structure.  One class of methods for detecting publication bias is based on examination of the relation between standard error and effect size. These include graphical methods (the funnel plot, Light and Pillemer 1984), and formal tests for the correlation between sample size (or standard error of the effect-size estimate) and effect size (Begg and Mazumdar 1994).  However, in many ecological applications, the effects are often expected to vary substantially across experiments. Consequently, a relation between sample size and effect size may reflect rational experimental design rather than publication bias. In such situations, which may be typical of experimental ecology, tests for publication bias based on more elaborate selection models (Hedges 1992b), and procedures for establishing robustness of results (the fail-safe N, Rosenthal 1979) may be more appropriate. |
| (Hedges et al., 1999) | 27/4/22 | **The meta-analysis of response ratios in experimental ecology**  **Abstract**  The response ratio (the ratio of mean outcome in the experimental group to that in the control group) and closely related measures of proportionate change are often used as measures of effect magnitude in ecology.  **Intro**  In a meta-analysis, the result of each independent experiment is usually expressed as an index of effect; these effect estimates are then combined across studies to produce a summary of the findings. Subgroupings of experiments may be examined separately to determine whether experimental results differ across biologically meaningful groupings of experiments (e.g., whether the effects of a manipulation such as elevated atmospheric CO2 on total biomass differ across plant taxa).  Meta-analyses can only provide meaningful summaries if the effect-size index used is a meaningful summary of any one experiment. Several recent meta-analyses in ecology have used the standardized difference between means (the difference between the mean of the treated group and a control group, divided by the within-group standard deviation), also called a ‘‘d index,’’ as an index of effect. Although its statistical properties are well understood, and a substantial set of meta-analytic procedures using the d index are available (see Hedges and Olkin 1985), it is not always a meaningful way to summarize experiments in ecology (Osenberg et al. 1997).  The response ratio, the ratio of some measured quantity in experimental and control groups, is commonly used as a measure of experimental effect because it quantifies the proportionate change that results from an experimental manipulation. Examples of such ratios include relative competition intensity, relative yield, and relative crowding coefficient.  R = XE/XC where R is the response ratio, XE is mean of experimental group, XC is mean of control group.  However, it is desirable to perform statistical analyses in the metric of the natural logarithm of the response ratio. L = ln(R). The logarithm linearizes the metric. The distribution of L is much more normal in small samples than that of R.    If the between-experiment variation is too large (e.g., many times the average within-experiment sampling-error variation), one might question whether the experimental results are similar enough to warrant combination.  However the effect estimates from different experiments will typically differ in precision (standard error). Therefore a weighting of the individual study estimates giving greater weight to experiments whose estimates have greater statistical precision (smaller standard error) will increase the precision of the combined estimate. Consequently, a weighted mean is typically used in meta-analysis.  E.g. of its use: Nine of the studies had zero biomass in the control groups, making computation of a response ratio impossible. Therefore more than one third had values of the denominator that were too small for the normal approximation to be adequate.  - So can't use response ratio if substantial proportion of studies have zero as their control mean. |
| (Vetter et al., 2013) | 27/4/22 | **Meta-analysis: A need for well-defined usage in ecology and conservation biology**  **Abstract**  ‘‘Meta-analysis’’, however, is not well-defined in these fields, but is regularly confused with other summary analysis techniques, such as multiple regression methods, vote-counting or other quantitative analyses.  We argue that this vague and inconsistent utilization of the term is problematic, because a meta-analysis typically provides scientifically rigorous results.  In the first round of rating, we assessed the usage of four ‘‘technical’’ steps that are normally applied in meta-analytical software. In the second round, we only evaluated the highly rated articles from the first round. We considered three steps regarding more qualitative aspects of interpretation and results presentation. Of the 133 evaluated articles in the first round, only 45% fulfilled all technical requirements for a meta-analysis, while 25% did not fulfill any of the requisite steps. In the second round, only one article of 83 fulfilled all requisite steps, while 22% did not fulfill any requirement.  **Intro**  The benefits of meta-analysis are higher statistical power and better precision, as well as the ability to address a broader scope than the combined primary studies.  Meta-analysis is sometimes confused with systematic reviews (Nakagawa and Poulin 2012, The Cochrane Collaboration 2012), but in fact meta-analysis, as a statistical summary technique, is part of a systematic review.  A systematic review, in contrast to a traditional narrative literature review, requires a clearly formulated research question, an extensive literature search that ideally includes relevant unpublished research findings, transparent study inclusion and exclusion criteria, a quantitative synthesis of the data (normally by a meta-analysis), and interpretation of the results (Borenstein et al. 2009: xxiii, Centre for Evidence Based Conservation 2012). A systematic review differs substantially from a narrative review in its transparency and replicability.  Declaring a less powerful summary analysis technique to be a meta-analysis could therefore result in a form of deceptive packaging.  Precise and unambiguous usage of the term would help in avoiding misinterpretations among scientists, by the public and by decision makers.  The four requisite steps in the first rating included: (1) generating an effect size metric based on continuous data, binary data or correlations; (2) weighting effect sizes by sample size or precision; (3) pooling of effect sizes into a summary effect or reasoning against pooling (e.g., due to high variation between effect sizes); (4) calculating confidence intervals for each effect size and the summary effect.  In the second round, where we evaluated if these articles also interpreted and discussed the results in a broader context e.g. looking at heterogeneity/variability – quantifying it, and exploring it by considering explanatory variables. Also make forest plot (to visualise results).  In our evaluation, we could not address the quality of the raw data, if they were adequate for the question being asked, if the generation of effect sizes had been done correctly.  If a study in our evaluation received the rating score 0, we do not at all state that it is a poorly conducted study or applying flawed statistics. We only argue that this study, claimed to be a meta-analysis, is not a meta-analysis according to the standard literature from the medical sciences.  We are aware of the many differences between the Bayesian and the frequentist approach and found it difficult to apply the same set of criteria to both kinds of meta-analyses. We decided to base our rating on the classical frequentist approach, which was far more common among the evaluated articles.  Also, the vote-counting approach that utilizes p-values instead of effect sizes, is commonly confused with meta-analysis (Chalfoun et al. 2002, Na´jera and Simonetti 2010), although it has been pointed out several times that vote-counting is statistically problematic and may result in false conclusions.  Meta-analysis allows us to calculate the magnitude rather than the existence of an effect; an important difference when we want to know if e.g., an intervention improves the habitat for an endangered species by 20% or by 80%.  Moreover, meta-analysis offers the possibility to assess if effect sizes are homogeneous across studies (Higgins 2008). If the effect sizes vary across studies, i.e., if there is heterogeneity, the interpretation of results will be substantially different than in the case of consistent effect sizes, e.g., if the intervention improves the habitat for the endangered species consistently by 50% or within a range from 10% to 90%.  If variation between effect sizes is very high, the presentation of a summary effect might be inadequate. Subgroup analyses and meta-regression can help to explain existent heterogeneity by comparing the effect size between different subgroups and explore the relationship between variables and effect sizes, respectively.  In the face of the high complexity and heterogeneity in natural systems, it is worrying that most meta-analysts do not consider or even mention the heterogeneity/variability of effect sizes in their meta-analysis.  A forest plot enables the reader to quickly assess the number of studies that form the summary effect, the precision of the included studies and the homogeneity/heterogeneity across effect sizes. |
| (Stewart, 2010) | 27/4/22 | **Meta-analysis in applied ecology – opinion piece**  **Abstract**  Vote counting and pooling unweighted averages are widespread despite the superiority of syntheses based on weighted combination of effects.  Meta-analyses are required to generalize in ecology, and to inform evidence-based decision-making.  **The use of meta-analysis**  Meta-analysis has multiple applications in applied ecology, but is particularly valuable for increasing power, exploring heterogeneity, identifying large-scale patterns and facilitating evidence-based decision making. Such applications are not possible with methods such as vote counting.  Such meta-analyses also illustrate the danger of over-reliance on information from a limited range of studies or sites.  **Criticisms of meta-analysis**  t is a legitimate concern that ecological meta-analysis may overestimate effect size (Stewart et al. 2009) because of publication bias induced by failure to publish negative studies  Ecological studies always differ and judgement is required about how similar they must be for pooled effects to be meaningful.  Ecological meta-analysis inevitably involves synthesis of studies measured on different spatio-temporal scales, requiring a focus on exploration of heterogeneity in almost all cases. Only by exploring heterogeneity can consistency, and hence generalizability, be empirically assessed.  Nevertheless, it remains true that all syntheses are constrained by the quality of available data and the standards of reporting in primary research.  **The future**  Establishment of a global register of environmental monitoring and primary research that requires submission of objectives and methods prior to the commencement of data collection would minimize publication bias.  Ecological meta-analysts have recognized the benefit of using hierarchical Bayesian models to explore complex data (e.g. Myers 2001), but their full potential remains untapped. |
| (Lortie et al., 2015) | 28/4/22 | **How to critically read ecological meta-analyses**  **Abstract**  The most important issues in critically assessing a meta-analysis initially include transparency, replicability, and clear statement of purpose by the authors.  Specific to ecology, more so than other disciplines, tests of the same hypothesis are generally conducted at different study sites, have variable ecological contexts (i.e., seasonality), and use very different methods. Clear reporting and careful examination of heterogeneity in ecological meta-analyses is thus crucial.  **Intro**  But the reviews in ecology are shifting from narrative descriptions to systematic reviews and meta-analyses. These forms of synthesis provide quantitative, replicable insights that can accelerate progress within ecology by providing the reader with a general framework of research completed to date and ideally insights into future directions.  Meta-analysis is an effective tool for synthesizing independent research efforts, comparing the relative success of treatments associated with groups of studies, testing whether mean treatment effects are significantly different than zero, and testing whether the effects are homogeneous or heterogeneous among and within groups or categories of studies.  Meta-analyses thus offer readers in ecology additional capacity to assess attributes of a hypothesis such as testability, generality, consistency, accuracy, and bias in the studies published on a topic at the time of analysis.  Failure of a meta-analysis to support a hypothesis may also be a product of the particular set of studies used in the synthesis.  Ongoing shift away from p-values is an important disciplinary transition.  In summary, a meta-analysis is not an experiment and does not test hypotheses but is a means to explore the strength of evidence associated with hypotheses. This is a critical clarification for ecology because there are often large collections of studies documenting only pattern, randomized controlled trials are not used, and even the exact replication of experiments or general protocols is unfortunately relatively infrequent.  **Reader guidelines**  The form of evidence used in the synthesis should be assessed to ensure that the review possesses the capacity to describe the ecological process or hypothesis of interest.  Publication bias is discussed at great length in the technical literature associated with meta-analysis and the reader should expect at least some examination to ensure that the studies included were not unduly skewed to only those that reported positive findings.  Hence, a useful meta-analysis should provide a list of studies for the reader, preferably in the main body of the paper and not in the appendix. This reporting approach also has the additional benefit of formally crediting the studies used in main citation list of the meta-analysis.  Meta-analytical statistical relevance does not necessarily map directly onto biological significance, and smaller mean effect sizes may be highly relevant and ecologically important in complex, diffuse natural systems.  Finally, the reader should also consider the breadth of confidence intervals when considering the magnitude of mean effect size because variation is such an important aspect of ecology in general. Forest plots are the most common and basic element of most, if not all, meta-analyses in ecology (Lortie et al., 2013). These plots depict both mean effect size estimates and the variances thereby providing a rapid assessment tool for the reader of patterns in relative variation.  The reader should also assess whether the authors explored potential dependencies between moderators.  Ecologists frequently use very different methods to test hypotheses because we measure populations, communities, ecosystem properties, and organisms. There is both methodological heterogeneity (i.e., qualitative) and statistical heterogeneity (Higgins and Green, 2011, Higgins et al., 2003). The reader should expect some treatment and discussion of patterns in both forms, and a common statistical metric, I2 , which ranges from 0% to100% has accepted benchmarks of 15% as low, 50% as moderate, and 75% as high (Higgins and Green, 2011).  A ‘great’ meta-analysis may focus more extensively on exploring the variation between studies and associated implications (Humphrey, 2011) and less on rejection of a method or hypothesis.  These include general guidelines associated with meta-analyses such as literature and scope (transparency, replicability, and purpose) and critical appraisal of the results (focus on strength of evidence and not statistical significance). |
| (Koricheva and Gurevitch, 2014) | 28/4/22 | **Uses and misuses of meta-analysis in plant ecology**  We found many cases of imprecise and inaccurate usage of the term ‘meta-analysis’ in plant ecology, particularly confusion between meta-analysis and vote counting and incorrect application of statistical techniques designed for primary studies to meta-analytical data, without recognition of the violation of statistical assumptions of the analyses.  The wealth of data available in ecology can threaten to overwhelm our ability to process it in an objective, unbiased manner, and one of the most basic contributions of meta-analysis is helping the researcher to make sense of and generalize from this information.  While it is possible to test for the effects of some covariates within individual primary studies, meta-analysis also allows testing for covariates which are logistically difficult or impossible to test within a single empirical study, for example, due to study scale. These include comparisons of effects of different experimental designs as well as comparisons across taxa, ecosystems or latitudinal gradients.  However, it is controversial whether one can use research syntheses to make causal inferences.  Marvier (2011) has recently reviewed the use of meta-analysis to assess environmental risk associated with GM crops. She pointed out that if meta-analyses of large data bases of completed studies were to become a routine part of risk assessment, it would reduce the risk of single experiments capturing media attention and inappropriately alarming or falsely comforting the public and policymakers.  Furthermore, the use of effect sizes and confidence intervals provide estimates of the precision and magnitude of an effect, rather than focusing exclusively on null-hypothesis significance tests.  Finally, a very important application of systematic reviews and meta-analyses in any scientific field, including plant ecology, is identification of knowledge gaps.  The most commonly used metrics of effect size in plant ecology meta-analyses are those based on comparison of means, that is, response ratios and standardized mean differences (Hedges’ and Cohen’s d). Effect sizes based on binary data (odds and risk ratios) are seldom used in plant ecology, presumably because most variables or interest to plant ecologists are continuous rather than binary.    Effect sizes in meta-analyses are normally weighted by study precisions, most commonly by inverse of study variance, so that studies with high precision (lower variance) are weighted more heavily than studies with higher variance. Weighting is important because it increases the precision of the mean effect estimate and the power of the tests and improves sampling distribution of the test statistics (Gurevitch & Hedges 1999). However, a quarter of meta-analyses in plant ecology (26%) used unweighted meta-analyses, largely because they reported that estimates of variance (standard deviations and standard errors) were not available in primary studies.  Alternatively, effects of several moderators and interactions between them can be tested in Proc Mixed in SAS, in the new open-access ecological meta-analysis package OpenMEE (Dietz et al. 2013) or in the metafor package in R (Viechtbauer 2010). So far less than a quarter of meta-analyses in plant ecology use multifactorial models to test for multiple moderators (Table 1).  One-third of meta-analyses (33%) in plant ecology do not specify the statistical model used. Fixed- and random-effect models make different assumptions about the data, and it is not uncommon for the outcomes of meta-analysis to differ depending on the model used.  However, the majority (61%) of meta-analyses in plant ecology did not include any tests for publication bias (Table 3) or mention the term ‘publication bias’ in the paper.  Among the remaining third of meta-analyses which did test for publication bias, the majority used funnel plots (scatter plots of effect sizes vs. sample size or variance). However, funnel plots are an inaccurate and unreliable method for assessing publication bias. Nakagawa & Santos (2012) recommended the use of a modification of the funnel plots which has better properties (Peters et al. 2008). Jennions & Møller (2002a) tested for publication bias in ecological meta-analyses by using ‘trim and fill’ method (Duval & Tweedie 2000), which allows one not only to test, but also to adjust for publication bias.  Sensitivity analyses should always be conducted to test the robustness of the findings of meta-analyses. E.g. would the results differ if MA was conducted in a different year?  In addition, we recommend that each meta-analysis is accompanied by PRISMA flow diagram (template is available at http://prisma-statement.org/statement.htm), which shows how studies used in meta-analysis were identified, screened and assessed for eligibility, and indicates the number of studies excluded at each step as well as the reasons for exclusion, as well as by a list of eligible studies which were excluded, with reasons for exclusions.  But in the methods explain that they ‘used a vote-counting technique because few of the compiled studies were suitable for the calculation of effect size’. This is a common justification for use of these flawed statistical techniques.  Hence, methodological shortcomings of meta-analyses in plant ecology are not unique to this field and could be easily rectified if plant ecologists become more vigilant both as research synthesists and peer reviewers. |
| (Noble et al., 2017) | 28/4/22 | **Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses**  **Abstract**  Nonindependence can affect two major interrelated components of a meta-analysis: (i) the calculation of effect size statistics and (ii) the estimation of overall meta-analytic estimates and their uncertainty.  Using sensitivity analyses are extremely important in assessing the impact of nonindependence.  **Intro**  They can, for example, inform policy, permit researchers to revise and refine current theoretical paradigms, establish more rigorous empirical tests of theory and help determine new research directions.  The importance of moderator variables (predictor variables in a normal linear model) in explaining effect size heterogeneity.  Moderator variables can also be included in analyses to understand whether hypothesized differences between studies (or effect sizes) explain heterogeneity (variance) among effects (using a so-called meta-regression model).  Ideally, a meta-analysis would involve a single effect size estimate being derived for each study, making every effect size within the meta-analysis statistically independent. However, for meta-analyses in ecology and evolution, it is often the case that effect size estimates are related to each other (i.e. are correlated) at various hierarchical levels, possibly because they come from the same study, are derived through comparisons with the same control group or are from correlated traits.  We argue that sensitivity analyses are an integral part of the solution to deal with problems associated with nonindependence.  Publication bias analysis, a type of sensitivity analysis, is already commonly used to explore the impact of missing unpublished studies.  Despite their importance, sensitivity analyses are seldom utilized to deal with problems of nonindependence, even though they provide greater confidence in meta-analysis results.  Nonindependence will artificially inflate sample size, increasing the magnitude of the denominator in variance equations and, thus, decreasing the sampling error variance for an effect size.  Log response ratio effect size is less susceptible to non-independence than standardized mean difference.  This is because meta-analysts can choose the fixed-effect model over the random-effects model when little to no heterogeneity among effect sizes exists (e.g. tau 2 = 0 and I 2 = 0). It is notable, however, that Senior et al. (2016) recently showed total heterogeneity in ecological and evolutionary meta-analyses to be very high on average (~92%), which indicates that random-effects models are more appropriate for typical meta-analyses in ecology and evolution.  We then use effective sample sizes, rather than the total sample size, in the calculations of effect size statistics to deal with nonindependence.  Multilevel meta-regression models have developed rapidly and can now deal with many sources of nonindependence, while allowing one to explore moderator variables to test a rich set of methodological and biological hypotheses explaining variation among effect sizes. Many of these models can be run in widely available software packages including METAFOR (Viechtbauer 2010). |
| (Borenstein et al., 2010) | 28/4/22 | **A basic introduction to fixed-effect and random-effects models for meta-analysis**  **Abstract**  The models represent fundamentally different assumptions about the data. The selection of the appropriate model is important to ensure that the various statistics are estimated correctly.  **Intro**  Under the fixed-effect model we assume that there is one true effect size that underlies all the studies in the analysis, and that all differences in observed effects are due to sampling error. While we follow the practice of calling this a fixed-effect model, a more descriptive term would be a common-effect model.  By contrast, under the random-effects model we allow the true effect sizes to differ—it is possible that all studies share a common effect size, but it is also possible that the effect size varies from study to study. For example, the effect size might be higher (or lower) in studies where the participants are older, or more educated, or healthier than in other studies, or when a more intensive variant of an intervention is used.  **Motivating example**  Under the random-effects model the confidence interval will always be wider and the weights will always be more similar to each other than under the fixed-effect model.  **Deriving the combined effect**  Studies with a precise estimate of the population effect size (a low variance) are assigned more weight, while studies with a less precise estimate of the population effect size (a high variance) are assigned less weight. This scheme is used for both the fixed-effect and random-effects models.  Where the models differ is in what we mean by a precise estimate, or (more correctly) how we define the overall study error variance. As outlined above, under the fixed-effect model there is one level of sampling (we sample subjects within the college), and therefore one source of variance. By contrast, under the random-effects model there are two levels of sampling (we sample colleges from the population of colleges and then students within each college), and therefore two sources of variance.  **Differences between fixed-effect and random-effects meta-analyses**  While many researchers find it intuitive that the random-effects model yields a less precise estimate of the combined effect than does the fixed-effect model, some find it surprising that the random-effects model also yields a different estimate of the combined effect itself. Indeed, while the fixed-effect model is estimating a common mean and the random-effects model is estimating the grand mean, one might expect the two estimates, based on the same set of study means, to be the same. In fact, though, the estimate of the combined effect is almost always different under the two models (provided, of course, that Tau 2 is not zero). The reason is that the combined mean is computed as the weighted mean of the effect size in each study, and the weights are different under the two models.  Under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate. This means that we cannot discount a small study by giving it a very small weight (the way we would in a fixed-effect analysis). The estimate provided by that study may be imprecise, but it is information about an effect that no other study has estimated. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed-effect analysis). Our goal is to estimate the mean effect in a range of studies and we do not want that overall estimate to be overly influenced by any one of them. This is implemented in the weighting scheme when the weights are based on the within-study variance plus a constant (Tau 2, the between-study variance). Because Tau 2 is a constant, it reduces the relative differences among the weights, which means that the relative weights assigned to each study are more balanced under the random-effects model than they are under the fixed-effect model.  The random-effects model allows that there may be a distribution of true effects. It follows that the first step in the analysis should be to estimate the amount of variation and then use this to inform the direction of the analysis. If the variation is trivial, then we would focus on reporting the mean and its confidence interval. If the variation is non-trivial, then we might want to address the substantive implications of the variation, but the mean might still be useful as a summary measure. By contrast, if the variation is substantial, then we might want to shift our focus away from the mean and toward the dispersion itself. For example, if a treatment reduces the risk of mortality in some studies while increasing the risk of mortality in others (and the difference does not appear to be due to estimation error), then the focus of the analysis should not be on the mean effect. Rather, it should be on the fact that the treatment effect differs from study to study. Hopefully, it would be possible to identify reasons (differences in the study populations or methods) that might explain the dispersion.  If number of studies is very small, option C is to perform a Bayesian meta-analysis, where the estimate of tau 2 is based on data from outside of the current set of studies. This is probably the best option, but the problem is that relatively few researchers have expertise in Bayesian meta-analysis. Additionally, some researchers have a philosophical objection to this approach.  The strategy of starting with a fixed-effect model and then moving to a random-effects model if the test for heterogeneity is significant relies on a flawed logic and should be strongly discouraged. |
| (Anzures‐Cabrera and Higgins, 2010) | 28/4/22 | **Graphical displays for meta-analysis: An overview with suggestions for practice**  **Abstract**  We start with forest plots and funnel plots, and proceed to Galbraith (or radial) plots, L’Abbé (and related) plots, further plots useful for investigating heterogeneity, plots useful for model diagnostics and plots for illustrating likelihoods and Bayesian meta-analyses.  **Forest plots**  Forest plots, also known as confidence interval plots, are probably the most familiar method for presenting results of meta-analyses. A forest plot displays effect estimates and their confidence intervals for each study and, usually, the meta-analysis.  Each study is represented by a square at a point estimate of effect and a horizontal line extending either side of the block to depict a 95% confidence interval. The area of the block is proportional to the weight assigned to that study in the meta-analysis.  Forest plots may include the result of the overall effect from a meta-analysis, normally at the bottom of the graph, and often using a diamond to distinguish it from the individual studies. It is common to plot the effect measure on the horizontal axis, in contrast to the usual convention of plotting dependent variables on the vertical axis.  A reference line should be drawn at the position of no effect, if such a null value is relevant.  **Funnel plots**  A funnel plot is a scatter plot of effect estimate against a measure of precision.  Funnel plots are used primarily as a visual aid for detecting bias or heterogeneity.  An asymmetric funnel indicates a relationship between effect size and precision in the studies at hand. This is most likely to be due to reporting bias.  Some methods for identifying or addressing potential publication bias are based on the funnel plot. Regression tests focus on the slope of a straight-line fit to points on a funnel plot using standard error as a measure of precision [33, 41], and the trim-and-fill method ‘trims’ studies from one side of the plot and ‘fills’ them in on both sides of the plot, in an attempt to deduce the results from both available and unavailable studies under a particular type of publication bias.  **Other plots**  Galbraith plots might be particularly appropriate when there are more studies than can comfortably be displayed on a forest plot.  L’Abbé plots (and related ROC plots) are appropriate only for studies comparing two groups, but for this they are very informative.  **Plots for describing the relationship between effect estimates and covariates (meta-regression)**  Often a source of heterogeneity can be summarized as a study-level covariate, i.e. some varying characteristic of the studies.  Meta-regression is the statistical analysis of the association between effect size and the value of one or more study-level  covariates. A simple scatter plot with the covariate along the horizontal axis and the treatment effect along the vertical  axis provides a convenient visual impression of the relationship.  **Plots useful for sensitivity analysis and diagnostics**  It is well known that the heterogeneity variance in a random-effects meta-analysis is imprecisely estimated  when the number of studies is not large.  **Bayesian meta-analyses**  In Bayesian meta-analyses, likelihoods are combined with prior distributions to produce posterior distributions for each unknown quantity. An effective way to convey the impact of prior distributions is to plot them along with likelihoods and posterior distributions.  **Discussion**  Our recommendation that the forest plot is a generally useful graph concurs with other recent commentaries [6, 7], and we propose that it continue to be the first choice when it is viable. |
| (Senior et al., 2016) | 28/4/22 | **Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications**  **Abstract**  Together with estimating overall effect magnitudes, meta-analyses estimate differences between effect sizes via heterogeneity statistics. It is widely hypothesized that heterogeneity will be present in ecological/evolutionary meta-analyses due to the system-specific nature of biological phenomena.  We reviewed 700 studies, finding 325 that used formal meta-analysis, of which total heterogeneity was reported in fewer than 40%.  We used second-order meta-analysis to collate heterogeneity statistics from 86 studies. Our analysis revealed that the median and mean heterogeneity, expressed as I2, are 84.67% and 91.69%, respectively. These estimates are well above “high” heterogeneity (i.e., 75%), based on widely adopted benchmarks. We encourage reporting heterogeneity in the forms of I2 and the estimated variance components (e.g., τ2) as standard practice.  **Intro**  The second, random-effects meta-analysis (REMA), assumes that effects from different studies truly differ after accounting for sampling; this “true” variance is termed heterogeneity.  Conceptually, I2 may be thought of as the percentage of variance between effect sizes that cannot be attributed to sampling error.  **Discussion**  Our results are the first systematic quantification of this expectation, suggesting that mean I2 in these fields is between 83.61% and 92.07%, and I2 = 0 in only 4.65% of studies. This stands in contrast to medical fields where I 2 = 0 in 30–55% of studies, depending on the outcome being assessed. A potential shortfall of our analyses is that our estimate of average I2 may be upwardly biased because heterogeneity is unreported when low. However, 122 studies that failed to report heterogeneity did report significant meta-regression results, suggesting heterogeneity was present.  We also advocate the use of MLMA, also referred to as mixed-effects models. These models allow heterogeneity to be partitioned among random factors fitted to the model. For instance, in studies where effects come from different species, it is possible to estimate the variance between effect sizes that is attributable to evolutionary relationships as I2. |
| (Harrison, 2011) | 28/4/22 | **Getting started with meta-analysis**  **Summary**  Errors or omissions made at the planning stage can create weeks of extra work.  **Intro**  First, if a number of attempts have been made to measure the effect of one variable on another, then meta-analysis provides a method to calculate the mean effect of the independent variable, across all attempts.  Secondly, meta-analysis allows us to measure the amount of experimentally-induced change in the dependent variable across studies and to attempt to explain this variability using defined moderator variables.  Tests which are analogous to analysis of variance (anova) and weighted regression can then be applied to the population of effect sizes to identify dependent variables that explain a significant amount of variation between studies.  P-values reflect a dichotomous question (is the observed pattern of data likely to be due to chance, or not?) not an open-ended one (how strong is the pattern in the data?).  A formal meta-analysis ameliorates the problems with vote-counting. Not only are effect sizes more informative, they also represent continuous variables that can be combined and compared. Meta-analysis is more clearly needs driven and evidence based.  It should be noted that, like any statistical method, meta-analyses are only as good as the data used and can still suffer from both type I and type II errors.  The keys to making meta-analysis as stress-free as possible are organization and planning. In particular, your list of potential moderator variables (step 4) should be clearly defined before you begin: it is far preferable to produce a data base which includes information that you later decide not to use, than to produce a data base that excludes a variable you later decide to explore, as the latter may require a second (or third, or fourth) trawl through your collection of studies to extract the necessary information.  Continuous or ordinal data from two or more groups. Data in this form are exemplified by treatment vs. control group comparisons and are generally presented and analysed using averages and measures of variance (mean and standard deviation, median and interquartile range, etc). In such cases, a measure of the difference between the group means is an appropriate effect size. The raw difference in means can be standardized by the pooled standard deviation; two commonly-used measures of standardized mean difference are Cohen’s d and Hedges’ g: these differ in the method used for calculating the pooled standard deviation but it should be noted that the d and g notation has been used interchangeably by some authors.  Alternatively, when the data measure rates of change in independent groups (e.g. plant growth response in normal or elevated CO2, body mass gain after supplementary feeding), the response ratio can be used. This measures the ratio of the mean change in one group to the mean change in the other. Like the standardized mean difference, it takes the standard deviations in the two groups into account. The response ratio is generally log-transformed prior to meta-analysis in order to linearize and normalize the raw ratios.  Continuous or ordinal data which are a response to a continuous or ordinal independent variable. Any data which are analysed using correlation or regression fall into this category. In this case, the correlation coefficient itself can be used as a measure of effect size. Whichever type of correlation coefficient we use, Fisher’s z transformation is generally applied in order to stabilize the variance among coefficients prior to meta-analysis.  Binary response data. Data that take the form of binary yes/no outcomes, such as nest success or survival to the next breeding season, are generally analysed using logistic regression or a chi-squared test. In this case, an appropriate measure of effect size is given by calculating the risk ratio or odds ratio. These types of effect size have been very rarely used in ecology and evolution, though they are common in medical research.  It is possible to test whether our set of effect sizes shows more heterogeneity than would be expected due to sampling error alone by calculating the Q (often called QTotal) statistic we could assume that the variation between studies is random. In this case, we can use mathematical methods for estimating the random effects variance component, add this component to the variance statistic for each individual effect size and re-calculate the inverse variance weights in order to calculate a random-effects version of ES. Thirdly, we could assume that heterogeneity stems from both fixed and random sources. In this case, we can run meta-analytic models to test for the effects of fixed moderator variables, using inverse variance weights that have been adjusted for the estimated random effects component. This produces a mixed-effects model and a more robust test for the significance of moderator variables.   * Borenstein, 2010 says don't do fixed effect first and then move to random if seems appropriate – think about it first   A funnel plot of effect size vs. study size is one method of identifying publication bias in our set of studies: all things being equal, we would expect that the effect sizes reported in a number of studies should be symmetrically distributed around the underlying true effect size, with more variation from this value in smaller studies than in larger ones. |
| (Nakagawa and Poulin, 2012) | 28/4/22 | **Meta-analytic insights into evolutionary ecology: an introduction and synthesis**  **Abstract**  Although it is not without faults, we find that meta-analysis usually outperforms qualitative reviews with respect to testing hypotheses, identifying sources of heterogeneity among primary results, assessing publication bias, and even generating new hypotheses and future research directions.  **Meta-analysis and narrative reviews**  Whether quantitative or qualitative, the main role of a review piece is to provide an up-todate overview of the state of knowledge in an area of study.  Strictly speaking, a narrative review cannot be used to test a hypothesis. However, authors of such reviews have often attempted hypothesis testing by tallying up significant (positive/ supportive) and non-significant (negative/unsupportive) studies, an approach known as ‘vote-counting’. Early advocates of meta-analysis severely criticized vote-counting mainly because it completely ignores study quality in terms of sample sizes and methodologies, often leading to erroneous conclusions.  Meta-analysis, in contrast, can be used for testing hypotheses in as rigorous a way as empirical studies. The ability to detect small effects (sensu Cohen 1988), which any single study cannot reliably detect, is probably its most obvious strength.  The kinds of questions that one can ask with meta-analysis can be extended by the use of meta-regression, which is basically meta-analysis with predictors (often called moderators in the meta-analytic literature; reviewed in Thompson and Higgins 2002). Meta-regression is usually applied once meta-analysis detects statistically significant heterogeneity.  Obviously, meta-analysis is not suitable for appraising new hypotheses, for which no or few empirical studies exist.  The clearest capability difference between the two types of reviews may be that meta-analysis comes with a range of tools that can detect and even correct publication bias. Despite this advantage, Nakagawa and Santos (2012) report that ecologists and evolutionary biologists often do not perform the analyses required to deal with publication bias. Such poor practice clearly needs to change in the future.  We argue, however, that meta-analysis is more than capable of doing the same, or even a better, job of generating new hypotheses and future directions than narrative pieces, although we feel that, in our field, this functional aspect of meta-analysis is underutilized.  Nakagawa and Santos (2012) review methodological advancements in meta-analytic methods, especially contributions made by ecologists and evolutionary biologists. They show that new meta-analytic methods, which combine phylogenetic comparative analysis and mixed-effects modelling, have resolved a number of problems posed by meta-analytic data typical of ecology and evolution (Hadfield and Nakagawa 2010); such data have multiple correlated structures due to temporal and spatial replications, as well as phylogenetic relatedness (i.e. statistical non-independence).  This persistent issue perhaps needs to be resolved by a collective action of journals in ecology and evolution imposing clear guidelines of statistical reporting.  We believe that the PRISMA statement, which has been developed for the health sciences, can potentially be adopted for meta-analyses in ecology and evolution, and that it should be routinely used.   * Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) consists of a 27-item check list and a 4-stage flowchart (see Fig. 1) intended for systematic reviews and meta-analyses for medical and clinical sciences. |
| (Cadotte et al., 2012) | 28/4/22 | **Gauging the impact of meta-analysis on ecology**  **Abstract**  For the quantitative assessment, we conducted an analysis of 240 published meta-analyses to examine trends in ecological meta-analyses. Our examination shows that publication rates of meta-analyses in ecology have increased through time, and that more recent meta-analyses have been more comprehensive, including more studies and a greater temporal range of studies.  **Intro**  Conversely, meta-analyses explicitly account for sample error and sample size when comparing effect sizes and thus have the power to detect significant differences in the pooled data even when individual datasets fail to detect significant effects.  The increasing quantity of data over time and the development of methods and infrastructure to archive and share data (Jones et al. 2006; Michener 2006) increase data accessibility, further reinforcing the push to do meta-analyses.  **Examples of meta-analyses that changed ecological understanding**  The paper by Root et al. (2003) again reveals that the true power of a meta-analysis is not necessarily to test new hypotheses, but rather to definitively show that certain patterns are generalizable. The meta-analysis by Root et al. (2003) reinforces how meta-analyses can utilize data that have a limited geographical extent to search for broader generalities. |
| (Page et al., 2021) | 28/4/22 | **The PRISMA 2020 statement: An updated guideline for reporting systematic reviews**  **Abstract**  The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, published in 2009, was designed to help systematic reviewers transparently report why the review was done, what the authors did, and what they found.  **Intro**  The PRISMA 2009 statement comprised a checklist of 27 items recommended for reporting in systematic reviews and an “explanation and elaboration” paper.  The PRISMA 2020 statement consists of a 27-item checklist, an expanded checklist that details reporting recommendations for each item, the PRISMA 2020 abstract checklist, and revised flow diagrams for original and updated reviews.  **The PRISMA 2020 statement**  The PRISMA 2020 statement has been designed primarily for systematic reviews of studies that evaluate the effects of health interventions, irrespective of the design of the included studies. However, the checklist items are applicable to reports of systematic reviews evaluating other interventions (such as social or educational interventions), and many items are applicable to systematic reviews with objectives other than evaluating interventions (such as evaluating aetiology, prevalence, or prognosis).  **Discussion**  Complete reporting allows readers to assess the appropriateness of the methods, and therefore the trustworthiness of the findings.  We anticipate that the PRISMA 2020 statement will benefit authors, editors, and peer reviewers of systematic reviews, and different users of reviews, including guideline developers, policy makers, healthcare providers, patients, and other stakeholders. |
| (Lajeunesse, 2016) | 28/4/22 | **Facilitating systematic reviews, data extraction and meta-analysis with the METAGEAR package for R**  **Summary**  The R package ecosystem is rich in tools for the statistics of meta-analysis. However, there are few resources available to facilitate research synthesis as a whole.  Current functionalities of METAGEAR include the following: an abstract screener GUI to efficiently sieve bibliographic information from large numbers of candidate studies.  **Intro**  A systematic review is first used to carefully plan and implement a strategy to locate studies, assess their eligibility and extract relevant research findings in a repeatable and transparent way (Khan et al. 2003; Pullin & Stewart 2006). A meta-analysis is then used to statistically model sources of variability within and between studies with the aim to quantitatively weight and aggregate their findings.  Here, I introduce METAGEAR, an R package aimed at improving the reproducibility of systematic reviews and meta-analysis. It is a comprehensive toolbox that spans the entire research synthesis taxonomy, and supports a large diversity of functionalities to help screen abstracts, download articles, extract data from figures and model dependencies within and among effect sizes, and various other tools to improve the quality of statistics used in meta-analysis.  **Screening abstracts, delegating tasks and retrieving articles**  Identifying candidate studies is one of the first key stages in a systematic review. However, literature searches with bibliographic data bases such as Web of Knowledge or Google Scholar will generate thousands of study references – this is unavoidable given that search terms should aim to yield the most inclusive results (see Curtis et al. 2013). The challenge here is that the title and abstract of all of these references need to be screened as a first attempt to sift those that are relevant for the synthesis project.  Once all study references have been screened, the next stage is to retrieve their full texts. This is tedious when there are hundreds or thousands of studies to collect. Here, METAGEAR can be used to automate the download of these studies.  The plot\_PRISMA() function can be used to quickly generate and update flow diagrams as each phase or stage gets completed.  **Extracting and imputing data**  Extracting data from studies to calculate effect sizes (a common currency that quantitatively summarizes the magnitude and sign of study outcomes) is by far the most difficult and time-consuming stage of the entire research synthesis process. It is a manual activity where researchers must read and interpret the text/tables/figures of each study with the goals of unearthing all the quantitative data needed to calculate effect sizes.  The function figure\_scatterPlot() can then be used to extract the data from the image.  **Meta-analysis and the reproducibility of syntheses**  Ecologists and evolutionary biologists have historically relied heavily on METAWIN. This popularity is also remarkable considering the availability of more sophisticated meta-analysis software, such as COMPREHENSIVE META-ANALYSIS (reviewed by Schmid et al. 2013) and the free OPENMEE (Wallace et al. 2015), as well as the METAFOR R package which now dominates the statistical analyses of many disciplines (Viechtbauer 2010). METAWIN is no longer maintained (last updated in 2007).  **Conclusions and prospectus**  There is still room to grow in terms of developing tools for research synthesis – such as automating PDF annotations, text-mining tools to enhance data retrieval and extraction, methods to estimate statistical power (Lajeunesse 2013b), multiple-imputation tools (Lajeunesse 2013a; Ellington et al. 2015) and better approaches to assess publication bias. |
| (Duval and Tweedie, 2000) | 29/4/22 | **A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis**  We develop a simple rank-based data augmentation technique, formalizing the use of funnel plots, to estimate and adjust for the numbers and outcomes of missing studies. |
| (Assink and Wibbelink, 2016) | 29/4/22 | **Fitting three-level meta-analytic models in R: A step-by-step tutorial**  **Abstract**  Strong method for dealing with dependency of effect sizes.  Relatively unknown.  The rma.mv function of the metafor package provides an easy and flexible way of applying a multi-level structure to meta-analytic models in R.  Further, the multilevel meta-analytic models can be easily extended so that the potential moderating influence of variables can be examined.  **Intro**  An important requirement in traditional univariate meta-analytic approaches is that there is no dependency between effect sizes in the data set that is to be analyzed (e.g., Rosenthal, 1984). If there is dependency between effect sizes (i.e., effect sizes are correlated), there is overlap in information to which correlated effect sizes are referring to. In this way the available information is ‘inflated’ and consequently leads to an overconfidence in the results of a meta-analysis.  Becomes possible to deal with dependency of effect sizes in such a way that a research synthesist can extract all relevant effect sizes from each primary study without needing to reduce the number of effect sizes in any way. By performing the analyses using all relevant effect sizes, all information can be preserved and maximum statistical power can be achieved.  Applying a three-level structure to a meta-analytic model is a better approach for dealing with dependency of effect sizes than the methods just mentioned. This three-level meta-analytic model considers three different variance components distributed over the three levels of the model: sampling variance of the extracted effect sizes at level 1; variance between effect sizes extracted from the same study at level 2; and variance between studies at level 3. In short, this model allows effect sizes to vary between participants (level 1), outcomes (level 2), and studies (level 3).  If there is evidence for heterogeneity in effect sizes, moderator analyses can be conducted to test variables that may explain within-study or between-study heterogeneity. For these analyses, the three-level random effects model can easily be extended with study and effect size characteristics, making the model a three-level mixed effects model.  The rma.mv function is part of this package and makes it possible to fit multilevel meta-analytic models that can be extended by including moderators.  Goes through example/tutorial on how to do it. |
| Christina Raw's notes on meta-methods | 29/4/22 | **Meta-methods**  Must keep and provide a detailed record of the methods applied on each step to guarantee a sufficient level of reproducibility.  What is the research question? Apply the PICO framework: define the Population, Intervention, Comparator, and the Outcome of interest. Try and narrow this down as much as possible.  Include descriptions of the PICO elements in the protocol.  Develop the optimal string of search terms. For each iteration, record the string of search terms used, modifications made, number of hits and number of relevant papers in the first 20 abstracts.  Develop a systematic review protocol that will be included in supplementary material.  ROSES systematic review protocol includes the following elements:   * title * objective of the overview * definitions of the question components * search strategy * search string * bibliographic databases used * inclusion criteria * screening strategy * critical appraisal * meta-data extraction and coding strategy * data extraction strategy * approaches to missing data * effect modifiers * risk of bias * author contribution and competing interests   Download the studies retrieved by the data base using your optimal string of search terms (make sure to remove duplicates).  Screen list of studies following inclusion/exclusion criteria. Could use MetaGear package. Provide flow diagram that documents the screening process.  Extract the data  Build excel data extraction spreadsheet where extracted data will be recorded. Could use metaDigitise package to extract data from figures.  Data includes information about the study (e.g., study ID, authors, title, etc.), quantitative variables that will be used to calculate the effect size, and qualitative variables that will be used for moderator analysis.  Extract data from relevant papers and record in spreadsheet.  Assess quality of papers and assign score using sensitivity analysis.  Perform meta-analysis along with sensitivity analysis, and bias assessment e.g. assess publication bias with funnel plot.  It took me a total of 33 h 40 ' to extract data from 13 papers = ~ 2 h 30' / paper (mean)  This means that, in a 9 to 5 workday, you can extract a mean of ~ 3 papers / day |

References

ANZURES‐CABRERA, J. & HIGGINS, J. P. 2010. Graphical displays for meta‐analysis: an overview with suggestions for practice. *Research synthesis methods,* 1**,** 66-80.

ARNQVIST, G. & WOOSTER, D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution,* 10**,** 236-240.

ASSINK, M. & WIBBELINK, C. J. 2016. Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology,* 12**,** 154-174.

BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. & ROTHSTEIN, H. R. 2010. A basic introduction to fixed‐effect and random‐effects models for meta‐analysis. *Research synthesis methods,* 1**,** 97-111.

CADOTTE, M. W., MEHRKENS, L. R. & MENGE, D. N. 2012. Gauging the impact of meta-analysis on ecology. *Evolutionary Ecology,* 26**,** 1153-1167.

CÔTÉ, I. M. & REYNOLDS, J. D. 2012. Meta-analysis at the intersection of evolutionary ecology and conservation. *Evolutionary Ecology,* 26**,** 1237-1252.

DIRZO, R., YOUNG, H. S., GALETTI, M., CEBALLOS, G., ISAAC, N. J. & COLLEN, B. 2014. Defaunation in the Anthropocene. *science,* 345**,** 401-406.

DUVAL, S. & TWEEDIE, R. 2000. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the american statistical association,* 95**,** 89-98.

GUREVITCH, J. & HEDGES, L. V. 1999. Statistical issues in ecological meta‐analyses. *Ecology,* 80**,** 1142-1149.

GUREVITCH, J., KORICHEVA, J., NAKAGAWA, S. & STEWART, G. 2018. Meta-analysis and the science of research synthesis. *Nature,* 555**,** 175-182.

HARRISON, F. 2011. Getting started with meta‐analysis. *Methods in Ecology and Evolution,* 2**,** 1-10.

HEDGES, L. V., GUREVITCH, J. & CURTIS, P. S. 1999. The meta‐analysis of response ratios in experimental ecology. *Ecology,* 80**,** 1150-1156.

JÄHNIG, S. C., BARANOV, V., ALTERMATT, F., CRANSTON, P., FRIEDRICHS‐MANTHEY, M., GEIST, J., HE, F., HEINO, J., HERING, D. & HÖLKER, F. 2021. Revisiting global trends in freshwater insect biodiversity. *Wiley Interdisciplinary Reviews: Water,* 8**,** e1506.

KORICHEVA, J. & GUREVITCH, J. 2014. Uses and misuses of meta‐analysis in plant ecology. *Journal of Ecology,* 102**,** 828-844.

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.

LORTIE, C. J., STEWART, G., ROTHSTEIN, H. & LAU, J. 2015. How to critically read ecological meta‐analyses. *Research synthesis methods,* 6**,** 124-133.

NAKAGAWA, S., NOBLE, D. W., SENIOR, A. M. & LAGISZ, M. 2017. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC biology,* 15**,** 1-14.

NAKAGAWA, S. & POULIN, R. 2012. Meta-analytic insights into evolutionary ecology: an introduction and synthesis. *Evolutionary Ecology,* 26**,** 1085-1099.

NAKAGAWA, S. & SANTOS, E. S. 2012. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology,* 26**,** 1253-1274.

NOBLE, D. W., LAGISZ, M., O'DEA, R. E. & NAKAGAWA, S. 2017. Nonindependence and sensitivity analyses in ecological and evolutionary meta‐analyses. Wiley Online Library.

PAGE, M. J., MCKENZIE, J. E., BOSSUYT, P. M., BOUTRON, I., HOFFMANN, T. C., MULROW, C. D., SHAMSEER, L., TETZLAFF, J. M., AKL, E. A. & BRENNAN, S. E. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery,* 88**,** 105906.

SENIOR, A. M., GRUEBER, C. E., KAMIYA, T., LAGISZ, M., O'DWYER, K., SANTOS, E. S. & NAKAGAWA, S. 2016. Heterogeneity in ecological and evolutionary meta‐analyses: its magnitude and implications. *Ecology,* 97**,** 3293-3299.

STEWART, G. 2010. Meta-analysis in applied ecology. *Biology letters,* 6**,** 78-81.

VAN KLINK, R., BOWLER, D. E., GONGALSKY, K. B., SWENGEL, A. B., GENTILE, A. & CHASE, J. M. 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science,* 368**,** 417-420.

VETTER, D., RUECKER, G. & STORCH, I. 2013. Meta‐analysis: A need for well‐defined usage in ecology and conservation biology. *Ecosphere,* 4**,** 1-24.

VIECHTBAUER, W. 2010. Conducting meta-analyses in R with the metafor package. *Journal of statistical software,* 36**,** 1-48.

Christina Raw's notes on meta-methods