PLOS ONE

RESEARCH ARTICLE

A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Paula Pappalardo[1]*, **Chao Song**[2], **Bruce A. Hungate**[3], **Craig W. Osenberg**[4]

1 Smithsonian Environmental Research Center, Tiburon, California, United States of America, **2** State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems and College of Ecology, Lanzhou University, Lanzhou, Gansu, China, **3** Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, United States of America, **4** Odum School of Ecology, University of Georgia, Athens, Georgia, United States of America

* paulapappalardo@gmail.com

Citation: Pappalardo P, Song C, Hungate BA, Osenberg CW (2023) A meta-evaluation of the quality of reporting and execution in ecological meta-analyses. PLoS ONE 18(10): e0292606. https://doi.org/10.1371/journal.pone.0292606

Editor: Daniel de Paiva Silva, Instituto Federal de Educacao Ciencia e Tecnologia Goiano - Campus Urutai, BRAZIL

Received: September 26, 2022

Accepted: September 25, 2023

Published: October 12, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0292606

Data Availability Statement: All the data files and R code used to analyze and report data are provided in the Supporting information.

Abstract

Quantitatively summarizing results from a collection of primary studies with meta-analysis can help answer ecological questions and identify knowledge gaps. The accuracy of the answers depends on the quality of the meta-analysis. We reviewed the literature assessing the quality of ecological meta-analyses to evaluate current practices and highlight areas that need improvement. From each of the 18 review papers that evaluated the quality of meta-analyses, we calculated the percentage of meta-analyses that met criteria related to specific steps taken in the meta-analysis process (i.e., execution) and the clarity with which those steps were articulated (i.e., reporting). We also re-evaluated all the meta-analyses available from Pappalardo et al. [1] to extract new information on ten additional criteria and to assess how the meta-analyses recognized and addressed non-independence. In general, we observed better performance for criteria related to reporting than for criteria related to execution; however, there was a wide variation among criteria and meta-analyses. Meta-analyses had low compliance with regard to correcting for phylogenetic non-independence, exploring temporal trends in effect sizes, and conducting a multifactorial analysis of moderators (i.e., explanatory variables). In addition, although most meta-analyses included multiple effect sizes per study, only 66% acknowledged some type of non-independence. The types of non-independence reported were most often related to the design of the original experiment (e.g., the use of a shared control) than to other sources (e.g., phylogeny). We suggest that providing specific training and encouraging authors to follow the PRISMA EcoEvo checklist recently developed by O'Dea et al. [2] can improve the quality of ecological meta-analyses.

Introduction

Meta-analyses evaluate summary statistics from primary studies to obtain aggregate effects, assess the heterogeneity of those effects, and ascertain possible causes of the observed heterogeneity. For example, meta-analysis has been used to quantify the strength of density-dependence [3], to assess the response of ecosystems to climate change [4], and to evaluate the performance of different management strategies [5]. Through synthesis, meta-analysis not only advances basic ecological theory, but also facilitates the application of ecological data to inform environmental policy [6]. Moreover, meta-analysis can help identify knowledge gaps, and thus direct new research endeavors [7]. Along with these benefits, the number of published meta-analyses is rapidly increasing [8,9], due to increased data availability, and pressing ecological questions that require synthetic research.

Despite their importance and wide application, the quality of meta-analyses is highly variable [1,7,10,11]. If the quality of meta-analyses is poor, it is hard to know if "biological meta-analysis embodies 'mega-enlightenment', a 'mega-mistake', or something in between" [12]. One issue that can prevent readers from evaluating the overall quality of a published meta-analysis is the lack of details describing each step in the meta-analysis. We refer to this as reporting quality, which is the degree to which the meta-analysis explicitly reports the steps taken to conduct the meta-analysis, including details about the methods used to identify studies and extract data, the meta-analytic model, the number of effect sizes, and the sources of non-independence [7]. Good reporting quality means also that the meta-analysis provides the data used and describes each step of analysis in sufficient detail to replicate the results. Poor reporting quality hinders the readers from assessing if the meta-analysis was executed properly and if the results are reliable. Reporting quality it does not refer to whether those steps were the best available–only that the author(s) was explicit about the steps taken.

The quality of a meta-analysis also is affected by how well the study is implemented. We refer to this proper implementation as execution quality, which is the extent to which the analyses conform to expert recommendation. Examples of recommended execution steps are weighing effect sizes by study precision, testing for publication bias, quantifying heterogeneity in effect sizes, exploring temporal changes in effect size, controlling for phylogenetic non-independence (if applicable), and conducting sensitivity analyses [7].

New methodological guidelines specifically designed for ecology and evolutionary biology [PRISMA-EcoEvo, 2] provide authors, reviewers, and editors with a checklist of items with the goal of improving the overall quality of ecological meta-analysis. Wide adoption of these guidelines could greatly improve the quality of meta-analyses in ecology and evolutionary biology. Assessing the current compliance with recommended steps for reporting and execution in ecological meta-analyses and identifying places that need improvements can help guide the meta-analytic community towards more robust inference and reduce controversy.

In this paper, we reviewed the literature assessing the quality of ecological meta-analyses, collected new data to evaluate current practices, and highlight the areas that need more work. First, we compiled information from 18 studies in the last 20 years (between 2002 and 2022) that reviewed the quality of meta-analyses in ecology, evolution and related fields. These papers provided different insights on the compliance with different standards of reporting quality and with recommended execution steps that should be part of a meta-analysis. Second, we evaluated the recognition and treatment of non-independence for the ecological meta-analyses included in Pappalardo et al. [1], and extracted new data on quality criteria to compare it with the other reviews. Finally, we summarized the level of compliance for different quality criteria across these 18 previous meta-analysis reviews and the new data taken from papers reviewed in Pappalardo et al [1].

Funding: During the development of this project CWO received support from the National Science Foundation (DEB-1655426 and OCE-1851032); BH, PP and CWO were supported by a grant from the US Department of Energy, Office of Science, Biological and Environmental Research Program (DE-SC-0010632). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Methods

Literature search

To evaluate current practices when conducting and reporting ecological meta-analyses, we surveyed the literature for quantitative assessments on criteria previously identified as best practices in meta-analysis. These criteria fall into two broad categories: 1) execution (i.e., methodological issues related with best practices for data analysis), and 2) reporting (i.e., details describing each step of the meta-analysis and providing the data and information needed to allow for reproducibility). Both categories of criteria aim to ensure appropriate and reproducible results. Our list of criteria was informed by Koricheva & Gurevitch [7, Table 3] and the PRISMA EcoEvo checklist [2]. To find relevant papers, we first performed an exploratory search in Google Scholar, using combinations of keywords including "meta-analysis", "review", "quality", "ecology", "evolution". We then searched the Core Collection of the ISI Web of Science database including articles and reviews within the "Ecology", "Evolutionary Biology", "Biodiversity Conservation" and "Plant Sciences" categories (last search update on Sep 16, 2022). We used a search string for TOPIC as: (["meta-analyses" OR "metaanalyses" OR "meta analyses"] AND ["quality" OR "performance criteria"] AND ["reporting"]). The search resulted in 751 citations. We supplemented those with 11 articles published in the "Meta-analytic insights into evolutionary ecology" Special Issue of Evolutionary Ecology (2012, Volume 26, Issue 5), and 3 articles referenced by other articles or presented in scientific talks. The 765 papers that were obtained were then screened using the *metagear* [13] R package (additional details and R code in S1 Appendix) and based on titles and abstracts, this set was reduced to 61 papers. PDFs of the 61 papers were obtained and evaluated in more detail.

Of the 61 papers for which we screened the full text, the majority were excluded for not having quantitative data on the selected criteria (Fig 1). One paper was excluded for being out of scope. Two of the papers [14,15] overlapped considerably by evaluating many of the same studies in restoration ecology (S2 Appendix); Romanelli et al. [14] describe "many of the materials related to the dataset and methods used to collate evidence are similar to those presented in [15]". We checked their list of references and 92 citations were shared. To reduce non-

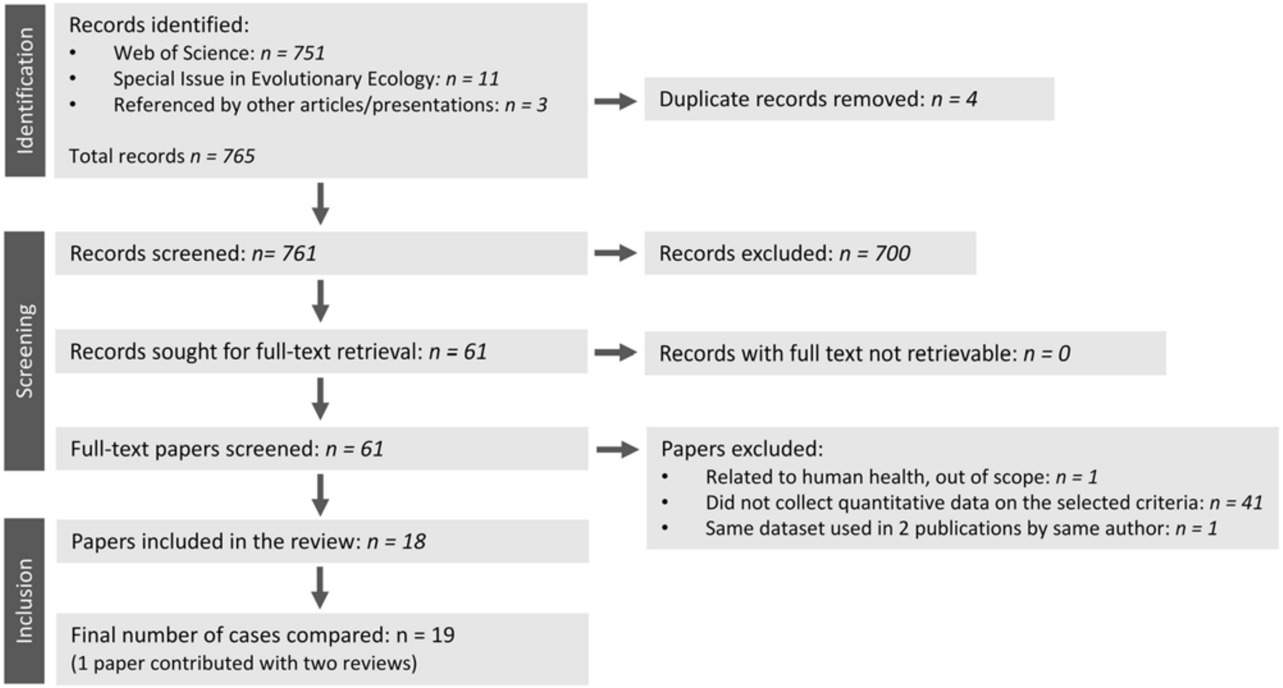

Fig 1. PRISMA diagram. The PRISMA diagram details the relevant literature sources identified, screened, excluded, assessed, and selected for the final analysis.

https://doi.org/10.1371/journal.pone.0292606.g001

Table 1. Compilation of 18 papers that reviewed the quality of reporting in ecological meta-analyses.

Publication	Review ID	Area	Time period	Number of meta-analysis papers reviewed
Archmiller et al. [18]	arch2015	molecular ecology	2003–2014	18
Beillouin et al. [19]	beil2022	biodiversity conservation, ecology; environmental sciences	2001–2020	217[a]
Cadotte et al. [8]	cado2012	ecology	1992–2008	240
Chamberlain et al. [20]	cham2012	ecology and evolutionary biology	1992–2010	56
Chaudhary et al. [21]	chau2010	ecology	1992–2006	188
Gates [10]	gate2002	ecology	1992–1998	29
Jennions et al. [22]	jenn2012	sexual selection	1996–2012	94
Koricheva & Gurevitch [7]	kori2014	plant ecology	1996–2013	322
Lodi et al. [16]	lodi2021_fe	freshwater ecology	1994–2017	114
	lodi2021_ee	ecology and evolution	1992–2014	86[b]
Nakagawa & Santos [23]	naka2012	ecology and evolution	2009–2011	100
Nakagawa et al. [24][c]	naka2022	ecology and evolutionary biology	2010–2019	102
Odea et al. [2]	odea2021	ecology and evolutionary biology	2010–2019	102
Pappalardo et al. [1]	papp2020	ecology, climate change	2013–2016	96
Philibert et al. [25]	phil2012	agronomy	2001–2011	73[d]
Roberts et al. [26]	robe2006	conservation, ecology, and environmental management	2003–2005	73
Romanelli et al. [15]	roma2021a	restoration ecology	2009–2019	63
Senior et al. [27]	seni2016	ecology and evolution	1992–2014	325
Vetter et al. [28]	vett2013	ecology and conservation biology	2002–2011	133 and 83[e]

In this table we provide details for each review paper included in our final analysis. The "Review ID" (first author initials plus publication year) was used to identify review papers in Figs 1 and 2, and tables and figures in the Supporting Information. "Area" indicates the subdiscipline(s) summarized in the review papers. Because Lodi et al. [16] reviewed the quality of meta-analyses in two topic areas, we distinguished them using the "Review ID" (lodi2021_fe and lodi2021_ee). Because the two Romanelli et al. reviews [14,15] were based on a similar set of meta-analyses, we included only roma2021a [16] in our final dataset. "Time period" gives the range of publication dates of the meta-analyses that were reviewed.

[a] For Beillouin et al. [19], we counted the number of meta-analyses provided in the supplementary data table referred to as "retained meta-analyses" (which yielded 217 meta-analyses), even though the main text referred to 196 meta-analyses. In addition, most of the percentages mentioned in the main text agreed with the total being 217, rather than 196.

[b] Lodi et al. [16] reviewed n = 86 papers from the n = 325 papers in Senior et al. [27] for their subset of meta-analysis in ecology and evolution (they used the papers that Senior et al. [27] included in their second order meta-analysis). To avoid non-independence between these two review papers, we only collected data from Lodi et al. [16] for criteria that were not addressed by Senior et al. [27].

[c] Nakagawa et al. [24] reviewed the compilation of papers in O'Dea et al. [2] extracting additional information for our criteria of interest.

[d] Philibert et al. [25] analyzed 73 meta-analyses from 55 papers.

[e] Vetter et al. [28] reviewed 133 papers, from which they quantified the percentage that did not report using weights. Then, from the 83 papers that used weights, they quantified the percentage of papers that quantified and explored heterogeneity.

https://doi.org/10.1371/journal.pone.0292606.t001

independence in our compilation, we kept only [15] that had information for six of our selected criteria; i.e., we excluded [14] from our analyses. We compiled information from the 18 papers (Table 1) that provided quantitative data on the quality of reporting or execution of the meta-analyses. Because Lodi et al. [16] provided metrics for two separate reviews in two topic areas, our final analyses were based on 19 cases, taken from 18 papers (Fig 1). A PRISMA plot [17] is presented in Fig 1, which details the number of papers in each screening step.

Overlap between review papers

If there is overlap in the meta-analyses evaluated in these review papers, comparisons between review papers may not be independent; however, because each review used a different set of search algorithms and often targeted a specific topic, such overlap might be small. We quantified the overlap between meta-analysis reviews for all cases in which the full list of references was available (in the main text or in the supplementary material), or when the authors replied to our requests for this information (S2 Appendix). We used the first author's last name, journal, and year as the identification string to measure overlap in the number of publications shared between review papers.

Compliance with reporting and execution criteria

Our final list of criteria for analysis is presented in Table 2, where we detail which paper contributed data to each criterion. Because different review papers used slightly different criteria (or different names to refer to similar criteria), we matched similar criteria and provided details on which information was extracted for each review paper in S3 Appendix. We obtained the proportion of the meta-analyses that complied with a particular criterion (data are available in the supplementary data file "compilation-of-previous-review-papers"). For the final analysis, we included only those criteria for which we could gather information from at least two review papers.

In addition, because we had access to the full set of meta-analyses reviewed by Pappalardo et al. [1], we expanded on their results by adding additional criteria detailed below and highlighted in Table 2 and S3 Appendix. Pappalardo et al. [1] analyzed 96 meta-analyses related to global change (PRISMA diagram available in their S1 Fig). For the criteria related to Reporting, we collected new information on: inclusion/exclusion criteria, the number of papers and the number of effect size estimates, the types of non-independence, and if the software, specific functions, and code used for the analyses were provided (when applicable). For the criteria related to Execution, we compiled new data to evaluate if the publication explored temporal changes in effects, conducted sensitivity analyses, controlled for phylogenetic non-independence, and tested for publication bias. More details on the calculations for each criterion are provided in S3 Appendix.

We compiled the percentage of meta-analyses in each review paper that complied with each Reporting and Execution criterion. We classified performance for each criterion as "high" when the percentage of papers complying with a criterion was $\geq$75%, "moderate" when compliance was $\geq$50% but $<$75%, "low" when compliance was $\geq$25% but $<$50%, and "very low" when compliance was $<$25%.

Non-independence

A portion of the new data we collected from the meta-analyses reviewed by Pappalardo et al. [1] focused on non-independence. When a publication acknowledged non-independence (e.g., described some type of non–independence), we also recorded the source of the non-independence that was acknowledged, if the authors attempted to account for it, and the methods used to address non-independence. Non-independence arising from non-independent within-study error (e.g., multiple measurements of the same individual, shared control or treatments) was coded as "sample", whereas non-independence arising from study-level correlation (e.g., multiple effect sizes from each publication which could generate random paper effects) was coded as "study". If a study reported both sources of non-independence, we recorded both (e.g., coding the study as "sample, study"). To code if a publication addressed non-independence, we used: "yes", when the publication described one or more sources of non-independence and addressed at least one; and "no" when the publication did not address non-independence. We coded the methods used to address non-independence as: 1) "average", when the non-independent values were averaged (e.g., [29] averaged repeated measurements,

Table 2. List of reporting and execution criteria compiled from reviews of meta-analyses.

Criteria	Definition	Reviews with data available
Reporting		
Full details of bibliographic searches	Described details such as the databases searched, if specific filters were used, the key words used, and the time span of the review.	[1,2,7,10,15,16,18,19,25,26]
Inclusion/exclusion criteria	Clearly described the process of screening and study selection, detailing the criteria used to include (or exclude) studies.	[1 new data,2,7,10,15,16,18,19,26]
Reference list of primary studies	Provided the full citation for all the primary studies included in the meta-analysis.	[1,2,7,10,16,18,25,26]
Meta-analytical model	Explained the type of meta-analytic model used to analyze the effect sizes (e.g., a random-effects model) and the type of factors and model structure for more complex models.	[1,2,7,16,18,19,27]
Dataset used in the meta-analysis	Provided the data used for the meta-analysis: the effect sizes and their variances (when applicable) and moderators (i.e., moderators), if used.	[1,2,7,16,18,25]
Data used to calculate effect sizes (raw data)	Provided the data used to calculate effect sizes (e.g., the mean and number of replicates for treatment and control from each comparison).	[1,15,19]
The number of papers and the number of effect size estimates	Provided the final number of papers included in the literature review and the number of effect size estimates included in the meta-analysis.	[1 new data,2,8]
The software used	Identified the software used to conduct the meta-analysis.	[1,2,7,16,18,23,25,27]
The packages used (if applicable)	Identified the packages used to conduct the meta-analysis, if applicable (e.g., if a scripting program like R was used and *metafor* package was used).	[1 new data,2]
The functions used (if applicable)	Identified the functions used for data analysis, if applicable (e.g., the rma.mv function from the *metafor* R package).	[1 new data,2]
The code used (if applicable)	Provided the code used to conduct the meta-analysis, if applicable (e.g., if a scripting program like R was). If the full code is provided, packages and functions are available.	[1 new data,2]
The types of non-independence	Described the sources of non-independence. For example, a non-independent within-study error could occur when there are multiple measurements of the same individual or shared control or treatments; and non-independence could also emerge from study-level correlation (e.g., when there are multiple effect sizes from each publication which could generate random paper effects).	[1 new data,2,16,18]
Execution		
Weighted effect sizes by study precision	The meta-analysis weighted effect sizes by study precision. The most used weight is the inverse of the variance, but weights can also be based on sample size.	[1,7,15,16,18,19,25,28]
Quantified heterogeneity in effect sizes	The meta-analysis provided heterogeneity statistics (e.g., Q statistics, I^2, τ^2).	[2,7,10,15,16,18,19,25–28]
Explored causes of heterogeneity	The causes of heterogeneity were explored using explanatory variables either through statistical analyses or graphical visualizations.	[1,2,7,10,16,18,19,21,27,28]
Conducted multifactorial analysis of moderators	When multiple moderators were included, the non- independence among moderators was accounted for by including all the moderators in the same model.	[7,16,21,27]
Tested for publication bias	Publication bias was assessed with any of the recommended methods (e.g., funnel plots).	[1 new data,2,7,10,15,16,18,19,23,25,26]
Conducted sensitivity analysis	Quantified the effect of different methodological choices by conducting a sensitivity analysis: e.g., comparing results of a weighted analysis with a reduced dataset versus an unweighted analysis with the full dataset.	[1 new data,2,7,10,16,18,25,26]
Explored temporal changes in effect size	Temporal changes in effect sizes were assessed with any of the recommended methods (e.g., a cumulative meta-analysis).	[1 new data,7,16,24]
Controlled for phylogenetic non-independence	When multiple species were included in the meta-analysis, their phylogenetic relatedness was considered. The effect of phylogenetic relatedness can be assessed using a phylogeny, if available, or by using taxonomy as a proxy.	[1 new data,7,8,16,20,22,23]

In this table we provide broad definitions of Reporting and Execution criteria and detail the meta-analysis reviews from which we extracted information from. Each review may have defined the criteria slightly differently; in S3 Appendix we detail how data was extracted and matched for each criterion and for each review paper. For Pappalardo et al. [1] we highlighted the cases in which we re-reviewed the meta-analyses to compile new data. The list of criteria and definitions were informed by Koricheva & Gurevitch [7, Table 3] and the PRISMA EcoEvo checklist [2, Table 1]. In the Discussion section, we expand on best practices for each issue and available solutions.

https://doi.org/10.1371/journal.pone.0292606.t002

[30] averaged across species); 2) "choose", when the authors chose one value from multiple non-independent values (e.g., [31] used last sampling point, [32] used one response variable per study); and 3) "model", when the authors accounted for non-independence within the meta-analytic model (e.g., [33] included paper ID as a random effect, [34] included variance covariance matrix obtained from phylogenetic distances); and 4) "tested", when non-independence was assessed, found not to be demonstrable, and was subsequently ignored (e.g., [35–37]). If a test was done and non-independence was supported, then the paper was coded according to the method used to address non-independence, and not as "tested" (e.g., [38]). When multiples methods to address non-independence were used, they were all listed (e.g., "choose, average").

We analyzed and visualized data using the R software [39] and packages *scales* [40], *flextable* [41], *pander* [42], *kableExtra* [43], *readxl* [44], *ggcharts* [45], and *tidyverse* [46]. All the data files and the code used to compile information, analyze data, and create figures and tables, are provided as Supporting Information.

Results

Overlap between review papers

The overlap between review papers was generally low. For Reporting criteria, the median number of shared papers was 2 and the mean was 3; for Execution criteria, the median number of shared papers was 2 and the mean was 4.7. In S2 Appendix we include the overlap matrices for each quality criterion that show the number of papers that overlapped between each review paper, and the distribution of the number of shared papers (S1 and S2 Figs). The two instances with largest overlap were: 1) overlap of 77 papers between Koricheva & Gurevitch [7] and Senior et al. [27], representing a 23.9% overlap for the "Meta-analytical model", "The software used", "Quantified heterogeneity in effect sizes", "Tested for publication bias", and "Multifactorial analysis of moderators criteria" criteria; 2) overlap of 74 papers between Koricheva & Gurevitch [7] and Cadotte et al. [8], representing 30.8% overlap for the "Controlled for phylogenetic non-independence" criterion.

Compliance with reporting and execution criteria

In our compilation of the 19 meta-analysis reviews from 18 papers, we found wide variability in the compliance within and between the different quality criteria. We did not observe any clear differences among different subdisciplines (S3 and S4 Figs, S4 Appendix), nor did we observe any temporal trends in compliance (S5 and S6 Figs, S4 Appendix).

In general, we observed better compliance in Reporting (Fig 2) than Execution (Fig 3). Across reviews, we observed high to moderate compliance with Reporting criteria such as: providing the list of references (Fig 2F), specifying the meta-analytic model (Fig 2D), detailing inclusion/exclusion criteria (Fig 2C), and identifying the packages (Fig 2J) and software (Fig 2K) used. On the other hand, Reporting criteria exhibited very low to moderate compliance in including full details on the literature search (Fig 2B), providing the data used to calculate effect sizes (Fig 2E), and providing the analytic code (Fig 2G) and functions used (Fig 2H).

For the Execution criteria, there was lower compliance with criteria such as conducting sensitivity analyses (Fig 3A), controlling for phylogenetic non-independence (Fig 3B), exploring temporal changes in effect sizes (Fig 3D), conducting a multifactorial analysis of moderators (vs. multiple single factor analyses) (Fig 3E), and testing for publication bias (Fig 3G). In contrast, most papers explored the possible causes of heterogeneity (Fig 3C). For the Execution criteria of weighting effect sizes by study precision (Fig 3H) and quantifying heterogeneity in

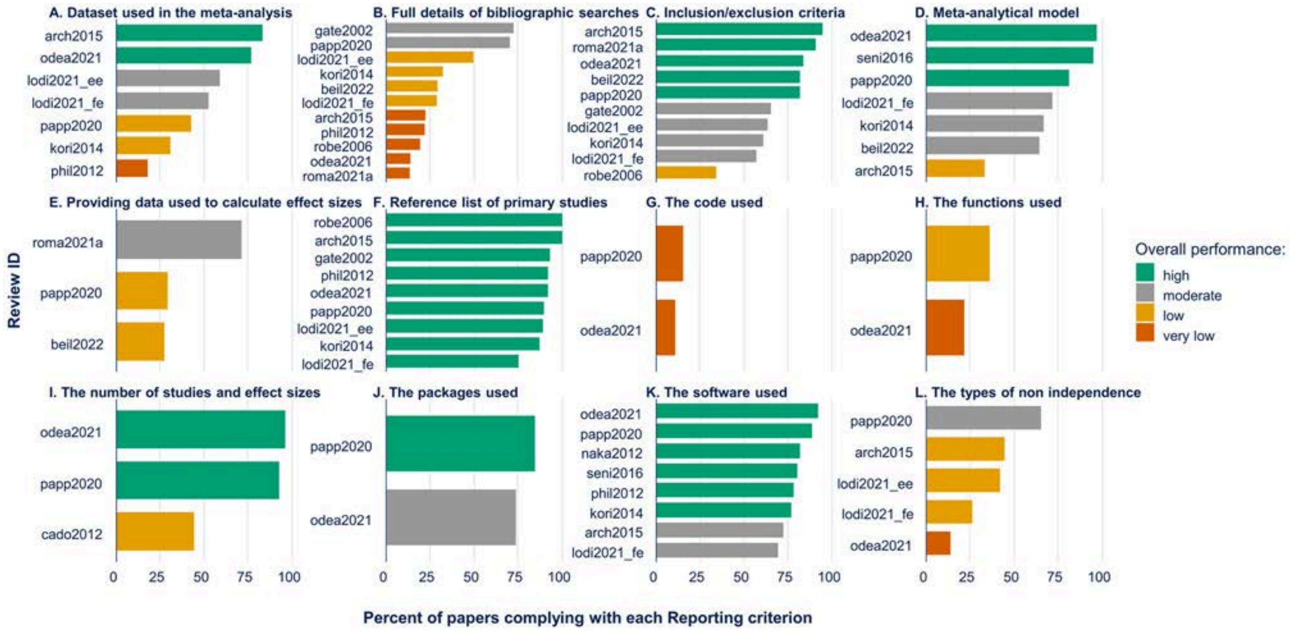

Fig 2. Percent of papers complying with reporting criteria. The percent of papers complying with each criterion is plotted for each synthesis paper. The colors indicate the overall performance for each criterion coded as: "High" (percentage compliance ≥ 75%), "moderate" (50 ≤ percentage compliance < 75), "low" (25 ≤ percentage compliance < 50), and "very low" (percent compliance < 25%). The Review ID corresponds to the papers listed in Table 1.

https://doi.org/10.1371/journal.pone.0292606.g002

effect sizes (Fig 3F), compliance was highly variable (ranging from 33% to 92% for weighting and 22% to 100% for quantifying heterogeneity).

Non-independence

In our review of the meta-analyses compiled by Pappalardo et al. [1], we found that in all meta-analyses but one, the number of effect sizes was larger than the number of papers (Fig 4). This variation was often of several orders of magnitude (Fig 4, note the log scale in both axis). This suggests the possibility of non-independence as effect sizes derived from the same source paper are more likely to be more similar than are those coming from different papers. 66% of the meta-analyses acknowledged some type of non-independence (Fig 5A). The source of non-independence acknowledged most often (68% of the time) was related to the design of the original experiment (e.g., a common control used for different treatments) and how data were collected (Fig 5B). Acknowledging non-independence from other sources of correlation (e.g., multiple effect sizes per publication) was less common (36% of the time, Fig 5B). Most papers (98%) that acknowledged non-independence took steps to address it (Fig 5C). The most common ways that non-independence was addressed (Fig 5D) were: choose (55%) and average (32%). Only 11% tested for the effects of non-independence, and only 16% explicitly modeled a potential source of non-independence. A few papers used a combination of these approaches (which is why the percentages sum to slightly more than 100%).

Discussion

Compliance with reporting criteria

Even though there was overall good compliance for Reporting criteria (e.g., providing the list of primary papers included in the meta-analysis), many issues remain widespread. Meta-

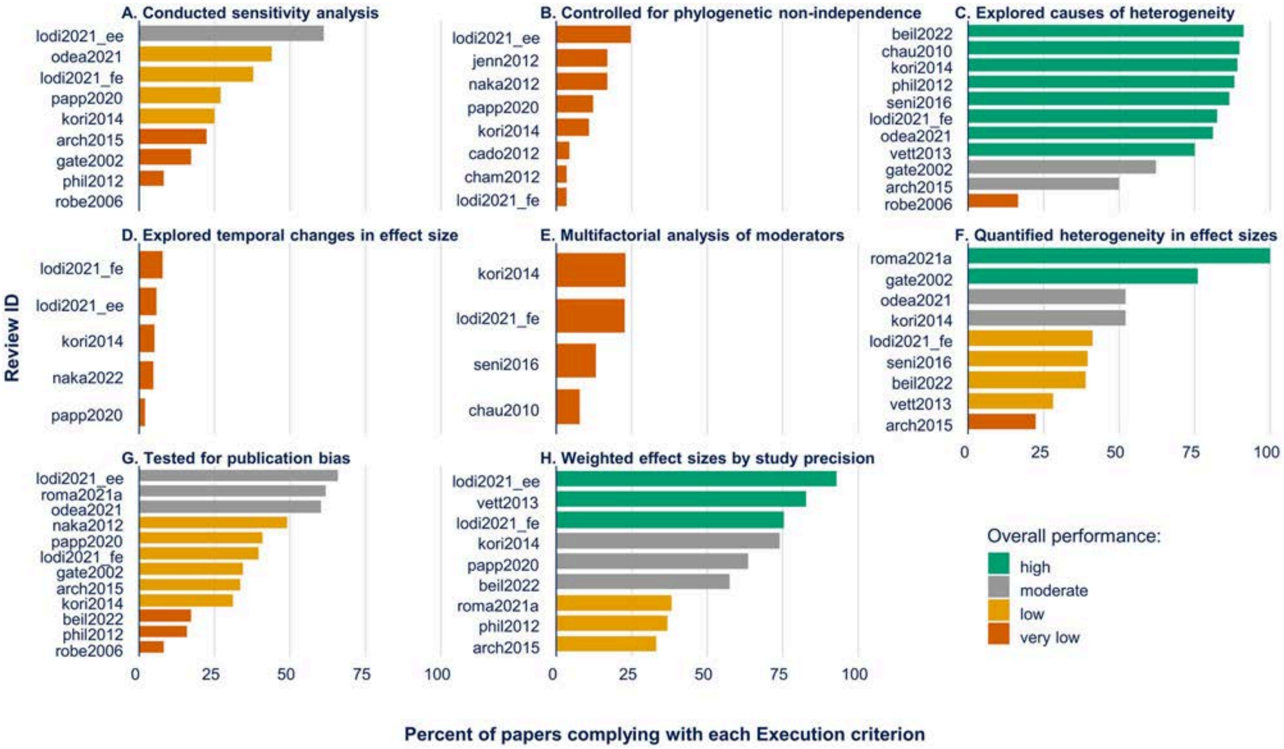

Fig 3. Percent of papers complying with execution criteria (i.e., recommendations during data analysis). The percent of papers complying with each criterion is plotted for each synthesis paper. The colors indicate the overall performance for each criterion coded as: "High" (percentage compliance $\geq$ 75%), "moderate" (50 $\leq$ percentage compliance $<$ 75), "low" (25 $\leq$ percentage compliance $<$ 50), and "very low" (percent compliance $<$ 25%). The Review ID corresponds to the papers listed in Table 1. In panel (A), Roberts [26] evaluated sensitivity analysis, and found 0% of papers reporting it.

https://doi.org/10.1371/journal.pone.0292606.g003

analysis papers were less consistent in their reporting of information that can critically affect the results of a meta-analysis (e.g., the inclusion/exclusion criteria [49]). Even minimal information such as the number of papers and effect sizes were not always included; for example, the review by Cadotte et al. [8] showed that fewer than 50% of the meta-analyses reported this basic information. Many of these meta-analyses are not reproducible because relatively few of the meta-analyses provided the data used to conduct their analyses (e.g., effect sizes, variances, moderators). Making the data available benefits the research community by supporting meta-research or integrative research that combines the data in some novel way without having to redo the data extraction [50]. Many studies also failed to specify the model used to analyze the data (e.g., random-effects model). To remedy this issue, we suggest academic journals adopt standard checklists for reporting items, such as the PRISMA-EcoEvo checklist [Tables 1 and 2]. Similarly, for the meta-analyses that reported using a programming language, very few reported the specific functions or the code used for data analysis, which are essential for reproducibility. Failure to share code is not exclusive to meta-analysis; even for research articles published in ecological journals that encourage or mandate code-sharing, only 27% provide all or some of the code used for the analyses [51].

To encourage code and data sharing, journals can develop incentives. Some cover the fee for publishing data in a repository. Discounts on open access fees could further encourage authors to share code and data. Most data repositories provide a separate DOI for the dataset so it can be properly cited. Most importantly, as reporting practices improve and data become

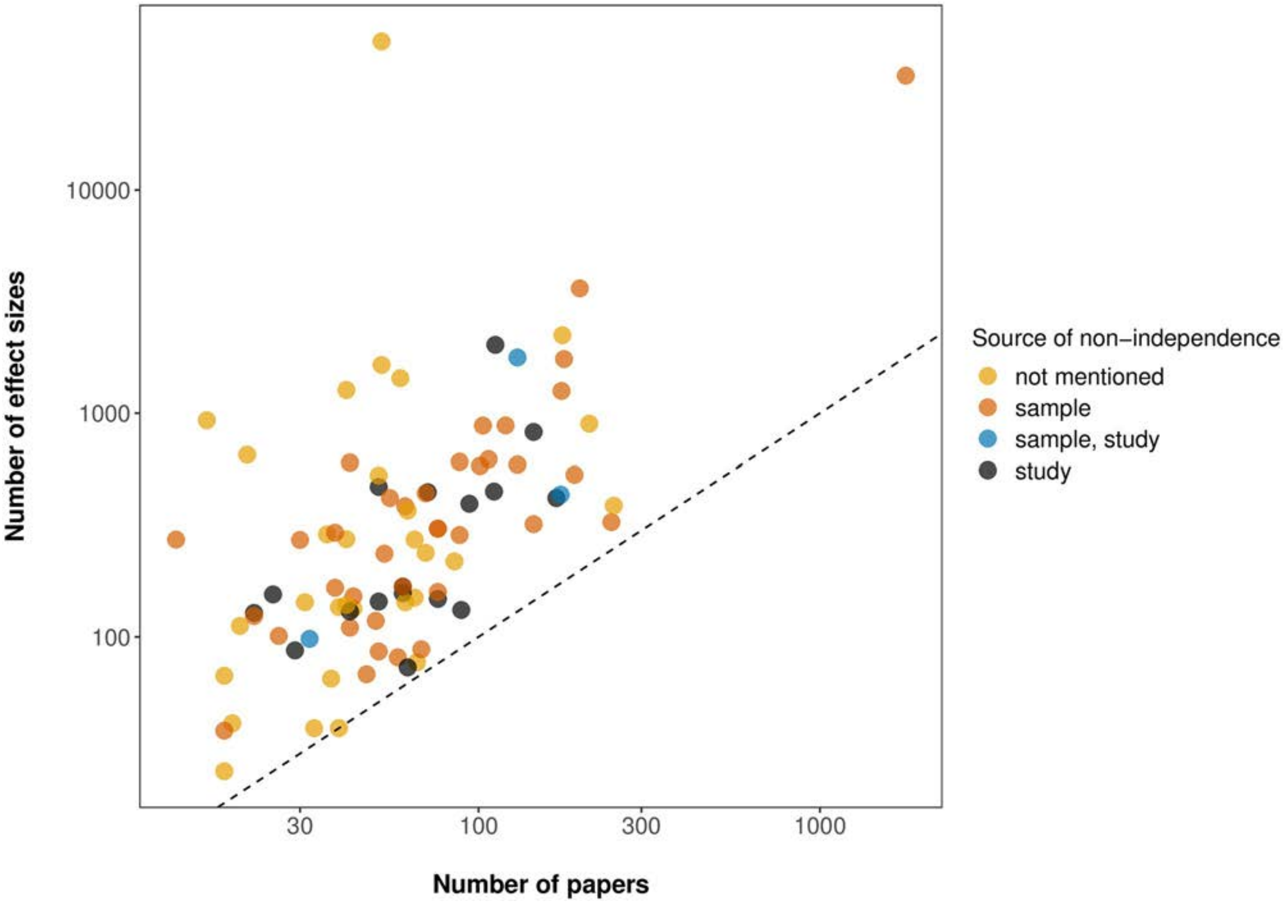

Fig 4. Relationship between the number of papers and the number of effect sizes included in ecological meta-analysis. The relationship between the number of papers and the number of effect sizes based on our re-analysis of meta-analyses in Pappalardo et al. [1]. The colors indicate when the source of non-independence was not mentioned (yellow) or was acknowledged at the sample level (orange), study level (black) or both (blue). Note that the axes are on log scales to accommodate two studies with an extreme number of effect sizes. One is a study reporting 52 papers that did not specify the number of effect sizes but provided a dataset with 46,347 rows [47]. The other is a study that analyzed 1,785 papers and reported a total of 32,567 effects for one of their meta-analyses [48]. The dashed line is the one-to-one line (x = y).

https://doi.org/10.1371/journal.pone.0292606.g004

available, reproducibility will improve. Achieving computational reproducibility will help ensure results are robust, transparent, and credible. This is particularly important for researchers working in applied science and conservation where stakes are high, and transparency can help maintain public trust [52].

Having commonly accepted guidelines for meta-analysis could improve the quality of meta-analyses, although empirical research on this topic often gives mixed results. Even before the PRISMA guidelines were initially developed [in 2009 by 17], systematic reviews in the medical field showed higher reporting quality compared with meta-analyses in ecology [10,26]. This was likely due to the early guidelines for systematic reviews in the medical field using a standard set of methods developed by the Cochrane Collaboration [53] and to ecological studies often being more complex and varied in terms of the types of questions, sources of data, and experimental design. How much the PRISMA guidelines improved the quality of reporting in medical meta-analysis is not clear. Some papers report a moderate increase in reporting quality after the publication of PRISMA guidelines [54], while others report no change [55, only reviewed abstracts]. Two syntheses of medical meta-analyses found that reporting quality improved after journals endorsed and implemented PRISMA guidelines

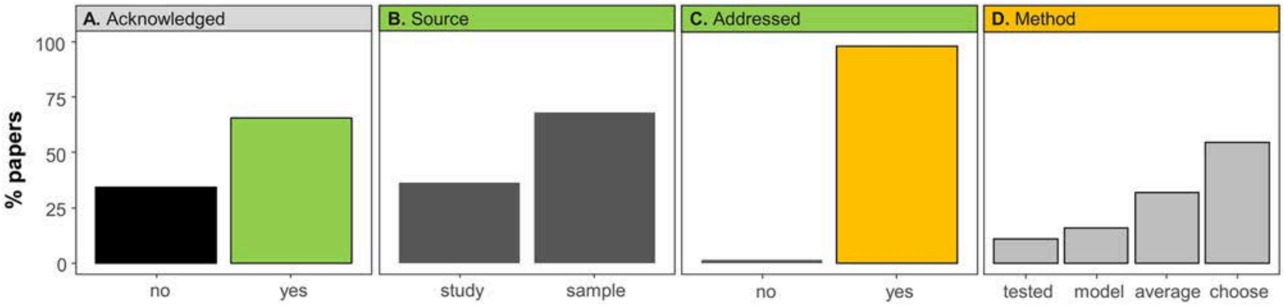

Fig 5. Percent of papers that acknowledged non-independence, addressed it, and which methods they used to deal with non-independence. (A) Percent of papers that acknowledged at least one type of independence in their data ("yes") or did not acknowledge non-independence ("no"). (B) For the papers that did acknowledge non-independence, the sources of non-independence were classified as "study" or "sample". (C) The percent of papers that addressed at least one type of non-independence ("yes"), or did not address non-independence ("no"). (D) For the papers that did address non-independence, we show the methods used to address non-independence, classified as: "Choose", when the authors chose one value from multiple non-independent values; "average", when the non-independent values were averaged; "model", when the authors accounted for non-independence within the meta-analytic model, and "tested" when the authors tested for the effects of non-independence. The papers that used more than one method (or source) were counted in each category, so the percentages between levels of panels B and D sum to greater than 100%.

https://doi.org/10.1371/journal.pone.0292606.g005

[56,57]. In their review of meta-analyses in ecology and evolution, O'Dea et al. [2] showed that meta-analyses that reported to have followed specific guidelines tended to have higher quality ratings. Lodi et al. [16], in their review of meta-analyses from freshwater ecology, found higher quality in more recent years and suggested that previous papers on reporting guidelines were the reason for the improvement. The recently published PRISMA Eco-Evo guidelines [2] could generate even bigger impacts on the quality of reporting if journals required those guidelines during the submission of meta-analyses. Some journals such as PLOS ONE already have a structure in place to detect if certain key aspects of meta-analysis are present (e.g., a PRISMA plot). We suggest that introducing the PRISMA Eco-Evo guidelines to a journal's submission process will greatly benefit the discipline, especially if the journal publishes a large number of meta-analyses as is the case for Ecology Letters, Global Change Biology, Ecology, Oecologia and American Naturalist (the top five according to our compilation, S5 Appendix).

Compliance with execution criteria

The low compliance with Execution criteria suggests that most meta-analyses do not follow recommended methods. One of the advantages of meta-analysis is that effect sizes are conventionally weighted by the precision of the observed effect size. The PRISMA-EcoEvo guidelines recommend using a weighted analysis because weighting generally yields more precise estimates of effects that unweighted analyses [2]. However, our compilation of reviews showed that the percentage of papers that weighted effect sizes varied widely (from 33% to 93%). By reanalyzing the meta-analyses from Pappalardo et al. [1], we found that only 42% weighted by the inverse of the variance (as recommended by [2]), 6% weighted by sample size, 16% used some non-traditional weight, and 36% of the meta-analyses did not weight effect sizes in any way to account for variation in their precision or quality. Several papers used unweighted analyses because of incomplete reporting in the primary publications (e.g., the original papers did not report standard deviations or sample sizes), and the meta-analysts did not want to greatly reduce the number of studies by excluding the studies without estimates of variance. New imputation techniques to estimate variances can provide an effective alternative to conducting an unweighted meta-analyses [58]. Under some situations, however, unweighted analyses can provide results as reliable as those obtained using weighted analyses, e.g., when among-study variance is large relative to within-study variances (Song et al, pers. comm.) or when effect sizes and their variances are independent and follow a normal distribution [58]. Conducting a sensitivity analysis with a smaller dataset that compares results from unweighted and weighted meta-analyses can be a way to check if results are robust to that decision [59].

A central purpose of ecological meta-analysis is to quantify heterogeneity and explore its causes. A fixed-effects model, which assumes no heterogeneity among true effect sizes, has been discouraged for ecological meta-analysis [60] and its use seems to be declining [27]. Pappalardo et al. [1] found that random-effects and mixed-effects models were the most popular in their review. Given that heterogeneity in ecological and evolutionary meta-analyses is high [27], it is encouraging we found high compliance in exploring the causes of heterogeneity (either by conducting statistical analysis of moderators or by graphical visualizations). However, our compilation showed high variability on providing metrics quantifying heterogeneity (e.g., using Q or I^2 statistics), with most reviews (seven of nine) reporting very low to moderate compliance on quantifying heterogeneity.

A common approach when meta-analysts explore heterogeneity is to evaluate individual covariates one at a time, rather than in a single analysis. This is an invalid approach because these explanatory variables may not be independent or because failure to simultaneously account for a factor may give rise to spurious results (e.g., via Simpson's paradox). A multifactorial analysis of moderators, which would address this issue, was reported only in a few meta-analyses. Gates [10] also mentioned that most meta-analyses did not correct for multiple testing when conducting subgroup analyses. This deficiency could reflect limitations imposed by available software. For example, MetaWin, a commonly used software in older meta-analyses, did not allow multifactorial analyses. Additionally, Nakagawa & Santos [23] noted that meta-analytic data are often sparse and including all moderators in the model may greatly reduce sample sizes, making such analyses problematic [see 61]. However, given the increase in data availability, and an increase in the sophistication of software capable of including multiple moderators (e.g., the R package *metafor* [62]), multifactorial analysis are more feasible and should become more common.

Sensitivity analyses evaluate the robustness of a meta-analysis to methodological choices, for example, by exploring how results change when removing influential points, altering the weighting schemes, or calculating different types of effect sizes. Sensitivity analyses can also be used to explore the consequences of non-independence [63]. Across reviews, we observed that a low percentage of meta-analyses reported conducting a sensitivity analysis. Although analysis for publication bias can be considered a type of sensitivity analysis, we followed previous reviews (e.g., [7]) and quantified them as separate criteria. It is possible that some researchers may have run additional explorations that could be considered sensitivity analyses in earlier stages of a publication, but these were not stated explicitly in their final manuscripts or supplementary information. We encourage authors of meta-analyses to include the results of sensitivity analyses, to showcase the different types of limitations related to their dataset, and to quantify if different methodological choices affect their conclusions. This will enhance the robustness of their meta-analysis.

A general problem in scientific research is that significant results are more likely to be published. For meta-analysis, this may bias the meta-analyst towards discovering more significant effect sizes, which in turn may bias the conclusion of the meta-analysis. Meta-analysis has tools to identify the existence of publication bias (e.g., via a funnel plot) and to assess its impact (e.g., by calculating a fail-safe number), but these methods have pros and cons, and meta-analysts are discouraged from relying on only one approach [12,23]. Despite the availability of methods, compliance for assessing publication bias was low ($<50\%$) in nine of the twelve reviews that quantified this criterion.

A different type of publication bias in meta-analysis of ecology and evolution arises from temporal trends in effect sizes. For example, a decrease in the magnitude of effect sizes over time has been observed in various areas of ecology and evolution [64,65], although the existence of such a general trend has been debated [66]. Possible non-biological causes for the decrease in the magnitude of effect sizes with time are time-lags, selective reporting, shifts in the choice of research organisms, and changes in statistical methods (reviewed in [65]). Not accounting for temporal trends may give a false sense that conclusions from meta-analyses are invariant through time [67]. Instead, meta-analyses should explore temporal trends, which would also help identify additional sources of heterogeneity in the effect sizes. Koricheva et al. [65] described graphical and statical methods available to analyze temporal trends and included examples analyzing real datasets. The simplest graphical method is a plot of effect sizes versus publication year. Another option is a cumulative meta-analysis, in which mean effect sizes are calculated starting with the oldest publication and adding in other studies chronologically [65]. The temporal effect can also be assessed in a statistical model by incorporating publication year as a moderator [65]. In the ecological meta-analyses reviewed by Pappalardo et al. [1], 40% of meta-analyses reported the range in publication years of the original papers and this time span averaged 41 years (min = 1, median = 34, max = 115). In their review, Cadotte et al. [8] found an average time span of 15 years and a maximum of 65 years. Despite the wide time span reported in many of the meta-analyses reviewed, the percent of meta-analyses that addressed temporal trends in effect size was very low (ranging from 1 to 8%).

Methods for detecting and quantifying the effects of publication bias, such as regression or correlation-based approaches for analyzing the asymmetry of funnel plots, may encounter challenges in ecology and evolution due to heterogeneity and non-independence, two characteristics commonly associated with data in ecology and evolution [24]. To address this issue, Nakagawa et al. [24] proposed using what they referred to as “conditional residuals” from hierarchical models instead of observed effect sizes in analyzing funnel plots. This approach accounts for heterogeneity and non-independence by subtracting the fixed effects and random effects that model the heterogeneity and non-independence from the observed effect sizes.

A Reporting criterion that we could not include in the main analysis was associated with the method used to obtain the uncertainty interval associated with the mean effect size–only one review paper assessed this criterion [1]. However, the method used to obtain an uncertainty interval (e.g., based upon a t-distribution or a z-distribution, or by bootstrapping) can affect the coverage and thus the inferences from the analyses. For example, when replication is low, the use of the bootstrap or z-distribution will generate confidence intervals that are much too narrow, resulting in more significant effects than expected [1]. 39% of the meta-analysis reviewed by Pappalardo et al. [1] did not specify this information, even though it can have dramatic effects on statistical inference. Furthermore, of the meta-analyses that did report their method, a vast majority (>90%) used either the bootstrap or z-distribution [1], which can lead to artificially small confidence intervals.

Non-independence

Non-independence is common in biological meta-analyses; if not addressed properly, non-independence can produce spurious results [68]. In particular, not properly accounting for non-independence often leads to wrong estimates of standard errors and thus invalid statistical inference. Non-independence may occur at the sampling level, such as through using a shared control or taking repeated measurements of the same individuals over time. Non-independence at the study level may occur by comparing species that are close phylogenetically or systems that are close spatially. Our review of the meta-analyses compiled by

Pappalardo et al. [1] showed a higher percent of papers (66%) that acknowledged some source of non-independence, compared with reviews by Archmiller et al. [18] and O'Dea et al. [2], which reported 44% and 14% respectively. Most non-independence arose at the sampling level, due, for example, to using a shared control or taking repeated measurements on replicates. In these cases, most meta-analysts addressed the non-independence before doing the analysis (one of the solutions mentioned in [63]), by either choosing a subset of the data (55%) or using the average of the non-independent measurements (32%). We did not observe any meta-analyses that tried to explicitly model the covariance of the sample-level non-independence even though formulae for covariance have been derived for many forms of non-independence [69,70].

In contrast, non-independence arising from study-level correlation are much less recognized and addressed in meta-analyses. Only 14% of the meta-analyses from Pappalardo et al. [1] attempted to address study-level non-independence by applying a multilevel model (e.g., using study ID as a random effect in the meta-analytic model). Including study effects in a multilevel model is one the simplest solutions, and different levels can be included in the model to account for non-independence due to other sources (e.g., species effects, discussed below) [12]. Study-level non-independence may arise in multiple ways in ecological meta-analyses [63]. In fact, we found that the number of effects far exceeds the number of papers in the meta-analyses we examined and many of the studies that have a large ratio of number of effect sizes to number of papers did not acknowledge sources of non-independence (Fig 4). Given that studies from the same source paper are more likely to share similar environments or methodology, it is very likely that study-level non-independence is common. Thus, the relatively low proportion of published meta-analyses addressing study-level non-independence is a source of concern.

Study-level non-independence frequently arises from phylogeny. Closely related species may have similar traits that could be associated with similar responses; thus, data from different species may not be independent. In paired analysis of the same dataset using traditional and phylogenetic meta-analyses, Chamberlain et al. [20] reported that 40% of random-effects meta-analyses changed from significant when not adjusting for phylogeny to non-significant when a phylogenetic meta-analysis was used. The influence of phylogenetic relatedness on the outcome of meta-analysis has also been studied using simulations. Cinar et al. [71] found that under moderately strong phylogenetic relatedness, failing to account for species-level variance generated biased estimates of mean effects and led to poor coverage (i.e., confidence intervals that were too small). This is troubling given that all the meta-analysis reviews found very low compliance with respect to controlling for phylogenetic non-independence. In some cases, a phylogenetic analysis may not be possible because a reliable phylogeny is not available. However, in those cases, taxonomic information (e.g., family or genus) can be used as a moderator in the analysis (e.g., as done in [35]).

How can we implement best practices?

As was highlighted in multiple sessions at the 2020 Ecological Society of America meeting, there is a need for data integration at multiple scales, data synthesis, and training of young investigators on computer programming and the use of appropriate statistical tools. To address this gap, it is important to train ecologists in meta-analysis techniques. This could involve including meta-analysis topics in the curriculum of Ecology/Evolution graduate programs, which could be done as part of courses focused on statical methods and data analysis or the subject could be required for qualifying exams. Training also could be provided in short workshops. The ever-increasing availability of ecological data and the scope of the questions we need to answer, require that we provide all researchers access to the tools necessary for synthesis research.

For researchers who wish to learn on their own, there are multiple resources available. Marc Lajeuneese has a YouTube channel (https://www.youtube.com/c/lajeunesselab) with multiple videos explaining techniques for the different steps to conduct a meta-analysis, and has also developed the R package *metagear* [13] that has functions to help with paper screening and data extraction (http://lajeunesse.myweb.usf.edu/metagear/metagear_basic_vignette.html). The Environmental Computing website (http://environmentalcomputing.net/meta-analysis/) provides tutorials to conduct meta-analysis with the R package *metafor* [62], and also general information on how to organize data that will be useful for meta-analysts (http://environmentalcomputing.net/data-entry/). The *metafor* website by Wolfgang Viechtbauer has detailed documentation and examples of data analysis and models to conduct a meta-analysis using *metafor* (https://www.metafor-project.org/doku.php/metafor). The CRAN task view for meta-analysis [72, https://cran.r-project.org/web/views/MetaAnalysis.html] provides a full list of R packages that have useful tools related with meta-analysis. Also available for R users is the online book *Doing meta-analysis with R*: *a hands-on guide* that is aimed at non-experts [73, https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/]. For Python users, there are also specific tools and resources focusing on meta-analysis [74,75]. For those who would prefer a friendly user interface, the software OpenMEE [76] provides advanced tools for meta-analysis in ecology and evolution (http://www.cebm.brown.edu/openmee/help.html). The Inter-Disciplinary Ecology and Evolution Lab (http://www.i-deel.org/links.html) provides several resources related to meta-analysis and systematic reviews. Finally, Briggs et al. [77, http://metaanalysis.ecology.uga.edu/] are developing a series of meta-analysis tutorials.

Now that specific guidelines are available with a focus on meta-analysis in Ecology and Evolution, authors can follow the PRISMA EcoEvo checklist [2] as a guide to plan their meta-analysis, and reviewers and editors can assess the quality of reporting in a meta-analysis. More importantly, improving the reporting quality and following guidelines will also improve the quality of the research. Although we showed that authors are better at following Reporting criteria than Execution criteria, compliance was highly variable suggesting there remains ample room for improvement. Making the PRISMA EcoEvo checklist mandatory for paper submission could help by 1) helping to identify if a paper is not a statistically-focused meta-analysis (e.g., papers that self-report as "meta" analysis because they analyzed a large dataset with multiple factors, but that do not use a meta-analytic framework, and 2) encouraging good reporting, reproducibility, and overall quality. A key component to future meta-analyses and synthesis studies are data sharing and good data management practices [78].

The learning curve to conduct a meta-analysis and follow all the detailed steps may appear steep and discouraging. The training opportunities mentioned above could help reduce the learning curve and facilitate improved reporting and execution. Some researchers argue that we should not let the perfect be the enemy of the good [79]. We agree, but also argue that the "good" should include clearly reporting methods, following best practices, and making data and code available to the community. Doing so will make the inferences from meta-analyses more robust and less controversial—the ultimate goal of a valuable statistical tool.

Supporting information

S1 Checklist. PRISMA 2020 checklist.
(DOCX)

S1 Fig. Distribution of paper overlap for reporting criteria. Distribution of the number of papers shared between reviews for all the Reporting criteria combined.
(TIF)

S2 Fig. Distribution of paper overlap for execution criteria. Distribution of the number of papers shared between reviews for all the Execution criteria combined.
(TIF)

S3 Fig. Percent of papers complying with Reporting criteria by review discipline. The percent of papers complying with each Reporting criterion is plotted for each review paper. The colors indicate different subdisciplines of the review papers. The Review ID corresponds to the papers listed in Table 1 of the main manuscript.
(TIF)

S4 Fig. Percent of papers complying with Execution criteria by review discipline. The percent of papers complying with each Execution criterion is plotted for each review paper. The colors indicate different subdisciplines of the review papers. The Review ID corresponds to the papers listed in Table 1 of the main manuscript.
(TIF)

S5 Fig. Percent of papers complying with Reporting criteria as a function of the time period analyzed by the review paper. Each panel represents a Reporting criterion. The line segment indicates the time period covered by each of the review papers that addressed a particular criterion.
(TIF)

S6 Fig. Percent of papers complying with Execution criteria as a function of the time period analyzed by the review paper. Each panel represents an Execution criterion. The line segment indicates the time period covered by each of the review papers that addressed a particular criterion.
(TIF)

S1 Appendix. Paper screening. Additional details and R code for paper screening using the package *metagear*.
(PDF)

S2 Appendix. Overlap between review papers. Additional details and R code used to calculate overlap between review papers, including overlap matrices for each criterion and R code for S1 and S2 Figs.
(PDF)

S3 Appendix. Details on the information extracted from each review paper for each performance criterion. For data from Pappalardo et al. [1], we indicated when their data was re-analyzed or when we collected new data in this study by re-reviewing their compilation of ecological meta-analyses with the tag “added”. When the number of publications complying (or not complying) with one of the criteria was reported, we used that information to calculate the percentage of papers complying; in other cases the reviews directly reported the information as a percentage. In a few papers in which we had the original review data for each criterion [e.g., 18], we summed the number of papers complying with each criterion, and then calculated the percentage of compliance based on the total number of papers relevant for that criterion.
(PDF)

S4 Appendix. Additional results. Details and R code for additional results about compliance of quality criteria by review discipline and temporal trends in compliance. Includes the R code to generate S3–S6 Figs.
(PDF)

S5 Appendix. Journals that publish the most meta-analyses. Number of meta-analyses per journal that had been included in the meta-analysis reviews. Because the distribution is strongly right skewed (with most journals publishing a few meta-analyses), we display only the journals with at least 5 meta-analyses.
(PDF)

S6 Appendix. R code used for data analysis. This rmarkdown file (.Rmd) includes the code to conduct the data analysis and create Figs 1 to 5.
(RMD)

S1 Data. List of references in reviews. This Microsoft Excel Worksheet (.xlsx) includes the compilation of all the references analyzed by previous reviews.
(XLSX)

S2 Data. New data from Pappalardo et al. [1]. This Microsoft Excel Worksheet (.xlsx) includes additional data collected by re reviewing the ecological meta-analysis compiled by Pappalardo et al. [1]. Please cite this publication and Pappalardo et al. [1] if you are using the data in this file for your research.
(CSV)

S3 Data. Data compilation from previous review papers. This Microsoft Excel Worksheet (. xlsx) includes information on the quality of Reporting and Execution criteria compiled from the review papers listed in Table 1.
(XLSX)

S4 Data. Papers screened and final classification. This Microsoft Excel Worksheet (.xlsx) includes the final list of papers screened using metagear from the Web Of Science search and also the additional papers found from additional sources.
(XLSX)

S5 Data. Journal names dictionary. This Comma Separated File (.csv) includes a conversion dictionary from short journal names to long journal names and was used when analyzing paper overlap.
(XLSX)

Acknowledgments

We are very thankful to Marc Cadotte, Scott Chamberlain, Steve Hovick, Julia Koricheva, Jessica Gurevitch, Shinichi Nakagawa, and João Paulo Romanelli for providing detailed additional methods or metadata associated with their reviews of meta-analyses. We also thank O'Dea et al. [2] for sharing a draft of the PRISMA EcoEvo guidelines before it was published. PP is grateful for the encouragement made by an anonymous reviewer of the Pappalardo et al. [1] publication to expand the analysis on the quality of reporting in meta-analysis in a separate publication. PP highly appreciates the feedback from Alyssa Gehman, Linsey Haram, Carrie Keogh, and Rachel Smith on an early draft of the manuscript.

Author Contributions

Conceptualization: Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg.

Data curation: Paula Pappalardo, Chao Song.

Formal analysis: Paula Pappalardo.

Funding acquisition: Craig W. Osenberg.

Investigation: Paula Pappalardo, Chao Song.

Methodology: Paula Pappalardo, Chao Song, Craig W. Osenberg.

Software: Paula Pappalardo.

Validation: Paula Pappalardo, Chao Song.

Visualization: Paula Pappalardo.

Writing – original draft: Paula Pappalardo.

Writing – review & editing: Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg.

References

1. Pappalardo P, Ogle K, Hamman EA, Bence JR, Hungate BA, Osenberg CW. Comparing traditional and Bayesian approaches to ecological meta-analysis. Methods Ecol Evol. 2020 Jul 26;2041-210X.13445.
2. O'Dea RE, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, et al. Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. Biol Rev. 2021 Oct; 96(5):1695–722.
3. Osenberg CW, St. Mary CM, Schmitt RJ, Holbrook SJ, Chesson P, Byrne B. Rethinking ecological inference: density dependence in reef fishes: Inference and density dependence. Ecol Lett. 2002 Nov 7; 5 (6):715–21.
4. Blankinship JC, Niklaus PA, Hungate BA. A meta-analysis of responses of soil biota to global change. Oecologia. 2011 Mar; 165(3):553–65. https://doi.org/10.1007/s00442-011-1909-0 PMID: 21274573
5. Thébault A, Mariotte P, Lortie CJ, MacDougall AS. Land management trumps the effects of climate change and elevated CO_2 on grassland functioning. Gibson D, editor. J Ecol. 2014 Jul; 102(4):896–904.
6. Stewart G. Meta-analysis in applied ecology. Biol Lett. 2010 Feb 23; 6(1):78–81. https://doi.org/10.1098/rsbl.2009.0546 PMID: 19776064
7. Koricheva J, Gurevitch J. Uses and misuses of meta-analysis in plant ecology. Gómez-Aparicio L, editor. J Ecol. 2014 Jul; 102(4):828–44.
8. Cadotte MW, Mehrkens LR, Menge DNL. Gauging the impact of meta-analysis on ecology. Evol Ecol. 2012 Sep; 26(5):1153–67.
9. Lortie CJ. Formalized synthesis opportunities for ecology: systematic reviews and meta-analyses. Oikos. 2014 Aug; 123(8):897–902.
10. Gates S. Review of methodology of quantitative reviews using meta-analysis in ecology. J Anim Ecol. 2002 Jul; 71(4):547–57.
11. Gurevitch J, Curtis PS, Jones MH. Meta analysis in ecology. Adv Ecol Res. 2001; 32:199–247.
12. Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. BMC Biol. 2017 Dec; 15(1):18. https://doi.org/10.1186/s12915-017-0357-7 PMID: 28257642
13. Lajeunesse MJ. Facilitating systematic reviews, data extraction, and meta-analysis with the metagear package for R. Methods Ecol Evol. 2016; 7:323–30.
14. Romanelli JP, Silva LGM, Gonçalves MCP, Naves RP, Almeida DRA, Resende AF, et al. Repeatability of the searching process in reviews of restoration outcomes. Restor Ecol [Internet]. 2021 Nov [cited 2022 Sep 18]; 29(8). Available from: https://onlinelibrary.wiley.com/doi/10.1111/rec.13496
15. Romanelli JP, Meli P, Naves RP, Alves MC, Rodrigues RR. Reliability of evidence-review methods in restoration ecology. Conserv Biol. 2021 Feb; 35(1):142–54. https://doi.org/10.1111/cobi.13661 PMID: 33347737
16. Lodi S, Godoy BS, Ortega JCG, Bini LM. Quality of meta-analyses in freshwater ecology: A systematic review. Freshw Biol. 2021 May; 66(5):803–14.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med. 2009; 151(4):264–9. https://doi.org/10.7326/0003-4819-151-4-200908180-00135 PMID: 19622511

18. ArchMiller AA, Bauer EF, Koch RE, Wijayawardena BK, Anil A, Kottwitz JJ, et al. Formalizing the definition of meta-analysis in *Molecular Ecology*. Mol Ecol. 2015 Aug; 24(16):4042–51.

19. Beillouin D, Cardinael R, Berre D, Boyer A, Corbeels M, Fallot A, et al. A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon. Glob Change Biol. 2022 Feb; 28(4):1690–702. https://doi.org/10.1111/gcb.15998 PMID: 34873793

20. Chamberlain SA, Hovick SM, Dibble CJ, Rasmussen NL, Van Allen BG, Maitner BS, et al. Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis: Phylogenetic meta-analysis. Ecol Lett. 2012 Jun; 15(6):627–36.

21. Chaudhary VB, Walters LL, Bever JD, Hoeksema JD, Wilson GWT. Advancing synthetic ecology: a database system to facilitate complex ecological meta-analyses. Bull Ecol Soc Am. 2010 Apr; 91 (2):235–43.

22. Jennions MD, Kahn AT, Kelly CD, Kokko H. Meta-analysis and sexual selection: past studies and future possibilities. Evol Ecol. 2012 Sep; 26(5):1119–51.

23. Nakagawa S, Santos ESA. Methodological issues and advances in biological meta-analysis. Evol Ecol. 2012 Sep; 26(5):1253–74.

24. Nakagawa S, Lagisz M, Jennions MD, Koricheva J, Noble DWA, Parker TH, et al. Methods for testing publication bias in ecological and evolutionary meta-analyses. Methods Ecol Evol. 2022 Jan; 13(1):4–21.

25. Philibert A, Loyce C, Makowski D. Assessment of the quality of meta-analysis in agronomy. Agric Ecosyst Environ. 2012 Feb; 148:72–82.

26. Roberts PD, Stewart GB, Pullin AS. Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. Biol Conserv. 2006 Oct; 132 (4):409–23.

27. Senior AM, Grueber CE, Kamiya T, Lagisz M, O'Dwyer K, Santos ESA, et al. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. Ecology. 2016 Dec; 97(12):3293–9. https://doi.org/10.1002/ecy.1591 PMID: 27912008

28. Vetter D, Rücker G, Storch I. Meta-analysis: A need for well-defined usage in ecology and conservation biology. Ecosphere. 2013 Jun; 4(6):1–24.

29. Shantz AA, Burkepile DE. Context-dependent effects of nutrient loading on the coral–algal mutualism. Ecology. 2014 Jul; 95(7):1995–2005. https://doi.org/10.1890/13-1407.1 PMID: 25163130

30. Yoon S, Read Q. Consequences of exotic host use: impacts on Lepidoptera and a test of the ecological trap hypothesis. Oecologia. 2016 Aug; 181(4):985–96. https://doi.org/10.1007/s00442-016-3560-2 PMID: 26820566

31. Gamfeldt L, Lefcheck JS, Byrnes JEK, Cardinale BJ, Duffy JE, Griffin JN. Marine biodiversity and ecosystem functioning: what's known and what's next? Oikos. 2015 Mar; 124(3):252–65.

32. Kuebbing SE, Nuñez MA. Negative, neutral, and positive interactions among nonnative plants: patterns, processes, and management implications. Glob Change Biol. 2015 Feb; 21(2):926–34. https://doi.org/10.1111/gcb.12711 PMID: 25142018

33. McDevitt-Irwin J, Iacarella J, Baum J. Reassessing the nursery role of seagrass habitats from temperate to tropical regions: a meta-analysis. Mar Ecol Prog Ser. 2016 Sep 28; 557:133–43.

34. Wooliver R, Pfennigwerth AA, Bailey JK, Schweitzer JA. Plant functional constraints guide macroevolutionary trade-offs in competitive and conservative growth responses to nitrogen. Fox C, editor. Funct Ecol. 2016 Jul; 30(7):1099–108.

35. Martinson HM, Fagan WF. Trophic disruption: a meta-analysis of how habitat fragmentation affects resource consumption in terrestrial arthropod systems. Hill J, editor. Ecol Lett. 2014 Sep; 17(9):1178–89. https://doi.org/10.1111/ele.12305 PMID: 24866984

36. Feng Z, Rütting T, Pleijel H, Wallin G, Reich PB, Kammann CI, et al. Constraints to nitrogen acquisition of terrestrial plants under elevated CO_2. Glob Change Biol. 2015 Aug; 21(8):3152–68.

37. He Q, Silliman BR. Consumer control as a common driver of coastal vegetation worldwide. Ecol Monogr. 2016 Aug; 86(3):278–94.

38. Anderson JT. Plant fitness in a rapidly changing world. New Phytol. 2016 Apr; 210(1):81–7. https://doi.org/10.1111/nph.13693 PMID: 26445400

39. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2021. https://www.R-project.org/.

40. Wickham H, Seidel D. scales: Scale Functions for Visualization. R package version 1.2.0. [Internet]. 2022. https://CRAN.R-project.org/package=scales.

41. Gohel D. flextable: Functions for Tabular Reporting. R package version 0.6.10. [Internet]. 2021. https://CRAN.R-project.org/package=flextable

42. Daroczi G, Tsegelskyi R. pander: An R "Pandoc" Writer. R package version 0.6.5. [Internet]. 2022. https://CRAN.R-project.org/package=pander

43. Zhu H. kableExtra: Construct Complex Table with "kable" and Pipe Syntax. R package version 1.3.4. [Internet]. 2021. https://CRAN.R-project.org/package=kableExtra.

44. Wickham H, Bryan J. readxl: Read Excel Files. R package version 1.3.1. [Internet]. 2019. https://CRAN.R-project.org/package=readxl.

45. Neitmann T. ggcharts: Shorten the Distance from Data Visualization Idea to Actual Plot. R package version 0.2.1. [Internet]. 2020. https://CRAN.R-project.org/package=ggcharts.

46. Wickham H. Welcome to the tidyverse. J Open Source Softw. 2019; 4(43):1686.

47. Albert A, Auffret AG, Cosyns E, Cousins SAO, D'hondt B, Eichberg C, et al. Seed dispersal by ungulates as an ecological filter: a trait-based meta-analysis. Oikos. 2015 Sep; 124(9):1109–20.

48. Yuan ZY, Chen HYH. Decoupling of nitrogen and phosphorus in terrestrial plants associated with global changes. Nat Clim Change. 2015 May; 5(5):465–9.

49. Englund G, Sarnelle O, Cooper SD. The importance of data-selection criteria: meta-analysis of stream predation experiments. Ecology. 1999 Jun; 80(4):1132–41.

50. Carpenter SR, Armbrust EV, Arzberger PW, Chapin FS, Elser JJ, Hackett EJ, et al. Accelerate Synthesis in Ecology and Environmental Sciences. BioScience. 2009 Sep; 59(8):699–701.

51. Culina A, van den Berg I, Evans S, Sánchez-Tójar A. Low availability of code in ecology: A call for urgent action. PLOS Biol. 2020 Jul 28; 18(7):e3000763.

52. Powers SM, Hampton SE. Open science, reproducibility, and transparency in ecology. Ecol Appl [Internet]. 2019 Jan [cited 2022 Jul 17]; 29(1). Available from: https://onlinelibrary.wiley.com/doi/10.1002/eap.1822 PMID: 30362295

53. Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews ofthe effects ofhealth care. 1994; 309:4.

54. Tunis AS, McInnes MDF, Hanna R, Esmail K. Association of Study Quality with Completeness of Reporting: Have Completeness of Reporting and Quality of Systematic Reviews and Meta-Analyses in Major Radiology Journals Changed Since Publication of the PRISMA Statement? Radiology. 2013 Nov; 269(2):413–26. https://doi.org/10.1148/radiol.13130273 PMID: 23824992

55. Maticic K, Krnic Martinic M, Puljak L. Assessment of reporting quality of abstracts of systematic reviews with meta-analysis using PRISMA-A and discordance in assessments between raters without prior experience. BMC Med Res Methodol. 2019 Dec; 19(1):32. https://doi.org/10.1186/s12874-019-0675-2 PMID: 30764774

56. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the Endorsement of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement on the Quality of Published Systematic Review and Meta-Analyses. Derrick GE, editor. PLoS ONE. 2013 Dec 26; 8(12): e83138. https://doi.org/10.1371/journal.pone.0083138 PMID: 24386151

57. Nawijn F, Ham WHW, Houwert RM, Groenwold RHH, Hietbrink F, Smeeing DPJ. Quality of reporting of systematic reviews and meta-analyses in emergency medicine based on the PRISMA statement. BMC Emerg Med. 2019 Dec; 19(1):19. https://doi.org/10.1186/s12873-019-0233-6 PMID: 30744570

58. Kambach S, Bruelheide H, Gerstner K, Gurevitch J, Beckmann M, Seppelt R. Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. Ecol Evol. 2020 Oct; 10 (20):11699–712. https://doi.org/10.1002/ece3.6806 PMID: 33144994

59. Shantz AA, Lemoine NP, Burkepile DE. Nutrient loading alters the performance of key nutrient exchange mutualisms. Knops J, editor. Ecol Lett. 2016 Jan; 19(1):20–8. https://doi.org/10.1111/ele.12538 PMID: 26549314

60. Gurevitch J, Hedges LV. Statistical issues in ecological meta-analyses. Ecology. 1999; 80 (4):1142–9.

61. Trager MD, Bhotika S, Hostetler JA, Andrade GV, Rodriguez-Cabal MA, McKeon CS, et al. Benefits for Plants in Ant-Plant Protective Mutualisms: A Meta-Analysis. Raine NE, editor. PLoS ONE. 2010 Dec 22; 5(12):e14308. https://doi.org/10.1371/journal.pone.0014308 PMID: 21203550

62. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw [Internet]. 2010 [cited 2019 Jan 10]; 36(3). Available from: http://www.jstatsoft.org/v36/i03/

63. Noble DWA, Lagisz M, O'dea RE, Nakagawa S. Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. Mol Ecol. 2017 May; 26(9):2410–25. https://doi.org/10.1111/mec.14031 PMID: 28133832

64. Jennions MD, Møller AP. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. Proc R Soc Lond B Biol Sci. 2002 Jan 7; 269(1486):43–8.

65. Koricheva J, Jennions MD, Lau J. Temporal trends in effect sizes: causes, detection, and implications. In: Handbook of meta-analysis in ecology and evolution. Princeton, New Jersey: Princeton University Press; 2013. (Chapter 15).

66. Costello L, Fox JW. Decline effects are rare in ecology. Ecology. 2022; 103(6):e3680. https://doi.org/10.1002/ecy.3680 PMID: 35302660

67. Koricheva J, Kulinskaya E. Temporal instability of evidence base: a threat to policy making? Trends Ecol Evol. 2019 Oct; 34(10):895–902. https://doi.org/10.1016/j.tree.2019.05.006 PMID: 31196571

68. Song C, Peacor SD, Osenberg CW, Bence JR. An assessment of statistical methods for nonindependent data in ecological meta-analyses. Ecology. 2020; 101(12):e03184. https://doi.org/10.1002/ecy.3184 PMID: 32893349

69. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Handbook of research synthesis and meta-analysis. New York: The Russell Sage Foundation; 1994. (2nd edition).

70. Lajeunesse MJ. On the meta-analysis of response ratios for studies with correlated and multi-group designs. Ecology. 2011 Nov; 92(11):2049–55. https://doi.org/10.1890/11-0423.1 PMID: 22164829

71. Cinar O, Nakagawa S, Viechtbauer W. Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny. Methods Ecol Evol. 2022 Feb; 13(2):383–95.

72. Dewey M, Viechtbauer W. CRAN Task View: Meta-Analysis [Internet]. 2022. https://CRAN.R-project.org/view=MetaAnalysis

73. Harrer M. Doing meta-analysis with R: a hands-on guide. First edition. Boca Raton: CRC Press; 2022. 1 p.

74. Masoumi S, Shahraz S. Meta-analysis using Python: a hands-on tutorial. BMC Med Res Methodol. 2022 Dec; 22(1):193. https://doi.org/10.1186/s12874-022-01673-y PMID: 35820854

75. Paracha UZ. Basics of meta-analysis with basic steps in Python (Programming language). Independently published; 2021. 116 p.

76. Wallace BC, Lajeunesse MJ, Dietz G, Dahabreh IJ, Trikalinos TA, Schmid CH, et al. OpenMEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. Poisot T, editor. Methods Ecol Evol. 2017 Aug; 8(8):941–7.

77. Briggs, A, Peacor SD, Osenberg CW. Meta-analysis for Ecology and Natural Sciences [Internet]. 2019. http://metaanalysis.ecology.uga.edu/.

78. Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. Front Ecol Environ. 2013 Apr; 11(3):156–62.

79. Munafò M. Don't let the perfect be the enemy of the good. PLOS Biol. 2021 Jul 15; 19(7):e3001327. https://doi.org/10.1371/journal.pbio.3001327 PMID: 34264931

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where each item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title, we used “meta-evaluation” to be more inclusive of our goals
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods and supplementary files
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Methods and supplementary files
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods and supplementary files
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods and supplementary files
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods and supplementary files
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	N/A
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods and supplementary files (overlap between review papers)
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods

Section and Topic	Item #	Checklist item	Location where each item is reported
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods and Appendix S1
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Methods
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Table 1 and Figure 1
Study characteristics	17	Cite each included study and present its characteristics.	Table 1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Results on overlap between review papers
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion
	23b	Discuss any limitations of the evidence included in the review.	N/A
	23c	Discuss any limitations of the review processes used.	N/A

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where each item is reported
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Supporting Information includes data files and R code used to analyze data

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Appendix S1 - Paper screening

Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg

From: A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Additional details and code for paper screening

We screened the 751 abstracts of citation downloaded from Web Of Science using the package **metagear** [13]. We used the function *effort_initialize()* to add columns “study_id”, “reviewers” and “include”. After that, the *abstract_screener()* function brings up an interactive window to rate the abstract for inclusion or not, the options are: “yes”, “maybe”, and “no”. The rating column gets added. The package adds a new column in the citation data with this decision. It can also split the effort among different coauthors; if so, it is necessary to include a column with the reviewer/s name and assign reviewer’s effort (see code below). Paper screening and data extraction from the reviews of meta-analysis was done by Paula Pappalardo; additional data extraction for the non-independence section was done by Chao Song.

When in doubt if to include a paper, they were tagged as “maybe” and the full text was evaluated. Here are the R code and **metagear** specific functions we used:

```
# load data downloaded from Web of Science and select columns we need

wos <- as.data.frame(read_excel("WOS/meta-analysis_WOS_751.xls",
                                sheet= "savedrecs", range= cell_cols("A:BT"))) %>%
  select("Publication Type", "Authors", "Article Title",
                     "Source Title", "Volume", "Issue", "Start Page",
                     "End Page", "DOI", "Publication Date", "Publication Year",
                     "Abstract", "WoS Categories")

# rename columns

names(wos) <- c("pub_type", "authors", "title", "pub_name","vol", "issue",
                "start_page", "end_page", "doi", "pub_date", "pub_year",
                "abstract","wos_categ")
```

```
library(metagear)

# prepare the file for the screening effort

wos_scan <- effort_initialize(wos)

# save file with the IDs as a backup

write.csv(wos_scan, "WOS/wosWithIds.csv")

# randomly distribute screening effort to a team
```

```r
theTeam <- c("Paula")
theRefs_unscreened <- effort_distribute(wos_scan, reviewers = theTeam,
    effort = c(100), save_split = T)

# start the abstract viewer to do first pass

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

# get the summary of your work

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

# the effort file got saved by default in the main project, I moved it to the WOS folder
```

```r
# load revision file

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

# subset by inclusion decision

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

# try getting pdfs from R

PDFs_collect(theRefs_included, DOIcolumn = "doi", FileNamecolumn = "STUDY_ID",
    directory = "C:/Users/Paula/Dropbox/Meta-analysis/ReportingQuality/WOS/pdfs")
```

We manually added the papers found with other sources to the “effort_Paula.csv”, and provide the file *“papers-screened_Final-classification.csv”* as supplementary data. This file has the final list of papers screened, the final decision, and the reasons to exclude papers after reading the full text.

Appendix S2

Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg

From: A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Setup

```
knitr::opts_chunk$set(echo = T, eval = T, warning = F,
                      message = F, comment = "")

# load libraries we need

library(kableExtra)
library(tidyverse)
library(ggcharts)
library(rcartocolor)
library(readxl)
library(flextable)
```

```
# load objects we need, generated with the R code provided

load("objects/reporting.R")
load("objects/execution.R")
load("objects/toplot.R")

# nice format for plots

niceplot <- theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_text(face = "bold", size = 12,
                                    margin = margin(t = 20, r = 0, b = 0, l = 0)),
        axis.title.y = element_text(face = "bold", size = 12,
                                    margin = margin(t = 0, r = 20, b = 0, l = 0)),
        axis.text.x = element_text(size = 11),
        axis.text.y = element_text(size = 11),
        legend.text = element_text(size = 11),
        legend.title = element_text(size = 11),
        legend.position = "right")
```

Overlap between review papers

We compiled the list of references analyzed by each review paper, by extracting them from the main text, the references section, supplementary materials, or by directly emailing the authors. For the following papers we were unable to obtain the list of references, and were not included in the tables of overlap:

- The references for Chaudhary et al. [21] paper were not accessible, we checked their supplementary information and data packages, but the web link gives an error. We also contacted the authors and had no answer.
- The references for Nakagawa & Santos [23] were not provided. We emailed the authors, S. Nakagawa promptly responded, and contacted E. Santos. The authors apologized for being unable to find the data after so many years.

In some cases, the journal names were not identical between lists (e.g., American Naturalist versus The American Naturalist), or had different spellings (e.g., 'Soil Biology & Biochemistry' versus 'Soil, Biology & Biochemistry'). Before analysis, we standardized all the journals names to the long format, and cross checked (and fixed when necessary) the journal names across reviews. We also checked for authors names that had different spelling across reviews (e.g., Castro-vila, vs Castro Vila) and followed the spelling used by most review authors. After the quality check, we created a **review code** by combining: 1) the first author last name (in caps), 2) the year of publication, and 3) the journal (in caps). We standardized the text to uppercase to correct for inconsistencies in spelling (e.g., Van Groenigen versus van Groenigen for authors, or PLoS ONE versus PLOS ONE for journals). Here an example of a final reference id: BYERS_1999_ANNUAL REVIEW OF ECOLOGY AND SYSTEMATICS. The list of references is available in the supplementary data file "DataFiles_list-of-references-in-Reviews".

```r
# function to select columns of interest

selectColumns <- function(mydf){
  mydf_ed <- mydf %>% select(first_author, year, journal)
  return(mydf_ed)
}

# function to add a standarized reference ID to compare accross reviews
addID <- function(mydf){
  mydf_ed <- mydf %>%
    mutate(ref_id = paste(str_to_upper(first_author), year,
                          str_to_upper(journal), sep = "_"),
           journal = str_to_upper(journal))
  return(mydf_ed)
}


# load short journals dictionary

journal_dic <- read.csv("data/DataFiles_journal-names-dictionary.csv", as.is = T)
```

```r
# load references for each publication

arch <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "arch2015", range= cell_cols("A:D"))) %>%
  selectColumns() %>%
```

```
  addID()

cado <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "cado2012", range= cell_cols("A:D"))) %>%
  selectColumns() %>%
  addID()

cham <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "cham2012", range= cell_cols("A:C"))) %>%
  addID()

gate <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "gate2002", range= cell_cols("A:C"))) %>%
  addID()

jenn <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "jenn2012", range= cell_cols("A:B"))) %>%
  left_join(journal_dic, by = "journal_short") %>%
  rowwise() %>%
  mutate(year = as.numeric(str_extract(reference, "\\d{4}")[[1]]),
         first_author = str_split_fixed(reference, " ", 2)[,1]) %>%
  ungroup() %>%
  selectColumns() %>% addID()

kori <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "kori2012", range= cell_cols("A:C"))) %>%
  addID()

odea <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "odea2021", range= cell_cols("A:C"))) %>%
  addID()

papp <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "papp2020", range= cell_cols("A:C"))) %>%
  addID()

phil<- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                sheet= "phil2012", range= cell_cols("A:C"))) %>%
  left_join(journal_dic, by = "journal_short") %>%
  selectColumns() %>% addID()

robe <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "robe2006", range= cell_cols("A:C"))) %>%
  addID()

seni <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "seni2016", range= cell_cols("A:C"))) %>%
  addID()

vett <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "vett2013", range= cell_cols("A:C"))) %>%
  addID()
```

```
lodi_fe <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                    sheet= "lodi2021_fe", range= cell_cols("A:C"))) %>%
  mutate(first_author =  str_split_fixed(Citation, ",", 2)[,1],
         journal = ifelse(journal == "The American Naturalist",
                          "American Naturalist", journal)) %>%
  addID()

lodi_ee <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                    sheet= "lodi2021_ee", range= cell_cols("A:C"))) %>%
  addID()

beil <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "beil2022", range= cell_cols("A:D"))) %>%
  addID()

roma_a <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                  sheet= "roma2021a", range= cell_cols("A:D"))) %>%
  addID()

roma_b <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                  sheet= "roma2021b", range= cell_cols("A:C"))) %>%
  rowwise() %>%
  mutate(year = as.numeric(str_extract(Citation, "\\d{4}")[[1]])) %>%
  ungroup() %>%
  addID()
```

Compare references between the two Romanelli papers:

```
# number of papers shared
length(union(roma_a$ref_id, roma_b$ref_id))

# number of unique papers in Romanelli_a
length(setdiff(roma_a$ref_id, roma_b$ref_id))

# number of unique papers in Romanelli_b
length(setdiff(roma_b$ref_id, roma_a$ref_id))
```

```
# Compile all first author names

allnames <- c(arch$first_author, cado$first_author, cham$first_author, gate$first_author,
              jenn$first_author, kori$first_author, odea$first_author, papp$first_author,
              phil$first_author, robe$first_author, seni$first_author, vett$first_author,
              lodi_fe$first_author, beil$first_author, roma_a$first_author)

View(data.frame(sort(unique(allnames))))

alljournals <- c(arch$journal, cado$journal, cham$journal, gate$journal,
              jenn$journal, kori$journal, odea$journal, papp$journal,
              phil$journal, robe$journal, seni$journal, vett$journal,
              lodi_fe$journal, beil$journal, roma_a$journal)

View(data.frame(sort(unique(alljournals))))
```

```r
load("objects/toplot.R")

# prepare list of criteria

rep_criteria <- toplot %>%
  filter(criteria != "Type of uncertainty interval was described") %>%
  filter(criteria != "Describe effect size used") %>%
  filter(criteria != "The number of studies excluded at each stage of screening") %>%
  filter(criteria.type == "Reporting") %>%
  distinct(criteria)
repvec <- rep_criteria$criteria

exe_criteria <- toplot %>%
  filter(criteria.type == "Execution") %>%
  filter(!criteria == "Standard metrics") %>%
  distinct(criteria)
exevec <- exe_criteria$criteria

# prepare list of references for each review paper

vecs <- list(arch$ref_id, cado$ref_id, cham$ref_id, gate$ref_id,
          jenn$ref_id, kori$ref_id, odea$ref_id, papp$ref_id,
          phil$ref_id, robe$ref_id, seni$ref_id, vett$ref_id,
          lodi_fe$ref_id, lodi_ee$ref_id, beil$ref_id, roma_a$ref_id)

names(vecs) <- c("arch2015", "cado2012", "cham2012", "gate2002", "jenn2012",
                 "kori2014", "odea2021", "papp2020", "phil2012", "robe2006",
                 "seni2016", "vett2013", "lodi2021_fe", "lodi2021_ee",
                 "beil2022", "roma2021a")

# function to get matrix of overlap

getPercentOverlapMatrix <- function(toplot, criterion, vecs){
   # separate the criterion of interest
    thiscriteria <- toplot %>%
    filter(criteria == criterion)
    # subset the vecs list to papers that measure this criterion
    shortlist <- vecs[names(vecs) %in% unique(thiscriteria$paper.id)]
    # create matrix
    mymat <- matrix(ncol = length(shortlist),
                    nrow = length(shortlist),
                    dimnames = list(names(shortlist), names(shortlist)))
  # run loop to calculate the differences between papers
  for(i in 1:length(shortlist)){
    thisvec <- shortlist[[i]]
    for(j in 1:length(shortlist)){
      mymat[i, j] <- length(intersect(thisvec, shortlist[[j]]))
      }
  }
  return(mymat)
}
```

Because the overlap only matters within each criteria used to compare review papers, we calculated the matrix of paper overlap for each criteria.

Matrix of overlap for Reporting criteria

Full details of bibliographic searches:

```
# calculate number of papers shared by reviews
rp1 <- getPercentOverlapMatrix(toplot, criterion = repvec[1], vecs = vecs)
rp1
```

```
            arch2015 gate2002 kori2014 odea2021 papp2020 phil2012 robe2006
arch2015          18        0        2        0        0        0        0
gate2002           0       29        2        0        0        0        0
kori2014           2        2      322        5        3        5        2
odea2021           0        0        5      101        3        0        0
papp2020           0        0        3        3       96        0        0
phil2012           0        0        5        0        0       55        0
robe2006           0        0        2        0        0        0       73
lodi2021_fe        0        4        5        2        3        0        0
lodi2021_ee        2        4        5        4        3        0        1
beil2022           0        0        6        1       13        6        0
roma2021a          0        0        9        0        2        0        0
            lodi2021_fe lodi2021_ee beil2022 roma2021a
arch2015              0           2        0         0
gate2002              4           4        0         0
kori2014              5           5        6         9
odea2021              2           4        1         0
papp2020              3           3       13         2
phil2012              0           0        6         0
robe2006              0           1        0         0
lodi2021_fe         114           1        0         2
lodi2021_ee           1          86        1         0
beil2022              0           1      217         7
roma2021a             2           0        7        65
```

```
# extract overlap elements
rp1_o <- rp1[lower.tri(rp1)]
```

Inclusion/exclusion criteria:

```
# calculate number of papers shared by reviews
rp2 <- getPercentOverlapMatrix(toplot, criterion = repvec[2], vecs = vecs)
rp2
```

```
            arch2015 gate2002 kori2014 odea2021 papp2020 robe2006 lodi2021_fe
arch2015          18        0        2        0        0        0           0
gate2002           0       29        2        0        0        0           4
kori2014           2        2      322        5        3        2           5
odea2021           0        0        5      101        3        0           2
papp2020           0        0        3        3       96        0           3
robe2006           0        0        2        0        0       73           0
lodi2021_fe        0        4        5        2        3        0         114
```

```
lodi2021_ee        2        4        5        4        3        1          1
beil2022           0        0        6        1       13        0          0
roma2021a          0        0        9        0        2        0          2
            lodi2021_ee beil2022 roma2021a
arch2015              2        0         0
gate2002              4        0         0
kori2014              5        6         9
odea2021              4        1         0
papp2020              3       13         2
robe2006              1        0         0
lodi2021_fe           1        0         2
lodi2021_ee          86        1         0
beil2022              1      217         7
roma2021a             0        7        65
```

```
# extract overlap elements
rp2_o <- rp2[lower.tri(rp2)]
```

Meta-analytical model:

```
# calculate number of papers shared by reviews
rp3 <- getPercentOverlapMatrix(toplot, criterion = repvec[3], vecs = vecs)
rp3
```

```
            arch2015 kori2014 odea2021 papp2020 seni2016 lodi2021_fe beil2022
arch2015          18        2        0        0        3           0        0
kori2014           2      322        5        3       77           5        6
odea2021           0        5      101        3       12           2        1
papp2020           0        3        3       96       13           3       13
seni2016           3       77       12       13      325           8        7
lodi2021_fe        0        5        2        3        8         114        0
beil2022           0        6        1       13        7           0      217
```

```
# extract overlap elements
rp3_o <- rp3[lower.tri(rp3)]
```

The software used:

```
# calculate number of papers shared by reviews
rp4 <- getPercentOverlapMatrix(toplot, criterion = repvec[4], vecs = vecs)
rp4
```

```
            arch2015 kori2014 odea2021 papp2020 phil2012 seni2016 lodi2021_fe
arch2015          18        2        0        0        0        3           0
kori2014           2      322        5        3        5       77           5
odea2021           0        5      101        3        0       12           2
papp2020           0        3        3       96        0       13           3
phil2012           0        5        0        0       55        6           0
seni2016           3       77       12       13        6      325           8
lodi2021_fe        0        5        2        3        0        8         114
```

```
# extract overlap elements
rp4_o <- rp4[lower.tri(rp4)]
```

Reference list of primary studies:

```
# calculate number of papers shared by reviews
rp5 <- getPercentOverlapMatrix(toplot, criterion = repvec[5], vecs = vecs)
rp5
```

```
            arch2015 gate2002 kori2014 odea2021 papp2020 phil2012 robe2006
arch2015          18        0        2        0        0        0        0
gate2002           0       29        2        0        0        0        0
kori2014           2        2      322        5        3        5        2
odea2021           0        0        5      101        3        0        0
papp2020           0        0        3        3       96        0        0
phil2012           0        0        5        0        0       55        0
robe2006           0        0        2        0        0        0       73
lodi2021_fe        0        4        5        2        3        0        0
lodi2021_ee        2        4        5        4        3        0        1
            lodi2021_fe lodi2021_ee
arch2015              0           2
gate2002              4           4
kori2014              5           5
odea2021              2           4
papp2020              3           3
phil2012              0           0
robe2006              0           1
lodi2021_fe         114           1
lodi2021_ee           1          86
```

```
# extract overlap elements
rp5_o <- rp5[lower.tri(rp5)]
```

Dataset used in the meta-analysis:

```
# calculate number of papers shared by reviews
rp6 <- getPercentOverlapMatrix(toplot, criterion = repvec[6], vecs = vecs)
rp6
```

```
            arch2015 kori2014 odea2021 papp2020 phil2012 lodi2021_fe
arch2015          18        2        0        0        0           0
kori2014           2      322        5        3        5           5
odea2021           0        5      101        3        0           2
papp2020           0        3        3       96        0           3
phil2012           0        5        0        0       55           0
lodi2021_fe        0        5        2        3        0         114
lodi2021_ee        2        5        4        3        0           1
            lodi2021_ee
arch2015              2
kori2014              5
odea2021              4
papp2020              3
```

```
phil2012              0
lodi2021_fe           1
lodi2021_ee          86
```

```
# extract overlap elements
rp6_o <- rp6[lower.tri(rp6)]
```

The types of non independence:

```
# calculate number of papers shared by reviews
rp7 <- getPercentOverlapMatrix(toplot, criterion = repvec[7], vecs = vecs)
rp7
```

```
            arch2015 odea2021 papp2020 lodi2021_fe lodi2021_ee
arch2015          18        0        0           0           2
odea2021           0      101        3           2           4
papp2020           0        3       96           3           3
lodi2021_fe        0        2        3         114           1
lodi2021_ee        2        4        3           1          86
```

```
# extract overlap elements
rp7_o <- rp7[lower.tri(rp7)]
```

The number of studies and effect sizes:

```
# calculate number of papers shared by reviews
rp8 <- getPercentOverlapMatrix(toplot, criterion = repvec[8], vecs = vecs)
rp8
```

```
         cado2012 odea2021 papp2020
cado2012      240        0        0
odea2021        0      101        3
papp2020        0        3       96
```

```
# extract overlap elements
rp8_o <- rp8[lower.tri(rp8)]
```

The packages used:

```
# calculate number of papers shared by reviews
rp9 <- getPercentOverlapMatrix(toplot, criterion = repvec[9], vecs = vecs)
rp9
```

```
         odea2021 papp2020
odea2021      101        3
papp2020        3       96
```

```
# extract overlap elements
rp9_o <- rp9[lower.tri(rp9)]
```

The functions used:

```r
# calculate number of papers shared by reviews
rp10 <- getPercentOverlapMatrix(toplot, criterion = repvec[10], vecs = vecs)
rp10
```

```
         odea2021 papp2020
odea2021      101        3
papp2020        3       96
```

```r
# extract overlap elements
rp10_o <- rp10[lower.tri(rp10)]
```

The code used:

```r
# calculate number of papers shared by reviews
rp11 <- getPercentOverlapMatrix(toplot, criterion = repvec[11], vecs = vecs)
rp11
```

```
         odea2021 papp2020
odea2021      101        3
papp2020        3       96
```

```r
# extract overlap elements
rp11_o <- rp11[lower.tri(rp11)]
```

Providing data used to calculate effect sizes:

```r
# calculate number of papers shared by reviews
rp12 <- getPercentOverlapMatrix(toplot, criterion = repvec[12], vecs = vecs)
rp12
```

```
          papp2020 beil2022 roma2021a
papp2020        96       13         2
beil2022        13      217         7
roma2021a        2        7        65
```

```r
# extract overlap elements
rp12_o <- rp12[lower.tri(rp12)]
```

This is the summary of the vector including all the number of papers shared for all criteria:

```r
allrep <-c(rp1_o, rp2_o, rp3_o, rp4_o, rp5_o, rp6_o, rp7_o, rp8_o,
           rp9_o, rp10_o, rp11_o, rp12_o)

summary(allrep)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.000   0.000   2.000   3.009   3.000  77.000 
```

This is the distribution of overlap for all Reporting criteria combined:

```r
# create dataframe to plot with ggplot

allrep_df <- data.frame(overlap = allrep)

# make histogram

hist_rep <- ggplot(allrep_df, aes(x = overlap)) +
  geom_histogram(color="black", fill="gray70") +
  xlab("Number of papers shared between reviews") +
  ylab("Frequency of overlap for Reporting criteria") +
  niceplot

# save figure S1

ggsave("figures/Fig_S1.pdf", hist_rep, width = 8, height = 5, dpi = 300)

# display figure S1

hist_rep
```

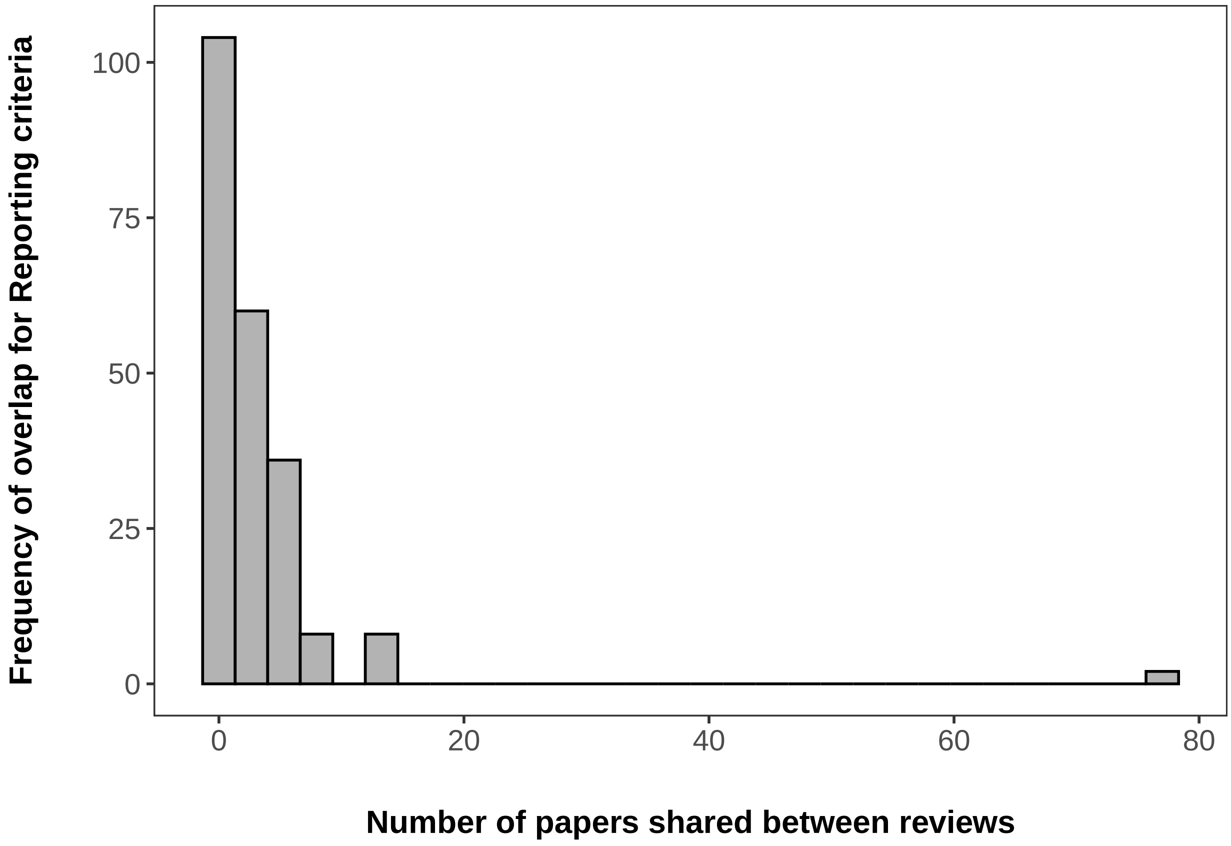

Fig S1. Distribution of paper overlap for Reporting criteria. Distribution of the number of papers shared between reviews for all the Reporting criteria combined.

Matrix of overlap for Execution criteria

Weighted effect sizes by study precision:

```
# calculate number of papers shared by reviews
ec1 <- getPercentOverlapMatrix(toplot, criterion = exevec[1], vecs = vecs)
ec1
```

```
            arch2015 kori2014 papp2020 phil2012 vett2013 lodi2021_fe
arch2015          18        2        0        0        0           0
kori2014           2      322        3        5       30           5
papp2020           0        3       96        0        0           3
phil2012           0        5        0       55        9           0
vett2013           0       30        0        9       83           4
lodi2021_fe        0        5        3        0        4         114
lodi2021_ee        2        5        3        0        7           1
beil2022           0        6       13        6        3           0
roma2021a          0        9        2        0        2           2
            lodi2021_ee beil2022 roma2021a
arch2015              2        0         0
kori2014              5        6         9
papp2020              3       13         2
phil2012              0        6         0
vett2013              7        3         2
lodi2021_fe           1        0         2
lodi2021_ee          86        1         0
beil2022              1      217         7
roma2021a             0        7        65
```

```
# extract overlap elements
ec1_o <- ec1[lower.tri(ec1)]
```

Quantified heterogeneity in effect sizes:

```
# calculate number of papers shared by reviews
ec2 <- getPercentOverlapMatrix(toplot, criterion = exevec[2], vecs = vecs)
ec2
```

```
            arch2015 gate2002 kori2014 odea2021 seni2016 vett2013 lodi2021_fe
arch2015          18        0        2        0        3        0           0
gate2002           0       29        2        0        5        0           4
kori2014           2        2      322        5       77       30           5
odea2021           0        0        5      101       12        2           2
seni2016           3        5       77       12      325       32           8
vett2013           0        0       30        2       32       83           4
lodi2021_fe        0        4        5        2        8        4         114
beil2022           0        0        6        1        7        3           0
roma2021a          0        0        9        0        5        2           2
            beil2022 roma2021a
arch2015           0         0
gate2002           0         0
kori2014           6         9
odea2021           1         0
```

```
seni2016           7         5
vett2013           3         2
lodi2021_fe        0         2
beil2022         217         7
roma2021a          7        65
```

```
# extract overlap elements
ec2_o <- ec2[lower.tri(ec2)]
```

Explored causes of heterogeneity:

```
# calculate number of papers shared by reviews
ec3 <- getPercentOverlapMatrix(toplot, criterion = exevec[3], vecs = vecs)
ec3
```

```
            arch2015 gate2002 kori2014 odea2021 phil2012 robe2006 seni2016
arch2015          18        0        2        0        0        0        3
gate2002           0       29        2        0        0        0        5
kori2014           2        2      322        5        5        2       77
odea2021           0        0        5      101        0        0       12
phil2012           0        0        5        0       55        0        6
robe2006           0        0        2        0        0       73        2
seni2016           3        5       77       12        6        2      325
vett2013           0        0       30        2        9        2       32
lodi2021_fe        0        4        5        2        0        0        8
beil2022           0        0        6        1        6        0        7
            vett2013 lodi2021_fe beil2022
arch2015           0           0        0
gate2002           0           4        0
kori2014          30           5        6
odea2021           2           2        1
phil2012           9           0        6
robe2006           2           0        0
seni2016          32           8        7
vett2013          83           4        3
lodi2021_fe        4         114        0
beil2022           3           0      217
```

```
# extract overlap elements
ec3_o <- ec3[lower.tri(ec3)]
```

Tested for publication bias:

```
# calculate number of papers shared by reviews
ec4 <- getPercentOverlapMatrix(toplot, criterion = exevec[4], vecs = vecs)
ec4
```

```
            arch2015 gate2002 kori2014 odea2021 papp2020 phil2012 robe2006
arch2015          18        0        2        0        0        0        0
gate2002           0       29        2        0        0        0        0
kori2014           2        2      322        5        3        5        2
odea2021           0        0        5      101        3        0        0
```

```
papp2020           0        0        3        3       96        0        0
phil2012           0        0        5        0        0       55        0
robe2006           0        0        2        0        0        0       73
lodi2021_fe        0        4        5        2        3        0        0
lodi2021_ee        2        4        5        4        3        0        1
beil2022           0        0        6        1       13        6        0
roma2021a          0        0        9        0        2        0        0
            lodi2021_fe lodi2021_ee beil2022 roma2021a
arch2015              0           2        0         0
gate2002              4           4        0         0
kori2014              5           5        6         9
odea2021              2           4        1         0
papp2020              3           3       13         2
phil2012              0           0        6         0
robe2006              0           1        0         0
lodi2021_fe         114           1        0         2
lodi2021_ee           1          86        1         0
beil2022              0           1      217         7
roma2021a             2           0        7        65
```

```
# extract overlap elements
ec4_o <- ec4[lower.tri(ec4)]
```

Conducted sensitivity analysis:

```
# calculate number of papers shared by reviews
ec5 <- getPercentOverlapMatrix(toplot, criterion = exevec[5], vecs = vecs)
ec5
```

```
            arch2015 gate2002 kori2014 odea2021 papp2020 phil2012 robe2006
arch2015          18        0        2        0        0        0        0
gate2002           0       29        2        0        0        0        0
kori2014           2        2      322        5        3        5        2
odea2021           0        0        5      101        3        0        0
papp2020           0        0        3        3       96        0        0
phil2012           0        0        5        0        0       55        0
robe2006           0        0        2        0        0        0       73
lodi2021_fe        0        4        5        2        3        0        0
lodi2021_ee        2        4        5        4        3        0        1
            lodi2021_fe lodi2021_ee
arch2015              0           2
gate2002              4           4
kori2014              5           5
odea2021              2           4
papp2020              3           3
phil2012              0           0
robe2006              0           1
lodi2021_fe         114           1
lodi2021_ee           1          86
```

```
# extract overlap elements
ec5_o <- ec5[lower.tri(ec5)]
```

Controlled for phylogenetic non-independence:

```
# calculate number of papers shared by reviews
ec6 <- getPercentOverlapMatrix(toplot, criterion = exevec[6], vecs = vecs)
ec6
```

```
            cado2012 cham2012 jenn2012 kori2014 papp2020 lodi2021_fe
cado2012         240       20        9       74        0          13
cham2012          20       56       11       13        0           2
jenn2012           9       11       94        0        0           0
kori2014          74       13        0      322        3           5
papp2020           0        0        0        3       96           3
lodi2021_fe       13        2        0        5        3         114
lodi2021_ee       12        2       16        5        3           1
            lodi2021_ee
cado2012             12
cham2012              2
jenn2012             16
kori2014              5
papp2020              3
lodi2021_fe           1
lodi2021_ee          86
```

```
# extract overlap elements
ec6_o <- ec6[lower.tri(ec6)]
```

Multifactorial analysis of moderators:

```
# calculate number of papers shared by reviews
ec7 <- getPercentOverlapMatrix(toplot, criterion = exevec[7], vecs = vecs)
ec7
```

```
            kori2014 seni2016 lodi2021_fe
kori2014         322       77           5
seni2016          77      325           8
lodi2021_fe        5        8         114
```

```
# extract overlap elements
ec7_o <- ec7[lower.tri(ec7)]
```

Explored temporal changes in effect size:

```
# calculate number of papers shared by reviews
ec8 <- getPercentOverlapMatrix(toplot, criterion = exevec[8], vecs = vecs)
ec8
```

```
            kori2014 papp2020 lodi2021_fe lodi2021_ee
kori2014         322        3           5           5
papp2020           3       96           3           3
lodi2021_fe        5        3         114           1
lodi2021_ee        5        3           1          86
```

```
# extract overlap elements
ec8_o <- ec8[lower.tri(ec8)]
```

This is the summary of the vector including all the number of papers shared for all Execution criteria:

```
allex <- c(ec1_o, ec2_o, ec3_o, ec4_o, ec5_o, ec6_o, ec7_o, ec8_o)

summary(allex)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.000   0.000   2.000   4.487   5.000  77.000 
```

This is the distribution of overlap for all Execution criteria combined:

```
# create dataframe to plot with ggplot

allex_df <- data.frame(overlap = allex)

# make histogram

hist_exe <- ggplot(allex_df, aes(x = overlap)) +
  geom_histogram(color="black", fill="gray70") +
  xlab("Number of papers shared between reviews") +
  ylab("Frequency of overlap for Execution criteria") +
  niceplot

# save figure S2

ggsave("figures/Fig_S2.pdf", hist_exe, width = 8, height = 5, dpi = 300)

# display figure S3
hist_exe
```

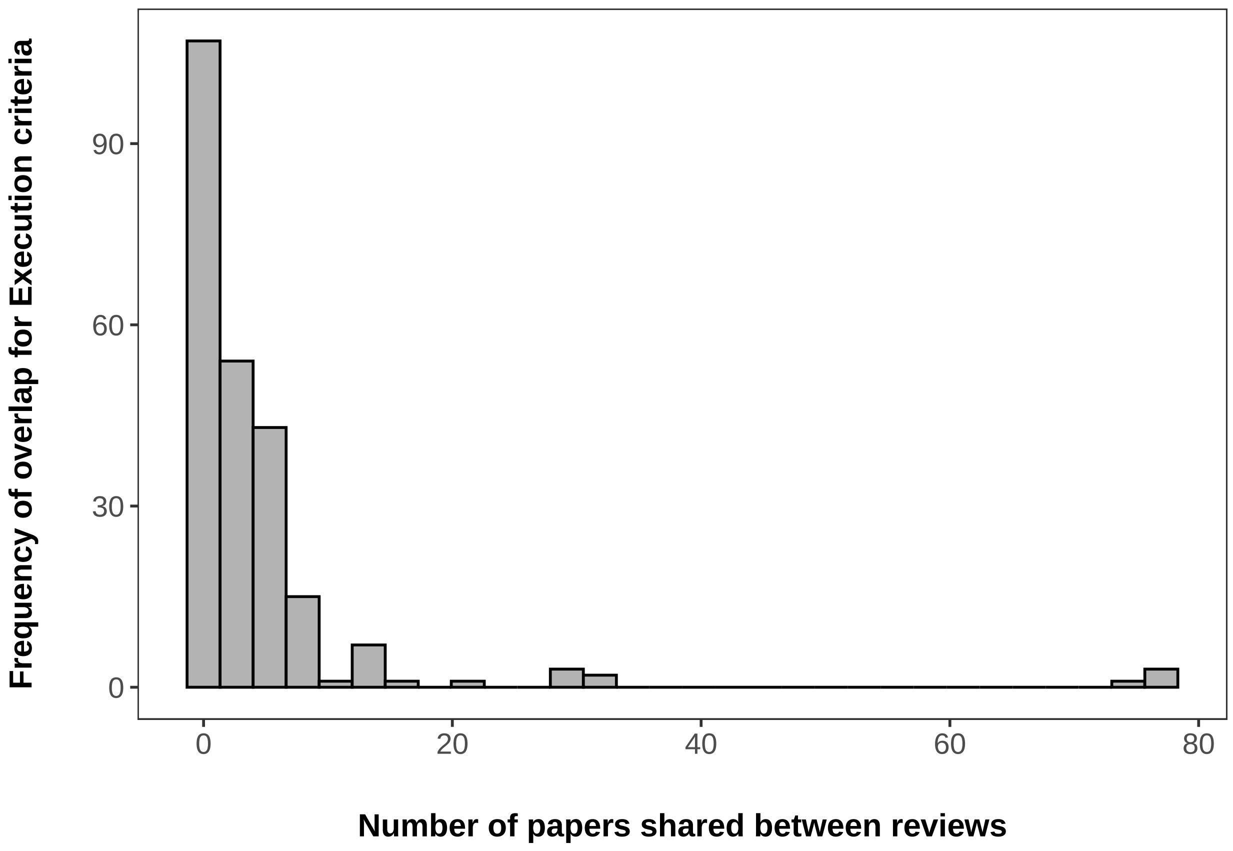

Fig S2. Distribution of paper overlap for Execution criteria. Distribution of the number of papers shared between reviews for all the Execution criteria combined.

Appendix S3

Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg

From: A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Details on the information extracted from each review paper for each performance criterion

For data from Pappalardo et al. [1], we indicated when their data was re-analyzed or when we collected new data in this study by re-reviewing their compilation of ecological meta-analyses with the tag "added". When the number of publications complying (or not complying) with one of the criteria was reported, we used that information to calculate the percentage of papers complying; in other cases, the reviews directly reported the information as a percentage. In a few papers in which we had the original review data for each criterion [e.g., 18], we summed the number of papers complying with each criterion, and then calculated the percentage of compliance based on the total number of papers relevant for that criterion.

Criterion/publication	**Details**
REPORTING	
Full details of bibliographic searches	
Archmiller et al. [18]	The authors evaluated separately if studies reported the "Boolean operators" and "Search databases and dates", we used both columns since they presented the raw data. In their Table S1, we counted as 1 when there was compliance for both items and calculated the percentage dividing by the total number of meta-analyses reviewed.
Beillouin et al. [19]	Data extracted from criterion "Search strings are clearly presented". Percentage calculated from the column "Quality_Lit_strings_OK" in the supplementary data package, n = 217 meta-analyses) and verified when possible with the main text and Figure 5.
Gates [10]	Data extracted from criterion "methods used to locate primary studies". We acknowledge that their concept is less strict, because compliance was considered as long as the study reported the source of the information but did not assess reproducibility.

Koricheva & Gurevitch [7]	Data extracted from criterion “Are details of bibliographic search (electronic data bases used, keyword combinations, years) reported in sufficient detail to allow replication?”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Searching Details” in Tables 3 and 4. Detailed description of each criteria appears in their Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
Philibert et al. [25]	Data extracted from the main text for criterion “Repeatable procedure: A repeatable procedure for the selection of papers for the meta-analysis is presented.”
ODea et al. [2]	Data extracted from criterion 5.4 of the PRISMA EcoEvo checklist (Table 1): “Provide enough information to repeat the equivalent search (if possible), including the timespan covered (start and end dates)”.
Pappalardo et al. [1]	Reanalyzed from Pappalardo et al [1] dataset looknig at information from two columns. We assigned 1 when columns *lit.search.explained* and *keywords.explained* were “yes”. The rest of the studies were coded as 0. The paper that used their own data for the meta-analysis was considered as not relevant.
Roberts et al. [26]	Data extracted from criterion "Defined search terms to be used to identify sources of evidence", Table 1.
Romanelli et al. [16]	Matched to the “2.2 Search strings clearly defined” criteria column in their supplementary data file. We only consider the 63 meta-analyses and did not included the systematic reviews. We assigned 0 to their score 0, 0.5 points to their score for partial compliance (1), and 1 point to their score of full compliance (3). We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
Inclusion/exclusion criteria	
Archmiller et al. [18]	Counted “1”s in Table S1 for item “Exclusion/inclusion criteria” and divided by number of articles to calculate the percentage of compliance. The item was described “Gave specific information about why papers were retained or rejected”.
Gates [10]	We extracted from the main text the number of studies that reported an explicit criterion for inclusion of studies, and use it to calculate the percentage of compliance.
Koricheva & Gurevitch [7]	Data extracted from criterion “Reporting inclusion/exclusion criteria”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Inclusion/exclusion” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]	Data extracted from criterion 4.1 of the PRISMA EcoEvo checklist (Table 1): “Report the specific criteria used for including or excluding studies when screening titles and/or abstracts, and full texts, according to the aims of the systematic review (e.g. study design, taxa, data availability)”.
Pappalardo et al. [1] (added)	Each study was coded 1 if the authors explained the inclusion/exclusion criteria to determine which papers to include in their meta-analysis. We gave 0.5 points to studies in which this was partially explained. The study that used their own data was tagged as “not applicable” and it was filtered out to calculate the percentage of studies complying with this criterion.

Roberts et al. [26]	Data extracted from criterion “Defined inclusion/exclusion criteria for identification of relevant (evidence) studies”, Table 1.
Romanelli et al. [16]	Matched to the “3.1 Inclusion criteria documented for all studies” criteria column in their supplementary data file. We only consider the 63 meta-analyses, and did not included the systematic reviews. We assigned 0 to their score 0, 0.5 points to their score for partial compliance (1), and 1 point to their score of full compliance (3). We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
Reference list of primary studies	
Archmiller et al. [18]	Data extracted from criterion “List of References”.
Gates [10]	Data extracted from number of papers that “gave lists of studies that were included”.
Koricheva & Gurevitch [7]	Data extracted from criterion “Have full bibliographic details of primary studies included in a meta-analysis been provided?”
Lodi et al. [17]	Data extracted from criterion “Bilbiographic details” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]]	Data extracted from criterion 27.1 of the PRISMA EcoEvo checklist (Table 1): “References”.
Pappalardo et al. [1]	We assigned 1 to each study that provided the full references (column *ref.provided* indicating “yes”); we assigned 0.5 to the few studies in which only partial information was provided (e.g., only providing first author and year or only providing references for one of the analysis conducted in the paper). The paper that used their own data for the meta-analysis was considered as not relevant.
Philibert et al. [25]	Data extracted from criterion “References: a list of the references used for the meta-analysis is provided.”
Roberts et al. [26]	Data extracted from criterion “Provides references of all studies within the review”.
Meta-analytical model	
Archmiller et al. [18]	Counted number of "1"s for "Model choice" in their Table S1. Item was described as “Stated if fixed-effect or random-effects model was used or discussed between- and within-study variation”.
Beillouin et al. [19]	Data extracted from criterion “Statistical models are fully described”; percentage calculated from the data available in the supplementary data package, column “Quality_Model” (n = 217 meta-analyses).
Koricheva & Gurevitch [7]	Data extracted from criterion “Specifying the meta-analytical model”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Meta-analytical model” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]	Data extracted from criterion 12.1 of the PRISMA EcoEvo checklist (Table 1): “Describe the models used for synthesis of effect sizes”.

Pappalardo et al. [1]	The cases in which the column *data.analysis* indicated “not mentioned”, or “not clear”, were coded 0. When the model was reported it was coded 1. For the few cases in which a non-traditional meta-analytic model was used to analyze the effect sizes, but it was explained, we also coded it as 1.
Senior et al. [27]	Percentage of papers that did not report the meta-analytic model used was extracted from the text.
Dataset used in the meta-analysis	
Archmiller et al. [18]	Data extracted from criterion “Dataset used in meta-analysis”, detailed as “Provided data set used for meta-analysis in article, SI, or online”. We counted the number of “1s” and divided it by the number of studies to obtain the percentage.
Koricheva & Gurevitch [7]	Data extracted from criterion “Has the data set used for meta-analysis, including effect sizes and variances/sample sizes from individual primary studies and moderator variables, been provided as electronic appendix?”.
Lodi et al. [17]	Data extracted from criterion “Data” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]	Data extracted from criterion 18.2 in the PRISMA EcoEvo checklist (Table 1): “Share data required to reproduce the results presented in the manuscript”.
Pappalardo et al. [1]	We coded 1 the studies with a “yes” in the column *original.data.provided*; we also coded 1 studies with a “yes” in both the column *eff.provided* and *var.provided*. We coded 0 studies where no data was provided, only partial data was provided, or those few cases in which the appendix links did not work and the authors did not reply to the emails requesting the original dataset.
Philibert et al. [25]	Data extracted from criterion “Availability of the dataset” item, described as “The dataset is available in an electronic format or published directly in the paper”.
Data used to calculate effect sizes (raw data)	
Beillouin et al. [19]	Data extracted from criterion “The full dataset is available” (also detailed in text as “data of the original studies”); percentage calculated from the data available in the supplementary data package, column “Quality_Data_Sharing” (n = 217 meta-analyses).
Pappalardo et al. [1]	Data extracted from column *original.data.provided*, coding 1 the studies with a “yes”; 0.5 the studies with a “partially”, and 0 studies with a “no” (or those few cases in which the appendix links did not work and the authors did not reply to the emails requesting the original dataset).
Romanelli et al. [16]	Data extracted from criterion “5.2 Are the extracted data reported?” in their supplementary data file. We only consider the 63 meta-analyses, and did not included the systematic reviews. We assigned 0 to their score 0, 0.5 points to their score for partial compliance (1), and 1 point to their score of full compliance (3). We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
The number of studies and effect sizes	

Cadotte et al. [8]	Data extracted from the Supplementary Material 2. To calculate percent compliance we coded 1 the cases in which there was information for the two columns *No_publications* and *No_datasets_or_studies*.
ODea et al. [2]	Data extracted from criterion 20.1 of the PRISMA EcoEvo checklist (Table 1): "Report the number of studies and effect sizes for data included in meta-analyses".
Pappalardo et al. [1] (added)	We collected information on the number of effect sizes reported in each meta-analysis from the main text if available, from the supplementary material, or by counting rows in the supplementary data tables. We combined this information with the number of studies included in the meta-analysis reported by Pappalardo et al. [1] and if both were reported we coded the item "Report the number of studies and effect sizes" as 1. We corrected one mistake in the dataset by Pappalardo et al. [1]; The publication by Luo et al. (2015) was reported as having included 216 papers, but that was before excluding some papers, the correct final number of papers should have been 31, which is the one we used in this paper.
The software used	
Archmiller et al. [18]	We counted number of "1"s for "Statistical software used" in Table S1 and divided it by the number of studies to obtain the percentage of compliance.
Koricheva & Gurevitch [7]	Data extracted from criterion "Specifying the software used", Table 3.
Lodi et al. [17]	Data extracted from criterion "Software" in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
Nakagawa & Santos [23]	The Appendix reported the number of papers that did not report the software used, that we used to calculate the percentage of compliance.
ODea et al. [2]	Data extracted from criterion 13.1 of the PRISMA EcoEvo checklist (Table 1): "Describe the statistical platform used for inference (e.g. R)".
Pappalardo et al. [1]	The cases in which the column *software* indicated "not mentioned" were coded as 0, and cases in which any software was reported coded as 1.
Philibert et al. [25]	Data extracted from Table 3 that reported the software used in the meta-analysis to calculate percentage of papers reporting the software. Note that Philibert evaluated if the software "was made available" which is more stringent criteria than generally reporting the software.
Senior et al. [27]	Number or papers reporting the software were extracted from the "Data_Package_Part_3_Survey_Data" excel spreadsheet.
The packages used (if applicable)	
ODea et al. [2]	Data extracted from criterion 13.2 of the PRISMA EcoEvo checklist (Table 1): "Describe the packages used to run models".
Pappalardo et al. [1] (added)	When papers used a programming language, we coded this criterion 1 if they described the packages used.
The functions used (if applicable)	
ODea et al. [2]	Data extracted from criterion 13.3 of the PRISMA EcoEvo checklist (Table 1): "Describe the functions used to run models".

Pappalardo et al. [1] (added)	When papers used a programming language, we coded this criterion 1 if they described the functions used.
The code (if applicable)	
ODea et al. [2]	Data extracted from criterion 18.4 of the PRISMA EcoEvo checklist (Table 1): “Share analysis scripts”.
Pappalardo et al. [1] added	When papers used a programming language, we coded this criterion 1 if they provided the code.
The types of non-independence	
Archmiller et al. [18]	Data extracted from criterion “Quantified or stated impacts of nonindependence”.
Lodi et al. [17]	Data extracted from criterion “Multiple effect sizes”, explained as “If more than one estimate of effect size per study was included in the analysis, has potential non-independence of these estimates been considered?”.
ODea et al. [2]	Data extracted from criterion 14.1 of the PRISMA EcoEvo checklist (Table 1): “Describe the types of non-independence encountered (e.g. phylogenetic, spatial, multiple measurements over time)”.
Pappalardo et al. [1] (added)	Each study was coded 1 if the authors discussed instances of non-independence in their data. Common cases of non-independence mentioned were repeated measures, multiple data per study, multiple experimental levels, multiple data per species.
EXECUTION	
Weighed effect sizes by study precision	
Archmiller et al. [18]	We counted the number of "1"s in their item “Individual study weight calculation” in Table S1 and divided it by the number of studies to obtain the percentage of compliance.
Beillouin et al. [19]	Data extracted from criterion “Studies are weighted according to their accuracy”; percentage calculated from the data available in the supplementary data package, column “Quality_Model_weights” (n = 217 meta-analyses).
Koricheva & Gurevitch [7]	Data extracted from criterion “Have effect sizes been weighted by study precision or has the rational for using unweighted approach been provided?”
Lodi et al. [17]	Data extracted from criterion “Weighted effect sizes” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
Pappalardo et al. [1]	To quantify if a study weighted the effect sizes we first re-classified all the levels of the column *weighting* into “yes”, “no”, or “not mentioned”. To calculate the percent of studies that weighted their effect sizes, we assigned 1 to studies with “yes”, and 0 to studies that did not weight or that did not mention if they used weights. The study that conducted both weighted and unweighted analysis and only reported unweighted was coded as 1. We also quantified separately the studies that used more traditional meta-analysis weights (sample size, inverse variance).
Philibert et al. [25]	Data extracted from criterion “Weighting: Observations are weighted according to their level of accuracy in the statistical model”.

Romanelli et al. [16]	Data extracted from criterion “4.2 Studies were weighted according to the methodological quality?” in their supplementary data file. We only consider the 63 meta-analyses, and did not included the systematic reviews. In this case, we assigned 0 to their score 0 and 1 (if weighted was not transparent or removed studies), and only assigned 1 to their score of 3 (metric for weighting was clearly stated). We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
Vetter et al. [28]	The total from this percentage is from the 60 articles that fulfilled all requirements plus the 23 that fulfilled all but weighting (total = 83). So, the 37 that did weight represent a 44%. Information extracted from the text.
Tested for publication bias	
Archmiller et al. [18]	Counted “1”s in Table S1 for item “Publication Bias” and divided by number of articles to calculate the percentage of compliance. The item was described as “Discussed or quantified publication bias”.
Beillouin et al. [19]	Data extracted from criterion “Publication bias is analyzed”; percentage calculated from the data available in the supplementary data package, column “Quality_Publication_bias” (n = 217 meta-analyses).
Gates [10]	The number of papers that reported to discussed publication bias and that calculated the “fail-safe” number was extracted from the text. Percentage calculated based on the total numbers of studies analyzed.
Koricheva & Gurevitch [7]	Data extracted from criterion “Testing for publication bias”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Publication bias” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
Nakagawa & Santos [23]	The percentage of papers that utilized some kind of procedure was extracted from the main text.
ODea et al. [2]	Data extracted from criterion 24.1 of the PRISMA EcoEvo checklist (Table 1): “Provide results for the assessments of the risks of bias (e.g. Egger’s regression, funnel plots)”.
Pappalardo et al. [1] (added)	Each study was coded 1 if the authors addressed publication bias (usually this was done using funnel plots or the fail-safe number); the study that used their own data was filtered out to calculate the percentage of studies complying with this criterion.
Philibert et al. [25]	Data extracted from criterion “Investigation of publication bias”, described as “Assessment of the publication bias, which occurs when only studies with highly significant results are published. In this case, a meta-analysis can lead to a biased conclusion and an overestimation of the effect of a given factor. Publication bias is a predominant issue in meta-analysis and several methods such as funnel plots (e.g., Borenstein et al. 2009; Light and Pillemer, 1984) have been developed to detect the presence of such bias in datasets including published results.”
Roberts et al. [26]	Data extracted from criterion “Estimation of publication bias”, Table 1.
Romanelli et al. [16]	Data extracted from criterion “6.3 Does the synthesis consider possible publication bias?” criteria column in their supplementary data file. We only consider the 63 meta-analyses, and did not included the systematic reviews. For this criterion, we consider any way of addressing publication bias as complying, since that seems to be how other papers have measured it (including our data extraction for meta-analyses from Pappalardo et al. [1]). So we assigned 0 to their score 0, and 1 to their scores for partial (1) and full (3) compliance. We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
Conducted sensitivity analysis	

Archmiller et al. [18]	Counted “1”s in Table S1 for item “Sensitivity Analysis” and divided by number of articles to calculate the percentage of compliance. The item was described as “Quantified impact of individual effect sizes with sensitivity analysis or discussed potential impacts of individual study effect sizes”.
Gates [10]	The number of papers that perform a sensitivity analysis was extracted from the text. Percentage calculated based on the total numbers of studies analyzed.
Koricheva & Gurevitch [7]	Data extracted from criterion “Sensitivity analysis”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Sensitivity analysis” in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]	Data extracted from criterion 24.2 of the PRISMA EcoEvo checklist (Table 1): “Provide results for the robustness of the review’s results (e.g. subgroup analyses, meta-regression of study quality, results from alternative methods of analysis, and temporal trends)”.
Pappalardo et al. [1] (added)	Each study was coded 1 if the authors conducted some type of sensitivity analysis, such as comparing their results with or without influential points, comparing weighted and weighted analysis, or comparing different types of effect size measures.
Philibert et al. [25]	Data extracted from criterion “Sensitivity analysis”, described as “Analysis of the sensitivity of the conclusions to any change in the dataset and/or in the statistical method used to analyze the data. Sensitivity analyses should be carried out to identify influential data and to assess the robustness of the main conclusions of a meta-analysis to the assumptions made in the statistical analysis.”
Roberts et al. [26]	Data extracted from criterion “Sensitivity analysis”, Table 1.
Controlled for phylogenetic non-independence	
Cadotte et al. [8]	Data was extracted from the data table presented in the Supplementary Material, using additional metadata provided by Marc Cadotte. If the column *No_species* was higher than 3 (following the methods in Chamberlain et al. [20]), we quantified if studies corrected for phylogenetic non-independence using the column *Phylogeny_included*. To calculate the percent compliance, we coded as 1 studies where Cadotte et al. [8] reported the meta-analysis used a real phylogeny (*Phylogeny_included == 1)*, and as 0.5 when it was reported that the meta-analysis was for congeneric species (*Phylogeny_included == 2)*, or that it used taxonomic levels (*Phylogeny_included == 3*). The 4 studies that had a large number of species and Phylogeny_included was empty were coded 0.
Chamberlain et al. [20]	Information extracted from figure 1, only 2 papers performed a phylogenetic meta-analysis (coded as 1 for full compliance), and 19 assessed wether effect sizes differed among taxonomic categories (coded as 0.5 for partial compliance), 35 did traditional meta-analysis (coded as zero). We calculated the percentage of compliance from the total number of relevant papers that were analyzed (n = 56).
Jennions et al. [22]	From their Table 1, we extracted the number of papers that controlled for phylogeny, the number of papers that did not, and the number of papers for which it was not relevant. We calculated the percentage of compliance from the total number of relevant papers that were analyzed.
Koricheva & Gurevitch [7]	Data extracted from criterion “Controlling for phylogeny”, Table 3.
Lodi et al. [17]	Matched to the “Controlling for phylogeny” criterion in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.

Nakagawa & Santos [23]	From their Appendix, we extracted the percentage of papers that did not control for phylogenetic relatedness, and transform it into percentage of compliance using the total number of papers analyzed.
Pappalardo et al. [1] (added)	Each study was coded as 1 if the authors corrected the analysis for phylogenetic relatedness (when applicable), and was coded 0.5 if phylogenetic effects were partially addressed (e.g., by using family or taxonomy as a covariate). Community or ecosystem studies were coded “not applicable” and filtered out before calculating the percentage of papers complying with this criterion.

Quantifying heterogeneity in effect sizes

Archmiller et al. [18]	Counted “1”s in Table S1 for item “Heterogeneity estimation” and divided by number of articles to calculate the percentage of compliance. The item was described as “Calculated heterogeneity statistics (e.g. Q, s, I)”.
Beillouin et al. [19]	Heterogeneity data was combined in their category “Heterogenity of results is analyzed”; but we were able to calculate the percentage that quantified heterogeneity by counting the level “YES, quantification (I2, Qb) and subgroup analysis” in the column “Quality_Model_heterogeneity (I2, PI, …)” available from the supplementary data package spreadsheet (n = 217 meta-analyses).
Gates [10]	We extracted from the text the number of studies that “contained some exploration of the heterogeneity of results, using the Q statistic or another measure of homogeneity of effect sizes.
Koricheva & Gurevitch [7]	Data extracted from criterion “Quantifying heterogeneity in effect sizes”, Table 3.
Lodi et al. [17]	Data extracted from criterion “Heterogeneity in effect sizes” in Table 3. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch [7] list of criteria.
ODea et al. [2]	Data extracted from criterion 22.1 of the PRISMA EcoEvo checklist (Table 1): “Report indicators of heterogeneity in the estimated effect (e.g. I2, tau2 and other variance components)”.
Romanelli et al. [16]	Data extracted from criterion “6.2 Is heterogeneity in the effect of the Intervention/Exposure investigated statistically?” in their supplementary data file. We only consider the 63 meta-analyses, and did not included the systematic reviews. In this case, all the studies were scored at the maximum value (3) and tallied by us as 1. We calculated percent compliance as the (sum of points*100)/total meta-analyses reviewed.
Senior et al. [27]	We extracted from the text the percentage of studies that reported statistics associated with heterogeneity.
Vetter et al. [28]	We extracted from the text the percentage of articles in which “authors quantified heterogeneity using an index measure”.

Exploring causes of heterogeneity

Archmiller et al. [18]	Counted “1”s in Table S1 for item “Sub-group analysis” and divided by number of articles to calculate the percentage of compliance. The item was described as “Summarized effect sizes for subgroups or categorical covariates”.
Beillouin et al. [19]	Heterogeneity data was combined in their category “Heterogenity of results is analyzed”; but we were able to calculate the percentage that explored heterogeneity by counting both the level “YES, quantification (I2, Qb) and subgroup analysis” and “YES, analysis by moderators/subgroups” in the column “Quality_Model_heterogeneity (I2, PI, …)” available from the supplementary data package spreadsheet (n = 217 meta-analyses).

Chaudhary et al. [21]	Percentage extracted from the text for “studies that collected data on multiple predictor variables” (not to be confused with multifactor analysis that was reported separately by the authors).
Gates [10]	Number of papers that reported results for subgroups was extracted from the main text “Eighteen reviews reported results for various subgroups of the studies or individuals included. In most cases these appeared to have been prespecified, as they were natural divisions into different classes of study (for example, different trophic groups, different habitats or different environmental stresses)”. We calculated the percentage of compliance based on the numbers of papers analyzed.
Koricheva & Gurevitch [7]	Data extracted from criterion “Exploring causes of heterogeneity”,Table 3. It seems to be specific to meta-regression based on their question: “Have the causes of existent heterogeneity in effect sizes been explored by meta-regression?”.
Lodi et al. [17]	Data extracted from criterion “Causes of heterogeneity” in Table 3. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch, 2014 list of criteria.
ODea et al. [2]	Data extracted from criterion 15.1 of the PRISMA EcoEvo checklist (Table 1): “Provide a rationale for the inclusion of moderators (covariates) that were evaluated in meta-regression models”.
Philibert et al. [25]	Information extracted from text for criterion 3, Heterogeneity was defined as "The origins of the variability of the results are analyzed", or more detailed explanation “Analysis of the variability of the results of individual studies, including checking to see whether the results vary between the selected individual studies and, when relevant, investigation of the sources of between-study variability (e.g., using random effects model).”
Roberts et al. [26]	Data extracted from criterion “Investigation of sources of heterogeneity”, Table 1, that included both “sub-group” and “meta-regression”. Data was extracted from the “sub-group” category.
Senior et al. [27]	The number of studies that reported “Analyses to identify sources of heterogeneity (e.g., meta-regression/sub-setting)” was extracted from the main text and converted to percentage of papers based on the total number of papers analyzed.
Vetter et al. [28]	The percentage of articles in which “authors explored heterogeneity by including explanatory variables” was extracted from the main text.
Multifactorial analysis of moderators	
Chaudhary et al. [21]	Number and percentage extracted from the text for studies that conducted “multifactor analysis”.
Koricheva & Gurevitch [7]	Data extracted from criterion “Multifactorial analysis of moderators”, Table 3.
Lodi et al. [17]	Matched to the “Collinearity analysis” criterion in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch, 2014 list of criteria.
Senior et al. [27]	We extracted from the text the number of studies “in which several moderators are fitted in one model.
Exploring temporal changes in effect size	
Koricheva & Gurevitch [7]	Data extracted from criterion “Exploring temporal changes in effect size”, Table 3.

Lodi et al. [17]	Data extracted from criterion "Changes in effect size" in Tables 3 and 4. Detailed description of each criteria appears in Table 2. The authors followed Koricheva & Gurevitch, 2014 list of criteria.
Nakagawa et al. [24]	Data extracted from criterion "time-lag bias tests" in their Figure 2.
Pappalardo et al. [1] (added)	Each study was coded as 1 if the authors explored temporal changes in effect size, or 0 if they did not.

Appendix S4

Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg

From: A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Setup

```
# load objects we need, generated with the R code provided

load("objects/reporting.R")
load("objects/execution.R")
load("objects/toplot.R")

# nice format for plots

niceplot <- theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.title.x = element_text(face = "bold", size = 14,
                                    margin = margin(t = 20, r = 0, b = 0, l = 0)),
        axis.title.y = element_text(face = "bold", size = 14,
                                    margin = margin(t = 0, r = 20, b = 0, l = 0)),
        axis.text.x = element_text(size = 13),
        axis.text.y = element_text(size = 13),
        legend.text = element_text(size = 13),
        legend.title = element_text(size = 13),
        legend.position = "right")
```

Review discipline

We compared the percent compliance from the different review papers according to the research discipline. We did not observe any clear patterns among the different sub disciplines for neither the Reporting or Execution criteria:

```
safe_colorblind_palette <- c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
                             "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")

#scales::show_col(safe_colorblind_palette)

# create Figure S3

fig_s3 <- ggcharts::bar_chart(
```

```
     data= reporting,
     x= paper.id,
     y= percent,
     fill = topic.area,
     facet= criteria,
 ) +
  scale_fill_manual(values = safe_colorblind_palette) +
  labs(x = "Review ID", y = "Percent of papers complying with Reporting criteria") +
  niceplot +
  theme(legend.position = "bottom")

# save figure S3

ggsave("figures/Fig_S3.pdf", fig_s3, width = 17, height = 10, dpi = 300)

# display figure S3

fig_s3
```

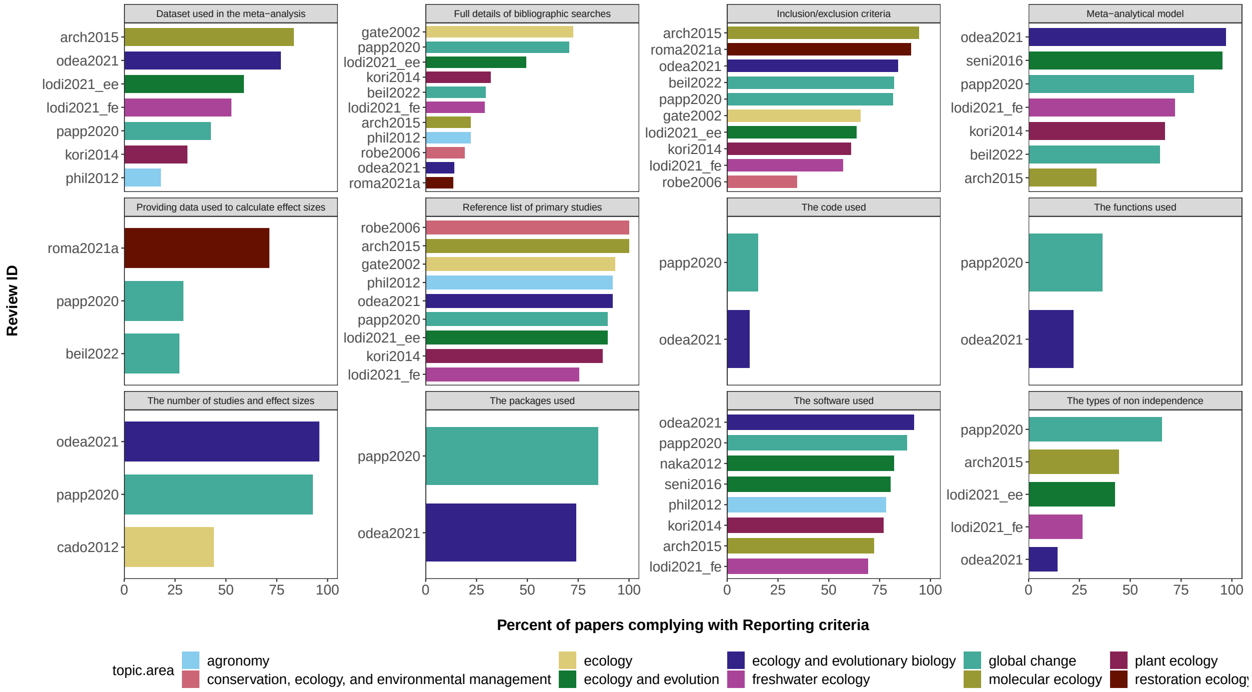

Fig S3. Percent of papers complying with Reporting criteria by review discipline. The percent of papers complying with each criterion is plotted for each synthesis paper. The colors indicate different subdisciplines of the review papers. The Review ID corresponds to the papers listed in Table 1.

```
# create Figure S4

fig_s4 <- ggcharts::bar_chart(
    data= execution,
    x= paper.id,
    y= percent,
    fill = topic.area,
```

```
    facet= criteria
  ) +
  scale_fill_manual(values = safe_colorblind_palette) +
  labs(x = "Review ID", y = "Percent of papers complying with Execution criteria") +
  ylim(0, 100)  +
  niceplot +
  theme(legend.position = "bottom")

# save figure S4

ggsave("figures/Fig_S4.pdf", fig_s4, width = 15, height = 10, dpi = 300)

# display figure S4

fig_s4
```

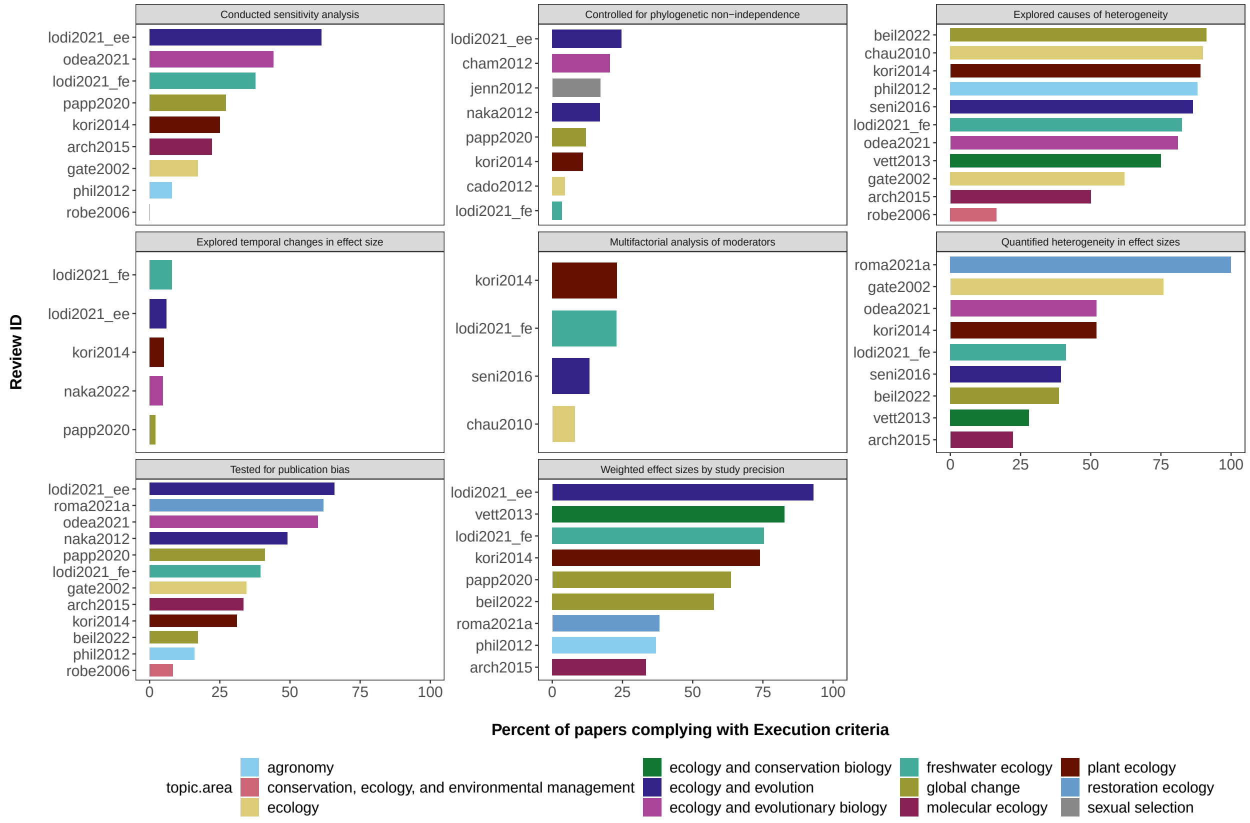

Fig S4. Percent of papers complying with Execution criteria by review discipline. The percent of papers complying with each criterion is plotted for each synthesis paper. The colors indicate different subdisciplines of the review papers. The Review ID corresponds to the papers listed in Table 1 of the main manuscript.

Temporal trends in compliance

To check for temporal trends in compliance we analyzed the time period of the papers included in each review paper and the percent compliance. We initially looked at:

1) the **mid-point** of the time period included in each review paper
2) the **earliest year** of the time period included in each review paper
3) the **most recent year** of the time period included in each review paper
4) the **full time period** included in each review paper

We considered that the **full time period** is the one that better represent the temporal data on a plot and present this below for the Reporting and Execution criteria. We did not observe any clear temporal trends.

```r
# create Figure S5

fig_s5 <- ggplot(reporting, aes(mid_point, percent)) +
  geom_segment(aes(x = first_year, y = percent,
                   xend = final_year, yend = percent), data = reporting) +
  niceplot +
  facet_wrap(~criteria) +
  #ggtitle("Reporting") +
  labs(x= "Time period", y= "Percent compliance with Reporting criteria") +
  niceplot

# save figure S5

ggsave("figures/Fig_S5.pdf", fig_s5, width = 10, height = 8, dpi = 300)

# display figure S5

fig_s5
```

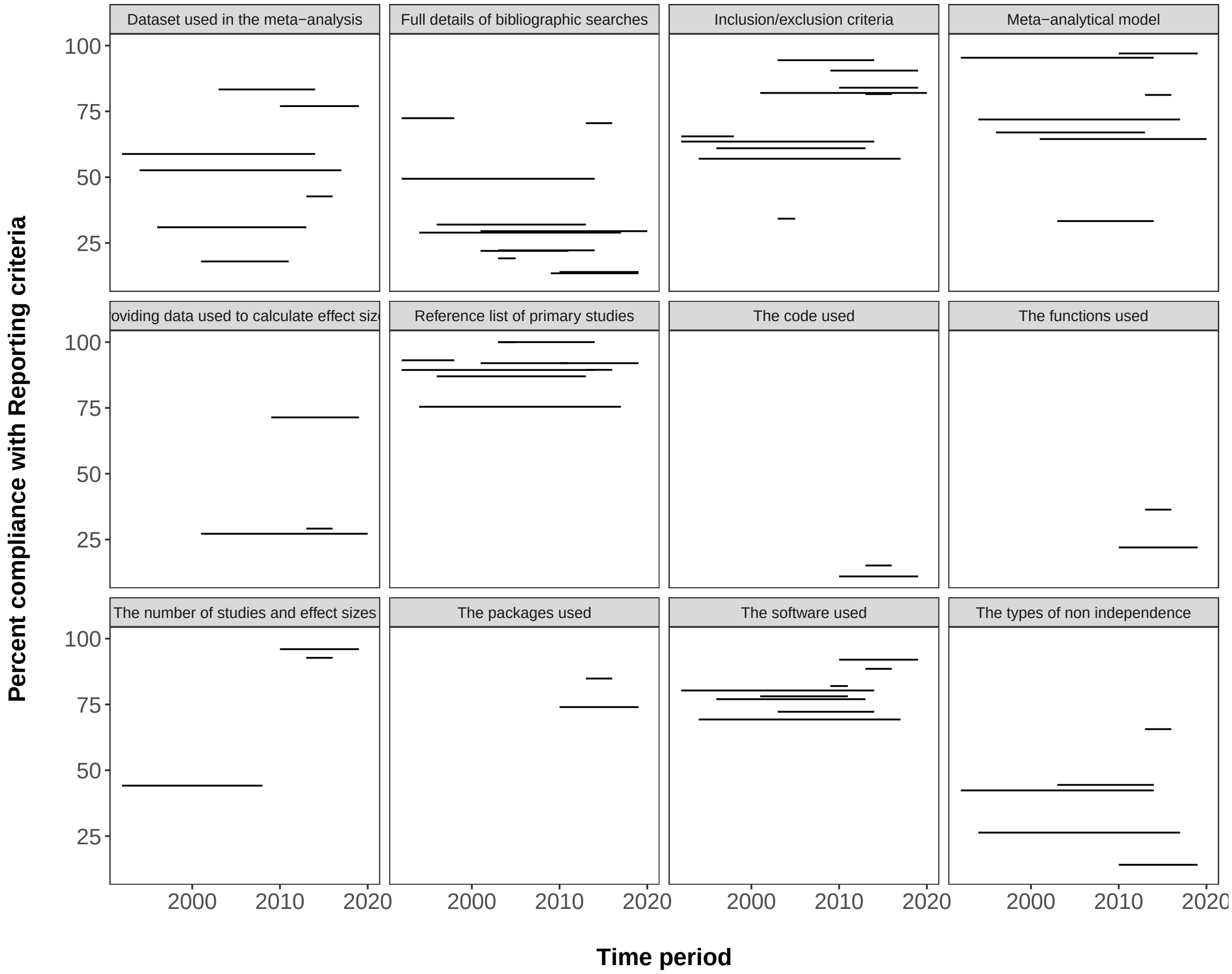

Fig S5. Percent of papers complying with Reporting criteria as a function of the time period analyzed by the review paper. Each panel represents a criterion. The line segment indicates the time period covered by each of the review papers that addressed a particular criterion.

```
# create Figure S6

fig_s6 <- ggplot(execution, aes(mid_point, percent)) +
  geom_segment(aes(x = first_year, y = percent,
                   xend = final_year, yend = percent), data = execution) +
  niceplot +
  facet_wrap(~ criteria) +
  #ggtitle("Execution") +
  labs(x= "Time period", y= "Percent compliance with Execution criteria") +
  niceplot

# save figure S6

ggsave("figures/Fig_S6.pdf", fig_s6, width = 10, height = 8, dpi = 300)
```

```
# display figure S6

fig_s6
```

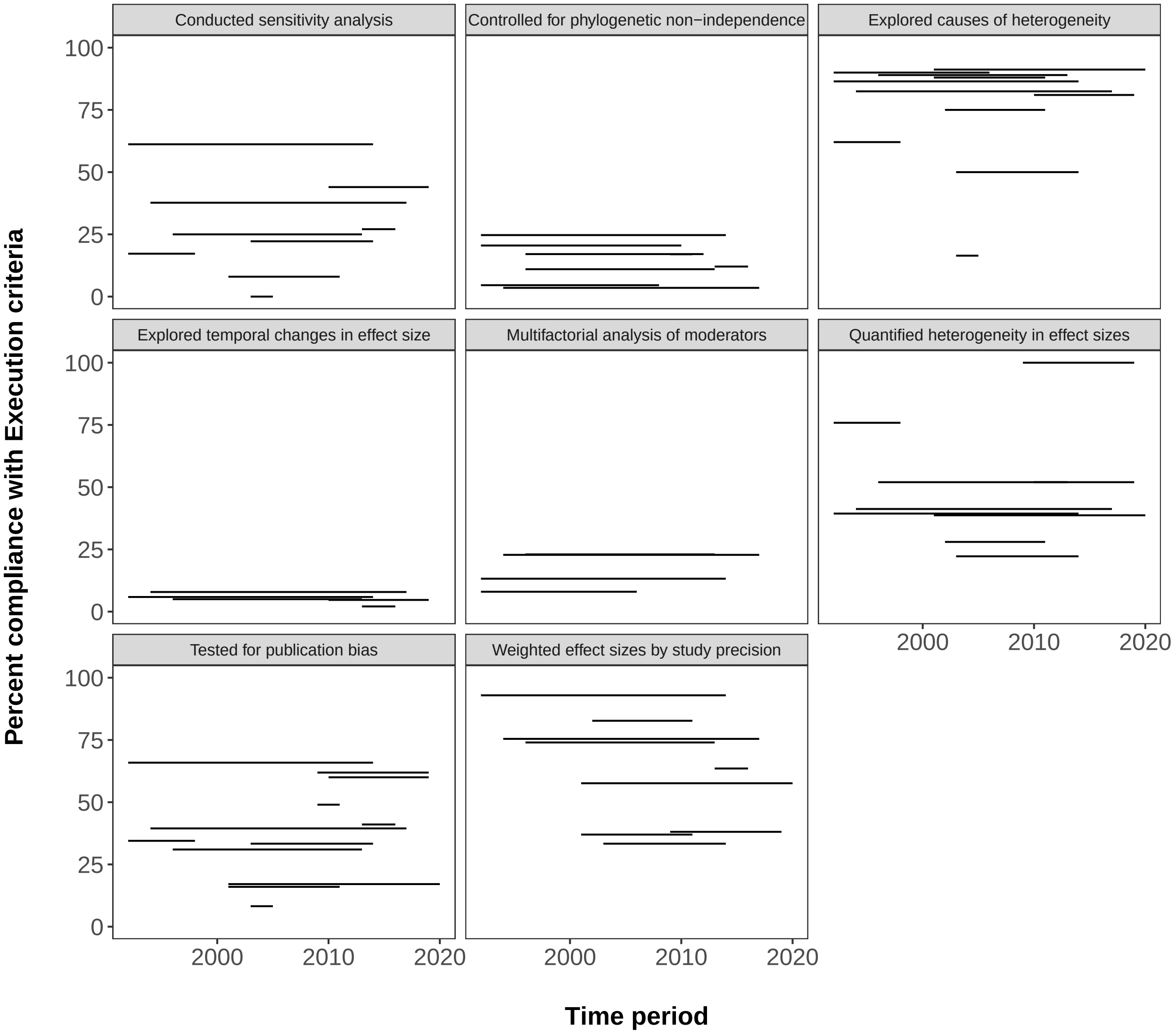

Fig S6. Percent of papers complying with Execution criteria as a function of the time period analyzed by the review paper. Each panel represents a criterion. The line segment indicates the time period covered by each of the review papers that addressed a particular criterion.

Appendix S5

Paula Pappalardo, Chao Song, Bruce A. Hungate, Craig W. Osenberg

From: A meta-evaluation of the quality of reporting and execution in ecological meta-analyses

Setup

```
knitr::opts_chunk$set(echo = T, eval = T, warning = F,
                      message = F, comment = "")

# load libraries we need

library(kableExtra)
library(tidyverse)
library(ggcharts)
library(rcartocolor)
library(readxl)
library(flextable)
```

Journals that most often publish meta-analyses

By pooling the reference list from all the review papers that made the references available (and without counting references twice), we were able to assess which journals tend to publish more meta-analyses. The top five journals were: 1) Ecology Letters (n= 91), 2) Global Change Biology (n= 89), 3) Ecology (n= 75), 4) Oecologia (n = 48), and American Naturalist (n = 43). Below you can see the number of meta-analyses per journal that had been included in the meta-analysis reviews. Because the distribution is strongly right skewed (with most journals publishing a few meta-analyses), we display only the journals with at least 5 meta-analyses.

```
# function to select columns of interest

selectColumns <- function(mydf){
  mydf_ed <- mydf %>% select(first_author, year, journal)
  return(mydf_ed)
}

# function to add a standarized reference ID to compare accross reviews
addID <- function(mydf){
  mydf_ed <- mydf %>%
    mutate(ref_id = paste(str_to_upper(first_author), year,
                          str_to_upper(journal), sep = "_"),
           journal = str_to_upper(journal))
```

```r
  return(mydf_ed)
}


# load short journals dictionary

journal_dic <- read.csv("data/DataFiles_journal-names-dictionary.csv", as.is = T)
```

```r
# load references for each publication

arch <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "arch2015", range= cell_cols("A:D"))) %>%
  selectColumns() %>%
  addID()

cado <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "cado2012", range= cell_cols("A:D"))) %>%
  selectColumns() %>%
  addID()

cham <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "cham2012", range= cell_cols("A:C"))) %>%
  addID()

gate <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "gate2002", range= cell_cols("A:C"))) %>%
  addID()

jenn <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "jenn2012", range= cell_cols("A:B"))) %>%
  left_join(journal_dic, by = "journal_short") %>%
  rowwise() %>%
  mutate(year = as.numeric(str_extract(reference, "\\d{4}")[[1]]),
         first_author = str_split_fixed(reference, " ", 2)[,1]) %>%
  ungroup() %>%
  selectColumns() %>% addID()

kori <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "kori2012", range= cell_cols("A:C"))) %>%
  addID()

odea <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "odea2021", range= cell_cols("A:C"))) %>%
  addID()

papp <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "papp2020", range= cell_cols("A:C"))) %>%
  addID()

phil<- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                sheet= "phil2012", range= cell_cols("A:C"))) %>%
  left_join(journal_dic, by = "journal_short") %>%
  selectColumns() %>% addID()
```

```r
robe <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "robe2006", range= cell_cols("A:C"))) %>%
  addID()

seni <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "seni2016", range= cell_cols("A:C"))) %>%
  addID()

vett <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "vett2013", range= cell_cols("A:C"))) %>%
  addID()

lodi_fe <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                    sheet= "lodi2021_fe", range= cell_cols("A:C"))) %>%
  mutate(first_author =  str_split_fixed(Citation, ",", 2)[,1],
         journal = ifelse(journal == "The American Naturalist",
                          "American Naturalist", journal)) %>%
  addID()

lodi_ee <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                    sheet= "lodi2021_ee", range= cell_cols("A:C"))) %>%
  addID()

beil <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                                 sheet= "beil2022", range= cell_cols("A:D"))) %>%
  addID()

roma_a <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                  sheet= "roma2021a", range= cell_cols("A:D"))) %>%
  addID()

roma_b <- as.data.frame(read_excel("data/DataFiles_list-of-references-in-Reviews.xlsx",
                  sheet= "roma2021b", range= cell_cols("A:C"))) %>%
  rowwise() %>%
  mutate(year = as.numeric(str_extract(Citation, "\\d{4}")[[1]])) %>%
  ungroup() %>%
  addID()
```

```r
# combine all unique ids (to avoid counting papers twice)

allids <- sort(unique(c(arch$ref_id, cado$ref_id, cham$ref_id, gate$ref_id,
              jenn$ref_id, kori$ref_id, odea$ref_id, papp$ref_id,
              phil$ref_id, robe$ref_id, seni$ref_id, vett$ref_id,
              lodi_fe$ref_id, lodi_ee$ref_id, beil$ref_id, roma_a$ref_id)))

# Compile all papers and create journal counts

alljournals <- data.frame(ids = allids) %>%
  mutate(journal = str_split_fixed(ids, "_", 3)[,3]) %>%
  add_count(journal) %>%
  distinct(journal, n) %>%
  arrange(desc(n)) %>%
  filter(n > 4)
```

```
flextable(alljournals, cwidth = 3)
```

journal	n
ECOLOGY LETTERS	91
GLOBAL CHANGE BIOLOGY	89
ECOLOGY	75
OECOLOGIA	48
AMERICAN NATURALIST	43
OIKOS	41
CONSERVATION BIOLOGY	40
JOURNAL OF ECOLOGY	34
BIOLOGICAL CONSERVATION	34
AGRICULTURE, ECOSYSTEMS & ENVIRONMENT	33
NEW PHYTOLOGIST	28
PROCEEDINGS OF THE ROYAL SOCIETY B	27
JOURNAL OF APPLIED ECOLOGY	26
GLOBAL ECOLOGY AND BIOGEOGRAPHY	24
ECOLOGICAL APPLICATIONS	22
FOREST ECOLOGY AND MANAGEMENT	22
EVOLUTION	21
MOLECULAR ECOLOGY	21
BEHAVIORAL ECOLOGY	20
ANIMAL BEHAVIOUR	18
PLANT AND SOIL	17
ANNUAL REVIEW OF ECOLOGY, EVOLUTION, AND SYSTEMATICS	16
MARINE ECOLOGY PROGRESS SERIES	15
JOURNAL OF ANIMAL ECOLOGY	15
SOIL BIOLOGY & BIOCHEMISTRY	15
FUNCTIONAL ECOLOGY	14
TRENDS IN ECOLOGY & EVOLUTION	13
BIOLOGY LETTERS	11

journal	n
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA	11
PLOS ONE	11
LAND DEGRADATION & DEVELOPMENT	10
JOURNAL OF EVOLUTIONARY BIOLOGY	10
BIOGEOSCIENCES	9
ECOLOGICAL MONOGRAPHS	9
AMERICAN JOURNAL OF BOTANY	9
ECOLOGY AND EVOLUTION	9
SCIENTIFIC REPORTS	8
ENVIRONMENTAL POLLUTION	8
SCIENCE	8
SCIENCE OF THE TOTAL ENVIRONMENT	8
BIOLOGICAL REVIEWS	8
ECOSYSTEMS	8
FIELD CROPS RESEARCH	8
FRESHWATER BIOLOGY	7
SOIL & TILLAGE RESEARCH	7
BEHAVIORAL ECOLOGY AND SOCIOBIOLOGY	7
CANADIAN JOURNAL OF FISHERIES AND AQUATIC SCIENCES	7
EVOLUTIONARY ECOLOGY	6
GLOBAL CHANGE BIOLOGY BIOENERGY	6
NATURE	6
BIOLOGICAL INVASIONS	6
ECOGRAPHY	6
PLANT, CELL AND ENVIRONMENT	5
RESTORATION ECOLOGY	5
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B	5
EVOLUTIONARY APPLICATIONS	5
ECOLOGICAL INDICATORS	5
PLANT	5

journal	n
CANADIAN JOURNAL OF FOREST RESEARCH	5
BIOGEOCHEMISTRY	5