Supporting Information

Preferred Reporting Items for Systematic reviews and Meta-Analyses in ecology and evolutionary biology: A PRISMA extension

Rose E. O'Dea[1], Malgorzata Lagisz[1], Michael D. Jennions[2], Julia Koricheva[3], Daniel W.A. Noble[1,2], Timothy H. Parker[4], Jessica Gurevitch[5], Matthew J. Page[6], Gavin Stewart[7], David Moher[8], Shinichi Nakagawa[1]

[1] Evolution & Ecology Research Centre and School of Biological and Environmental Sciences, University of New South Wales, Sydney, NSW, 2052, Australia.
[2] Research School of Biology, 46 Sulivans Creek Road, Australian National University, Canberra, 2600, Australia.
[3] School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, U.K.
[4] Department of Biology, Whitman College, Walla Walla, WA 99362, U.S.A.
[5] Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794-5245, U.S.A.
[6] School of Public Health and Preventative Medicine, Monash University, Melbourne, VIC, 3004, Australia
[7] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K.
[8] Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, General Campus, 501 Smyth Road, Room L1288, Ottawa, ON K1H 8L6, Canada.

ORCID IDs

Rose E. O'Dea: 0000-0001-8177-5075
Malgorzata Lagisz: 0000-0002-3993-6127
Michael D. Jennions: 0000-0001-9221-2788
Julia Koricheva: 0000-0002-9033-0171
Daniel W.A. Noble: 0000-0001-9460-8743
Timothy H. Parker: 0000-0003-2995-5284
Jessica Gurevitch: 0000-0003-0157-4332
Matthew J. Page: 0000-0002-4242-7526
Gavin Stewart: 0000-0001-5684-1544
David Moher: 0000-0003-2434-4206
Shinichi Nakagawa: 0000-0002-7765-5182

Contents

Methods: Development of PRISMA-EcoEvo

Formation of the PRISMA-EcoEvo working group

The process of extending PRISMA for ecology and evolutionary biology began with a meeting between D.M. [lead author of PRISMA 2009 (Moher *et al*., 2009)] and S.N., at the Ottawa Hospital Research Institute on 4 April 2018, and a virtual meeting between S.N., J.G., and T.H.P. on 25 April 25, 2018. After initial discussions (including one at the July 2018 meeting of the Society for Research Synthesis Methodology), S.N. assembled a working group with expertise in ecology and evolution systematic reviews and meta-analyses (M.L., M.D.J., J.K., D.W.A.N., T.H.P., J.G., and G.S.) to be led by his PhD student (R.E.O.). The PRISMA-EcoEvo project was formalised in an email invitation on 23 September 2018. On 27 September 2018, R.E.O. met with D.M. to discuss the development of PRISMA-EcoEvo. After the meeting with D.M., M.J.P. (lead author of the PRISMA 2020 update; Page *et al*., 2021*b*) was invited to join the project on 4 October 2018. At UNSW, Sydney, M.J.P. visited R.E.O. and S.N. on 6 December 2018, and D.M. visited on 28 February 2019. Discussions with all members of the working group occurred remotely.

Selection of items

R.E.O. began drafting the PRISMA-EcoEvo checklist in August 2018. First, R.E.O. created a checklist of previous reporting guidelines which were, in order of publication date: QUOROM (Moher *et al*., 1999); PRISMA (Moher *et al*., 2009); PRISMA for Abstracts (Beller *et al*., 2013); PRISMA for Protocols (Shamseer *et al*., 2015); PRISMA for Network Meta-analyses (Hutton *et al*., 2015); PRISMA for Individual Participant Data (Stewart *et al*., 2015); PRISMA for health inequity (Welch *et al*., 2016); PRISMA for harms outcomes (Zorzela *et al*., 2016); TTEE – Tools for Transparency in Ecology and Evolution (Parker *et* *al*., 2016*b*); and ROSES – RepOrting standards for Systematic Evidence Syntheses (Haddaway *et al*., 2018). Previous reporting checklists were used to create a longlist of potentially relevant items, which R.E.O. used as inspiration for the first draft of the PRISMA-EcoEvo checklist and manuscript outline.

Consensus on checklist items

From the initial draft of PRISMA-EcoEvo, the items were modified with iterative feedback from S.N., M.L., M.D.J., J.K., D.W.A.N., T.H.P., J.G., and G.S. After receiving comments from all members of the working group by 29 October 2018, R.E.O. prepared a second draft for approval. On 29 January 2019, RO invited members of the working group to complete a *Google* Form that presented each item of the revised PRISMA-EcoEvo checklist. The survey asked co-authors to select whether or not they approved each item and, if not, to indicate what about the item should be changed and why. All suggested changes were implemented. For the next six months, R.E.O. used the PRISMA-EcoEvo checklist to complete an assessment of reporting quality of meta-analyses published in ecology and evolutionary biology journals (methods detailed later in this supplementary information), along with S.N., M.L., M.D.J., D.W.A.N., J.K., and T.H.P. While developing this assessment we chose to separate components of each item into sub-items, as implemented in M.J.P.'s PRISMA update (Page *et al*., 2021*b*), as this granularity made it easier to assess reporting quality. While piloting the assessment process we made small modifications to the checklist, clarifying items where necessary. After the assessment of reporting quality, the checklist was sent back to all co-authors on the PRISMA-EcoEvo project, before opening it to comments from authors, reviewers, and editors of meta-analyses in ecology and evolutionary biology (details of this survey appear later in this supplementary information).

References

Beller, E.M., Glasziou, P.P., Altman, D.G., Hopewell, S., Bastian, H., Chalmers, I., Gøtzsche, P.C., Lasserson, T., Tovey, D. & for the PRISMA for Abstracts Group (2013). PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Medicine* **10**, e1001419.

Haddaway, N.R., Macura, B., Whaley, P. & Pullin, A.S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence* **7**, 7.

Hutton, B., Salanti, G., Caldwell, D.M., Chaimani, A., Schmid, C.H., Cameron, C., Ioannidis, J.P.A., Straus, S., Thorlund, K., Jansen, J.P., Mulrow, C., Catalá-López, F., Gøtzsche, P.C., Dickersin, K., Boutron, I., *et al*. (2015). The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Annals of Internal Medicine* **162**, 777.

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D. & Stroup, D.F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet* **354**, 1896–1900.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLoS Medicine* **6**, e1000097.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., *et al*. (2021*b*). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71.

PARKER, T.H., NAKAGAWA, S., GUREVITCH, J. & IIEE (IMPROVING INFERENCE IN EVOLUTIONARY BIOLOGY AND ECOLOGY) WORKSHOP PARTICIPANTS (2016*b*). Promoting transparency in evolutionary biology and ecology. *Ecology Letters* **19**, 726–728.

SHAMSEER, L., MOHER, D., CLARKE, M., GHERSI, D., LIBERATI, A., PETTICREW, M., SHEKELLE, P., STEWART, L.A. & THE PRISMA-P GROUP (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* **349**, g7647.

STEWART, L.A., CLARKE, M., ROVERS, M., RILEY, R.D., SIMMONDS, M., STEWART, G. & TIERNEY, J.F. (2015) Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: The PRISMA-IPD statement. *JAMA* **313**, 1657.

WELCH, V., PETTICREW, M., PETKOVIC, J., MOHER, D., WATERS, E., WHITE, H., TUGWELL, P., ATUN, R., AWASTHI, S., BARBOUR, V., BHUTTA, Z.A., CUERVO, L.G., GROVES, T., KOEHLMOOS-PEREZ, T., KRISTJANSSON, E., *ET AL*. (2016). Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. *Journal of Clinical Epidemiology* **70**, 68–89.

ZORZELA, L., LOKE, Y.K., IOANNIDIS, J.P., GOLDER, S., SANTAGUIDA, P., ALTMAN, D.G., MOHER, D., VOHRA, S. & PRISMA HARMS GROUP (2016). PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ* **352**, i157.

Examples for each item

For each sub-item of the PRISMA-EcoEvo checklist (see main text, Table 1), we here describe current reporting practices and provide an example from a published meta-analysis on an ecological or evolutionary topic. Most examples were found during our assessment of reporting quality (described later in this supplementary information). Each example (italicised) is listed alongside the corresponding sub-item number, or numbers, from Table 1 in bold. Some examples have been edited for the sake of brevity.

Item 1: Title and abstract

Estimated reporting quality across the literature: High

Nearly every assessed abstract summarised the aims and scope of the review (97%), the results of the primary outcome (96%), and conclusions (94%). Most described the data set (74%), and all identified the review as a meta-analysis (100%; although our sample of papers is biased towards this sub-item — details of the sample are provided below). The exception to high reporting quality was the reporting of limitations (only 17% of assessed papers mentioned limitations in the abstract).

Example: Kettenring & Adams (2011)

[Title: Lessons learned from invasive plant control experiments: a systematic review and meta-analysis: Invasive plant control experiments (**1.1**)]

Abstract:

[*Invasive plants can reduce biodiversity, alter ecosystem functions and have considerable economic impacts. Invasive plant control is therefore the focus of restoration research in invader-dominated ecosystems … In a systematic review and meta-analysis of invasive plant control research papers, we asked: (i) what control efforts have been most successful; and (ii) what invasive plant control research best translates into successful restoration application?*] (**1.2, 1.3**)] [*The literature evaluated typically described experiments that were limited in scope. Most plot sizes were small* ($<1 \text{ m}^2$)*, time frames were brief (51% evaluated control for one growing season or less) and few species and ecosystems (predominantly grasslands) were studied throughout much of the literature. The scale at which most experiments were* *conducted potentially limits relevance to the large scales at which restorations typically occur. Most studies focused on invasive species removal and lacked an evaluation of native revegetation following removal. Few studies (33%) included active revegetation even though native species propagule limitation was common. Restoration success was frequently complicated by re-invasion or establishment of a novel invader. Few studies (29%) evaluated the costs of invasive species control. Additionally, control sometimes had undesirable effects, including negative impacts to native species.* (**1.6**)] [*Despite a sizeable literature on invasive plant control experiments, many large-scale invasive plant management efforts have had only moderate restoration success. We identified several limitations to successful invasive species control including: minimal focus on revegetation with natives after invasive removal, limited spatial and temporal scope of invasive plant control research, and incomplete evaluation of costs and benefits associated with invasive species management actions* (**1.4**)]. [*We suggest that information needed to inform invasive plant management can be better provided if researchers specifically address these limitations. Many limitations can be addressed by involving managers in research, particularly through adaptive management* (**1.5**)].

Item 2: Aims and questions

Estimated reporting quality across the literature: High

All papers provided a rationale for the study (100%), and nearly all stated the primary aims and questions (96%), described the scope of the study (91%), and (where applicable) referenced previous reviews (93%). The exception to high reporting quality was that only 57% of assessed papers described whether effect sizes were derived from experimental and/or observational comparisons.

Example: Breed *et al*. (2015)

[*Habitat fragmentation is a globally pervasive problem that continues to drive changes to woody plant ecosystems. As most species of woody plants are animal-pollinated, studying the impacts of fragmentation on woody plant–pollinator interactions seems particularly important, especially because significant amounts of biodiversity rely on these plant–pollinator interactions. ... Pollen diversity should decline with lower conspecific density because as density declines the number of different pollen sources received into a given canopy is also expected to decline. This relationship is due to reduced numbers of pollen donors in the landscape and the fact that animal pollinators are less likely to shift from one plant to another because of the imposed costs of movement. As a consequence of reduced numbers of pollen sources, the correlated paternity within a given progeny array should increase* (**2.1**)].

[*A recent review of outcrossing rates in undisturbed versus disturbed plant populations across 27 species confirmed the expectation of decreased outcrossing rate in disturbed plant populations (Eckert et al., 2010). ... The effect of pollinator mobility on mating-pattern shifts in fragmented systems has been highlighted as a research gap (Ghazoul, 2005; Lowe et al., 2005; Eckert et al., 2010), but to the best of our knowledge no previous study has comprehensively explored this topic. ... No previous review has performed a quantitative assessment of habitat fragmentation or density effects on mating patterns of plants. Furthermore, previously published qualitative reviews have focussed on outcrossing rates of Neotropical trees, despite other mating-pattern data, taxa and regions being well represented (for example, pollen diversity, shrubs, southern Australia and east Asia)* (**2.2**)].

[*…we examine the relationships between habitat fragmentation and mating patterns in animal-pollinated woody plants. We present an in-depth study of mating patterns of three closely related eucalypt species that vary in the mobility of their pollinators ... We then investigate the generality of habitat fragmentation–mating-pattern relationships of animal-pollinated woody plants with a meta-analysis* (**2.3**)].

[Pollinator mobility:

…species with less mobile pollinators should theoretically be more sensitive to the drivers of these fitness effects. … One factor that may mitigate against decreased outcrossing in fragmented systems is mobile pollinators. Consequently, the generic expectation of decreased outcrossing with increased habitat fragmentation may need refining to be a function of pollinator mobility. In cases where pollinators have limited mobility (for example, small insects), these pollinators will tend to shift from plant to plant less freely than mobile pollinators, increasing the degree of pollen discounting and reducing the diversity of available pollen.

Selfing rate:

Pollen diversity (often measured by correlated paternity, rp) is a parameter that also describes the mating patterns of plants … Pollen diversity should decline with lower conspecific density because as density declines the number of different pollen sources received into a given canopy is also expected to decline … The effect of density on correlated paternity should again be a function of pollinator mobility, as more mobile pollinators should overcome greater distances between canopies more easily than less mobile pollinators, resulting in larger pollination neighbourhoods. (**2.4**)]

[*We sampled 199 progeny arrays from 13 groups of maternal plants within a single landscape in southern Australia across a habitat fragmentation gradient* (**2.5**)].

Item 3: Review registration

Estimated reporting quality across the literature: Very low

Only 3% of assessed papers provided a registration, and none of those papers reported or justified all deviations from registered aims or methods.

Example: Merkling *et al*. (2018)

[*In line with the recent push for more transparent and credible research, our literature search and subsequent analyses were preregistered on the Open Science Framework website to avoid post-hoc interpretation of our results (https://osf.io/vfnhg/register/565fb3678c5e4a66b5582f67)* (**3.1**)]

[*…although not originally pre-registered, we also tested for the effects of the timing of sex-ratio measure (primary or secondary) and the origin of animals (captive/wild).* (**3.2**)]

[*Yolk hormones have also been suggested to influence sex determination and a few studies have tested this hypothesis. To provide a more complete view of the influence of testosterone on sex determination, we included an exploratory meta-analysis investigating the correlation between yolk testosterone and offspring sex that was not part of the initial preregistration.* (**3.3**)]

Item 4: Eligibility criteria

Estimated reporting quality across the literature: High

Eligibility criteria were usually reported (84%), and justifications were given for around half of the eligibility criteria (54%).

Example: De Boeck *et al*. (2018)

Supplement:

[*Studies had to meet six specific criteria to be included in the meta-analysis. (1) Community-level above-ground biomass was measured in plots of different plant species richness. We excluded studies that only considered functional diversity, genetic diversity or a diversity index to avoid skewed comparisons. (2) The climate event had a return time of one in ten years (or more extreme). For studies with natural events, this meant a z-value of at least 1.28. Imposed droughts in all but one case involved complete removal of precipitation, leading to atmospheric droughts that would (if natural) be record-breaking in many cases, and substantially surpassing one in ten year return times in all cases. In the only study with an imposed wet event1, precipitation was experimentally manipulated in line with the wettest year in a local 17-year dataset (a similar method was used regarding the dry-year treatment in this study). (3) The climate event followed a non-extreme year. (4) Baseline data was available for a 'normal' reference year present before the first reported event or a control was available in the case of imposed events. (5) No post-event recovery period was included in the first biomass harvest after the extreme event. (6) Species composition was not confounded with species richness in artificially assembled communities. We further excluded studies where the artificial diversity gradient was compromised by invasion of species from the surrounding area at the moment of the climatic event. Studies were separated according to type of precipitation extreme (dry or wet).* (**4.1**)]

[*We refrained from further separation into subgroups (e.g. according to the manner of assembly) as the number of studies was deemed insufficient.* (**4.2**)]

Item 5: Finding studies

Estimated reporting quality across the literature: Moderate

Most papers stated the sources of information that were sought (89%), around half provided the Boolean search strings (49%), a quarter defined whether the search was comprehensive (25%), and very few provided sufficient information to repeat the equivalent search (14%).

Example: Davidson *et al*. (2017)

[*We comprehensively searched published literature using standard techniques (detailed in Appendix S1, Supporting Information).* (**5.1**)]

[*Keyword searches in Scopus and Web of Science were used to identify peer-reviewed literature quantifying the effects of livestock grazing on the EPs of salt marshes. … Additionally, the doctoral thesis repository EThOS was searched in November 2015 for PhD theses containing these terms. …Only papers published in English, or with a comprehensive English summary, were included.* (**5.2**)]

[*(saltmarsh OR salt marsh OR salt-marsh OR salt meadow OR coastal marsh) AND (graz OR herbivor OR livestock OR cow OR cattle OR sheep OR horse OR deer) in their title, abstract or keywords.* (**5.3**)]

[*We retrieved papers published from 1950 to November 2015* (**5.4**)]

Item 6: Study selection

Estimated reporting quality across the literature: Very low

Only 13% of assessed papers reported how studies were selected for inclusion during screening, and only 3% reported who was involved in screening.

Example: Merkling *et al*. (2018)

[*Two authors independently duplicate-screened 350 article abstracts with an 80% agreement rate, using AbstrackR software. The screening of the remaining abstracts was shared between the same two persons and in unclear cases both persons screened the abstract. … Inclusion criteria applied to the 194 full-texts were: (i) maternal testosterone (not yolk testosterone) was measured or manipulated before or shortly after the eggs were laid/offspring were conceived, (ii) offspring sex-ratio of mothers with known testosterone levels/subject to the testosterone manipulation was measured and (iii) suitable information about statistics and sample sizes (N > = 5) were reported in the paper or there was indication that the needed information could be provided by the authors. After applying these criteria, 16 studies with suitable data remained.* (**6.1**)]

[*Two authors independently duplicate-screened 350 article abstracts with an 80% agreement rate, using AbstrackR software. The screening of the remaining abstracts was shared between the same two persons and in unclear cases both persons screened the abstract.* (**6.2**)]

Item 7: Data collection process

Estimated reporting quality across the literature: Low

Just over half of the assessed papers described the construction of moderator variables from other data (56%), under half described how data were collected (42%) and where they were collected from (44%), and very few described who collected data (10%) or whether data were checked for errors (1%).

Example: O'Dea *et al*. (2019)

[*Data were extracted from text, tables, or figures.* (**7.1**)]

[*To extract data from figures, we used the metaDigitise package in R. … To minimize errors, data were entered into a relational database, built using Filemaker Pro software* (**7.2**)]

[*Phenotypic traits were … grouped into four broad categories: (a) behaviour (behaviour); (b) life history (growth); (c) morphology (bone number, condition, morphology, scale number, size); and (d) physiology (heart, metabolism, muscle fibre, swim performance).* (**7.3**)]

[*Where sample sizes or variance was missing, we attempted to contact authors for this information. All contacted authors (n = 8) were asked whether they could provide additional data that could be used in our meta-analysis. Five authors replied to requests for data, and two provided data used in analysis. … For each species represented in the database, we gathered information on thermal tolerance and life history from the websites Fishbase and Animal Diversity Web (when information from these two sources was conflicting, we took the average of the two values).* (**7.4**)]

[*All data were extracted by one author (RO)* (**7.5**)] … [*but to verify these extractions half of the data (50% of papers) were checked by other authors.* (**7.6**)]

Item 8: Data items

Estimated reporting quality across the literature: High

Nearly every paper described the key data that were extracted (96%), the majority described main assumptions or simplifications (62%) and the type of replication unit (73%), but less than half described additional data extractions (42%).

Example: Bateman & Bishop (2017)

[*…studies reporting the individual or total abundances of associated invertebrate taxa, and/or associated invertebrate species density (i.e. species richness per unit area) in otherwise similar habitat with (experimental) and without bivalves (control) were considered. To be included in analyses, papers also needed to include (1) means, (2) estimates of variation about the means and (3) sample sizes or raw data. ... contrasts were categorised according to bivalve taxon (mussel, oyster or pinnid) and the density of bivalves ... communities were categorised according to their habit, or mode of life, and the broader habitat context in which they were found. ... Habitats were categorised as soft sediment or rocky shore. ... we compared ecosystem engineering across environmental gradients of known stressors to intertidal invertebrates, including tidal elevation and latitude, over which abiotic stressors such as temperature and humidity may vary.* (**8.1**)]

[*High intertidal elevations could not be included due to insufficient data.* (**8.2**)]

[*Habitats were (1) infaunal — living in sediment under the bivalve matrix or in sediment trapped in the interstices of the bivalve matrix, (2) mobile epifaunal — living on the surface of the bivalve matrix or bare substratum with a mobile adult life-history stage, or (3) sessile epifaunal — also living on the surface but with a non-mobile adult life-history stage. ... Three tidal elevations were considered: mid intertidal (aerially exposed for ≥4 h but no more than 8 h per semi-diurnal tidal cycle), low intertidal (aerially exposed on every tidal cycle, but for <4 h per semi-diurnal tidal cycle) and subtidal elevations (permanently submerged). ... Analyses examining relationships between ecosystem engineering and latitude utilised only data collected from the intertidal zone of either sedimentary or rocky shores, as this is where invertebrates are exposed to the greatest variations in climatic factors, such as temperature and humidity, that can induce stress* (**8.3**)]

[*Within phyla, individual taxa were not considered as replicates due to issues of spatial non-independence arising from these being sampled from the same plots. ... Hence, for all other analyses, we used total invertebrate abundance or species density as our metric for analysis.* (**8.4**)]

Item 9: Assessment of individual study quality

Estimated reporting quality across the literature: Very low

Very few papers described whether the quality of studies included in the meta-analysis was assessed (7%) or how study quality was incorporated into analyses (6%).

Example: Ord *et al*. (2011)

[*…for every robust study we examined, we found an almost equal number of experiments (25 of 55; Table 4) that were poorly designed (10 of 55), lacked statistical power (13 of 55 experiments tested ≤ 10 subjects) or authors concluded species discrimination based on a subset of significant tests out of a larger set of nonsignificant results (2 of 55; Table 4)* (**9.1**)]

[*To assess the impact of these studies on our hypothesis tests, we excluded them from a second series of meta-analyses and obtained virtually identical results to those reported in Table 1 (see Table S3).* (**9.2**)]

Item 10: Effect size measures

Estimated reporting quality across the literature: High

Nearly every study described which effect size was used to synthesise results (97%) and the majority provided a reference to that effect size (63%), but very few studies that used a non-conventional effect size derived the equations and stated the assumed sampling distribution (7%).

Example: Gorné & Díaz (2017)

Section 1.2: *Derived effect sizes*

[*…in our database the log transformation of the unbiased SMD approximates well to a normal distribution. Therefore, it is reasonable to assume that meta-analysis assumptions are being met with use of the log10(*|g|*) (LG hereafter).* (**10.3**)]

[*Here we propose a new ES, called "LG" and its variance "V_{LG}" calculated according to the propagation errors theory*: (**10.1**)]

$$LG = \log_{10}(|g|)$$

$$= \frac{\ln(|g|)}{\ln(10)}$$

$$= \frac{1}{\ln(10)} \times \ln(|g|)$$

$$\sigma_{LG}^2 = \ln^{-2}(10) \times \frac{\sigma_g^2}{g^2} \approx 0.1886 \times \frac{V_g}{g^2}$$

…we computed SMD as in equations 2, 3, and 18 … When available, we used means, standard deviations (or standard error or variance), and sample size to compute the unbiased standardized mean difference (Hedges' g), as detailed above, and its variance (Vg) as follows (Nakagawa & Cuthill, 2007):

$V_g = V_d = \frac{\mathrm{n}_1+\mathrm{n}_2}{\mathrm{n}_1\times\mathrm{n}_2} + \frac{d^2}{2\times(\mathrm{n}_1+\mathrm{n}_2-2)}$ (**10.2**)]

Item 11: Missing data

Estimated reporting quality across the literature: Low

Less than half of studies with missing data described (37%) or justified (21%) how the missing data were analysed.

Example: Pearson *et al*. (2016)

[*For the 8% of studies included in our meta-analysis that did not report variance metrics necessary for calculation of d, we used the method of Wolf and Guevara (2001) to estimate variance based on that reported in other relevant studies. Specifically, for each response variable and abundance metric, we doubled the largest reported standard deviation and assigned this value to all missing standard deviation data.* (**11.1**)]

[*This conservative procedure ensured that all cases reporting valid means were utilized while ensuring that estimated values were down-weighted in the meta-analysis.* (**11.2**)]

Item 12: Meta-analytic model description

Estimated reporting quality across the literature: High

Almost every paper stated the type of model used to synthesise effect sizes (97%). Among papers that used a non-conventional model (*n* = 40), only half provided justification (50%).

Example: Thomsen *et al*. (2018)

[*Prior to meta-analyses, non-independent within-study effect sizes were averaged (1) across repeated measurements, (2) for multiple densities of the primary or secondary FS, (3) for multiple FS and inhabitants, (4) for multiple responses per inhabitant, and (5) across nested and orthogonal experimental designs, for example across seasons and water depth level … We used weighted random-effects models … In addition, we carried out a full set of matching unweighted analyses based on the log response ratio (ln(co-occurring primary and secondary FS/primary FS alone)), including studies that reported effects without associated data dispersion (using a unit variance of 1 and bootstrapped confidence intervals).* (**12.1**)]

[*These analyses were important because more primary studies could be included (104 papers for abundance and 57 for richness), increasing the taxonomic and spatio-temporal generality of our conclusions. However, unweighted analyses are less statistically robust than weighted analyses and should be interpreted more cautiously. Nevertheless, the ecological conclusions derived from the unweighted analyses fully support the conclusions based on the weighted analyses…* (**12.2**)]

Item 13: Software

Estimated reporting quality across the literature: Moderate

Almost every paper described the statistical platform (92%) and the majority described what packages were used (74%), but only a minority described functions (22%), version numbers (33%), and arguments that differed from the default settings (29%).

Example: Uller *et al*. (2013)

[*the statistical environment R (version 2.15, R Core Team, 2013)* (**13.1**)]

[*using the library MCMCglmm (version 2.16, Hadfield, 2010)* (**13.2, 13.3, 13.5**)]

[*For the random effects, we used an inverse Wishart prior with V = 0.002 and nu = 1, which is widely used in the statistical literature (Gelman & Hill, 2007). For each statistical model, we ran three MCMC chains (i.e. three independent runs of MCMCglmm models) to test for convergence of model parameters among the chains. For every chain, we used the same settings for sampling: (i) the number of iterations of 2 000 000, (ii) the thinning interval of 500 and (iii) the number of burn-in of 1 500 000.* (**13.4**)]

Item 14: Non-independence

Estimated reporting quality across the literature: Moderate

The majority of studies described how non-independence had been handled (74%), just under half justified this decision (47%), and only a third provided a complete description of the types of non-independence present in the data set (32%).

Example: Jiang *et al*. (2013)

[*A confounding factor in any meta-analysis of assortative mating is nonindependence (or pseudoreplication) in the data. There are several possible sources. The most obvious comes from multiple studies of the same trait in the same species. … A second source of pseudoreplication can arise from using separate estimates of assortment for multiple traits in the same species. These estimates will not be independent when the traits are phenotypically correlated. … A third source of pseudo-replication comes from phylogenetic relationships. Clearly, two sibling species that have recently diverged are likely to share similar patterns of assortative mating for purely historical reasons. The same effect occurs to different degrees at all levels of phylogenetic relationship.* (**14.1**)]

[*We controlled for [nonindependence from multiple studies of the same trait in the same species] by analyzing the mean values across studies for species-trait combinations.* (**14.2**)]

[*We were unable to correct for [separate estimates of assortment for multiple traits in the same species] because we lack data on correlations between traits tested for assortment. Further, most studies in our database include results for only a single trait. … In principle, it is possible to correct for phylogenetic dependencies using a phylogeny for all species in the database and a plausible null model for how assortative mating evolves. Since we lack both of those ingredients, we treated species as independent observations. In any event, we know of no reason why these possible causes of non-independence in our data might bias our general conclusions.* (**14.3**)]

Item 15: Meta-regression and model selection

Estimated reporting quality across the literature: High

The majority of applicable papers provided a reason for the inclusion of moderators (81%) and described model selection when it was used (80%). Only one in five described whether sample sizes were sufficient to justify the choice of moderators (20%).

Example: Des Roches *et al*. (2018)

[*…a thorough understanding of the ecological effects of intraspecific variation will be critical for predicting how rapid, widespread changes in biodiversity within species will impact communities and ecosystems … Our analysis generalizes across diverse response variables, such as population abundance, rates of ecological processes and community composition at different trophic levels. In addition, we incorporate both direct (consumption or excretion) and indirect interactions (mediated through another organism or nutrient).* (**15.1**)]

[*The limited sample size of studies meant that we did not have sufficient degrees of freedom to test the effect of different experimental design moderators on different response variables.* (**15.2**)]

[*Our base model specified focal species and study (by publication) as nested random effects (focal species (study)) to account for heterogeneity and non-independence of results from the same study or using the same focal species. We used this base random-effects model to estimate an overall Hedges' g and 95% CIs. We then included attributes of study design and characteristics of the ecological response variables as moderators in two separate mixed-effects models. Aspects of study design that we could consistently determine included habitat* *(aquatic or terrestrial) and experimental setting (natural, laboratory or field). We also recorded focal species' trophic level (primary producer, primary consumer or secondary consumer) and whether the species treatment was replacement or removal; however, given that all removal studies used consumer species and most replacement studies used producer species, these two moderators were largely confounded. In an additive model, no aspects of study design explained significant variation in the effect size (P > 0.05). Our final mixed-effects model therefore only included the following two moderators and their interaction: relationship with the focal species (direct or indirect) and type of ecological response (abundance, rate or composition). We excluded a third response characteristic (trophic level: nutrient, producer or consumer) through model selection (glmulti package version 1.0.7. using the Akaike information criterion).* (**15.3**)]

Item 16: Publication bias and sensitivity analyses

Estimated reporting quality across the literature: Moderate

Over half the assessed papers described publication bias tests (65%). When authors reported evidence of publication bias (in 30 papers), only half described steps taken to investigate the effect of that bias on the results (47%). A minority of papers described other assessments of the robustness of results (e.g. sensitivity analyses) (35%).

Example: Kornder *et al*. (2018)

[…*(b) funnel plots were investigated to assess distribution properties and potential for publication bias; … (d) meta-regression with publication year as independent variable was used to assess whether time-related factors have influenced outcomes…* (**16.1**)]

[…*(c) fail-safe analysis was performed to address robustness against publication bias;* (**16.2**)]

[…*(a) a random permutation test was designed to estimate the likelihood of committing type I errors; … and (e) an exclusion comparison as described by Kroeker et al. (2013) was performed to assess individual contributions of the most significant outcomes.* (**16.3**)]

Item 17: Clarification of *post hoc* analyses

Estimated reporting quality across the literature: Very low

Among papers where it seemed that hypotheses were formed after data collection (n = 28), 14% of papers included a *post-hoc* acknowledgement.

Example: Chaplin-Kramer *et al*. (2011)

[*Representational bias was investigated in post hoc analyses.* (**17.1**)]

Item 18: Metadata, data, and code

Estimated reporting quality across the literature: Moderate

The majority of papers provided data that produced the results presented in the manuscript (77%). Less than half of papers included meta-data (44%), or additional data beyond what was presented in the results of the main text (39%). Few studies provided code to reproduce the analyses (11%).

Example: Kinlock *et al*. (2018)

[*Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.rg5rd.* (**18.1, 18.2, and 18.3**)]

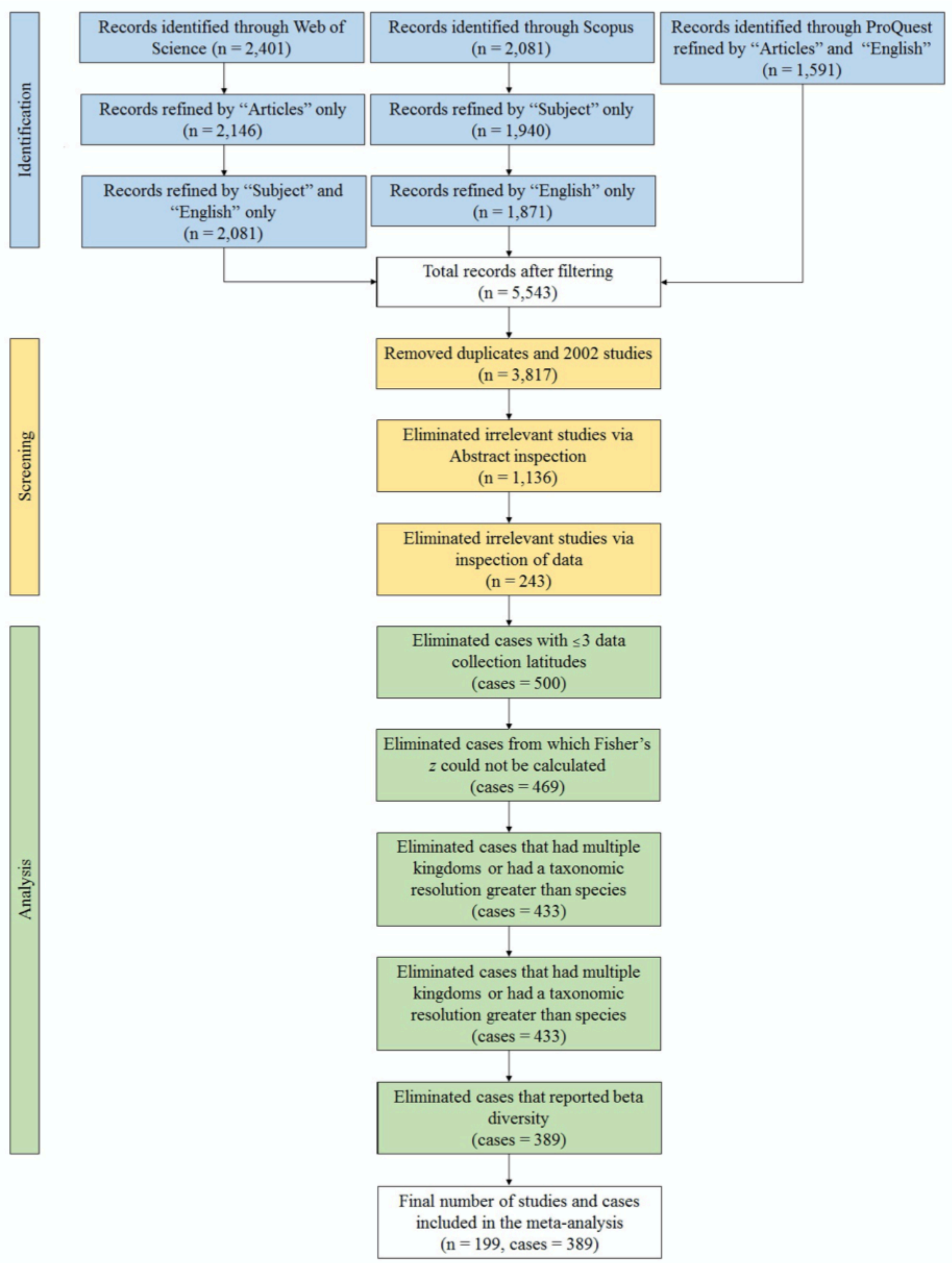
[*All R code used in this study can be accessed at https://github.com/ nlkinlock/LDGmeta-analysis.* (**18.4**)]

Item 19: Results of study selection process

Estimated reporting quality across the literature: Low

Less than half of papers reported the number of studies that were screened for inclusion (37%), and fewer reported the number of studies excluded (22%) and reasons for their exclusion (27%). In total, 19% of papers included a PRISMA-style flowchart.

Example: Kinlock *et al*. (2018)



[Appendix S2 *Figure S2.1* | *Prisma diagram. Explanation of the literature search and the subsequent elimination process of studies and cases used in the meta-analysis*. (**19.1, 19.2, 19.3, 19.4**)]

Item 20: Sample sizes and study characteristics

Estimated reporting quality across the literature: Moderate

Nearly every paper reported sample sizes for the meta-analysis (number of included studies and effect sizes; 96%), but just over half reported sample sizes for the subsets of data included in subgroup analyses or meta-regressions (57%). A minority of papers summarised the limitations of included moderator variables (22%). Among the small subset of papers that assessed individual study quality (risk of bias; $n = 5$), a majority provided a summary of this assessment (60%).

Example: Des Roches *et al*. (2018)

[*Our results summarize data from 25 different studies, which together focus on 12 genera at 3 trophic levels…* (**20.1**)]

[*Fig. 2 | Hedges' g values (± 95% cis) grouped by focal species (polygons) and studies for which intraspecific effects are larger (g < −0.3), similar to (−0.3 < g < 0.3) or smaller than (g > 0.3) species effects.* (**20.2**)]

[*Fig. 1 | Summary of studies used in the meta-analysis and the standardized intraspecific and species ecological effects. a, Focal species grouped by trophic level and showing the number of independent experiments. Focal species were placed in the lowest trophic level that they occupy. b, Standardized (mean and standard deviation) intraspecific versus species effects for each study with colour representing the trophic level, point size representing the number of responses measured and fill representing direct (filled) and indirect (unfilled) ecological effects. Points falling in the top left half of the graph represent larger intraspecific effects,* *while points falling on the bottom right represent larger species effects. c, Intraspecific versus species effects for all response variables with colour representing the trophic level, fill representing the direct (filled) and indirect (unfilled) ecological effects and shapes representing whether the response is a change in abundance (circle), process (triangle) or composition (square). Box plots show the interquartile range of the data.*

...

Table 1 | characteristics of studies (by author) included in our meta-analysis, including focal species, species ('replacement' versus 'removal') and intraspecific treatments, and categorization of all response variables measured as 'direct' or 'indirect' and as 'abundance', 'rate' or 'composition'. (**20.3**)]

[*…given that all removal studies used consumer species and most replacement studies used producer species, these two moderators were largely confounded.* (**20.4**)]

Example: Ord *et al*. (2011)

[*Table S3. Predictors of species recognition based on analyses in which data was excluded from those studies that suffered methodological problems* (**20.5**)]

Item 21: Meta-analysis

Estimated reporting quality across the literature: Very high

Nearly all assessed papers provided the overall meta-analytic mean with its corresponding confidence interval (94%).

Example: Jiang *et al*. (2013)

[*The mean value of r is 0.28 with a 95% confidence interval of 0.25–0.31, based on a random effects model with no fixed effects and species-trait means as the unit of replication.* (**21.1**)]

Item 22: Heterogeneity

Estimated reporting quality across the literature: Moderate

Around half of assessed papers reported an indicator of heterogeneity (52%).

Example: Thomsen *et al*. (2018)

Supplement:

[*Heterogeneity*

tau^2	*Q(d.f.=205)*	*Het. P*	*I^2*
0.887	*846.830*	*<0.001*	*81.541*

(**22.1**)]

Item 23: Meta-regression

Estimated reporting quality across the literature: High

Most papers provided estimates and confidence intervals for meta-regressions (78%) and outcomes from model selection (where applicable) (81%), and a smaller majority showed complete reporting for all moderators and interactions (59%).

Example: Moatt *et al*. (2016)

[*Fig. 3 Forest plots showing effect sizes (Cohen's d, standardised mean difference in reproduction between the control and restricted groups (see Methods and Additional file 1: Dialog S1)) of key moderators for the effect of dietary restriction (DR) on reproduction. Each point represents the Cohen's d value with the 95% credible intervals (CIs).* (**23.1**, **23.2**)]

[*Restriction:Model, represents the interaction between degree of restriction (%) and model or non-model species.* (**23.3**)]

[*Although the model including phylogenetic signal was a better fit by AIC score (phylogenetic AIC=577.33, non-phylogenetic = 579.86), the improvement was small and was not true for the model including all moderators (see below).* (**23.4**)]

Item 24: Outcomes of publication bias and sensitivity analyses

Estimated reporting quality across the literature: Moderate

Over half of assessed papers provided results for an assessment of publication bias (60%), and just under half gave results for other assessments of the robustness of the results (44%).

Example: Kornder *et al*. (2018)

[*Our overall effect sizes were robust to both publication bias and significant contributions of individual studies; that is, exclusion of the most significant studies did not change any of the results. Further, meta regression revealed no significant trend between response ratios and year of publication, suggesting that the datasets were not influenced by changes in methodology or assumptions (Appendix S1: Table S3).* (**24.1, 24.2**)]

Item 25: Discussion

Estimated reporting quality across the literature: High

Nearly every discussion touched on biological or practical relevance of the effect (98%), most papers discussed previous reviews, when they were available (88%), and most considered the magnitude of the effect (73%) and limitations of the review (72%). Over half of papers discussed the precision of effects (57%), and less than half of assessed papers summarised the main results in terms of their heterogeneity (47%).

Example: Wood *et al*. (2016)

[*Moreover, our results did not support that the magnitude, direction or form of selection systematically differs in small and large natural populations: it appears that natural populations of varying sizes experience a variety of environmental conditions, without consistently differing habitat quality at small population size.* (**25.1**)]

[*CI of the magnitude of linear coefficients overlapped for all N bins, N had no effect on directional selection, and N had little effect on the form of quadratic selection acting on different trait classes and taxa. While the direction of linear selection decreased weakly with increasing N across different trait classes, the effect was only significant for life-history traits and the effect size of the relationship was small.* (**25.2**)]

[*Our meta-analysis found little evidence for consistent, directional differences in* h^2 *or the extent of natural selection across a wide gradient of N in nature.* (**25.3**)]

[*If these results are not exceptional, response to selection at small N might be more extensive than previously assumed in evolutionary and conservation biology. Collectively, species* *conservation initiatives and priority setting should consider that (i) the evolutionary trajectories of some small populations appear to be very much affected by natural selection; (ii) different small and large populations of the same species may contain variation that is adaptive in a wide range of circumstances; and (iii) minimum viable population sizes for some species – genetically, strictly speaking – may not need to be as high as previously discussed…* (**25.4**)]

[*these are notable results given our very large databases and the general lack of research investigating patterns of selection in relation to population size in wild species.…long-term fluctuating spatial and temporal environmental conditions may have resulted in complex fluctuating selective pressures leading to a lack of relationship between* h^2 *and N in our meta-analysis ... Another possibility is that for some traits, phenotypic plasticity or fitness trade-offs might help to buffer loss of VA at small N (Rollinson and Rowe 2015; Wood and Fraser 2015).* (**25.5**)]

[*… data are heavily biased towards estimates of* h^2 *for morphology … responses to selection will likely depend on the nature of the selection pressures acting on particular traits, but information was not available for traits in the* h^2 *analysis to explore this. … N and not Ne was used as the population size metric in our analyses as Ne was rarely estimated on the same populations as selection, but it is Ne that dictates rates of genetic drift and inbreeding. ... multiyear N data were available for only a proportion of meta-analysis populations, so we cannot confirm long-term N for most populations. Fluctuating N could affect our conclusions by altering the relationship between* h^2 *and N…. more studies are required which measure selection and N for very large populations … there might be a systematic bias in the types of populations/taxa chosen for selection or* h^2 *studies.* (**25.6**)]

Item 26: Contributions and funding

Estimated reporting quality across the literature: High

Names, affiliations, funding sources and contact details were well reported (92%), but a contributions statement was given in only 31% of papers. For most assessed papers authors declared no conflict of interest or we did not have reason to believe there one was (92%). In the 8% of cases where it was ambiguous whether a conflict of interest was present, none was declared. All papers provided contact details for a corresponding author.

Example: Dunn *et al.* (2018)

[*Ruth E. Dunn[1], Craig R. White[2] and Jonathan A. Green[1]*
1School of Environmental Sciences, University of Liverpool, Liverpool L69 3GP, UK
2Centre for Geometric Biology, School of Biological Sciences, Monash University, Melbourne, Victoria 3800, Australia Funding. R.E.D. is supported by a NERC PhD studentship. C.R.W. is supported by the Australian Research Council (project nos FT130101493, DP180103925). (**26.1**)]

[*Authors' contributions. R.E.D. participated in the study design, data collection, analyses and writing of the manuscript; J.A.G. conceived the study, participated in data analysis and helped draft the manuscript; C.R.W. participated in the design of the study and provided essential guidance with the analyses. All authors contributed critically to the drafts, gave final approval for publication and agree to be held accountable for the content herein.* (**26.2**)]

[*Author for correspondence:*
Ruth E. Dunn

e-mail: ruth.dunn@liverpool.ac.uk (**26.3**)]

[*Competing interests. We declare we have no competing interests.* (**26.4**)]

Item 27: References

Estimated reporting quality across the literature: High

Nearly every assessed paper provided a list of all studies included in the meta-analysis (92%), but only 18% cited these studies in the reference list.

Example: Li *et al.* (2010)

[Page 6-9: *References to works included in the database* (**27.1, 27.2**)]

Reference list for examples

BATEMAN, D.C. & BISHOP, M.J. (2017). The environmental context and traits of habitat-forming bivalves influence the magnitude of their ecosystem engineering. *Marine Ecology Progress Series* **563**, 95–110.

BREED, M.F., OTTEWELL, K.M., GARDNER, M.G., MARKLUND, M.H.K., DORMONTT, E.E. & LOWE, A.J. (2015). Mating patterns and pollinator mobility are critical traits in forest fragmentation genetics. *Heredity* **115**, 108–114.

CHAPLIN-KRAMER, R., O'ROURKE, M.E., BLITZER, E.J. & KREMEN, C. (2011). A meta-analysis of crop pest and natural enemy response to landscape complexity. *Ecology letters* **14**, 922–932.

DAVIDSON, K.E., FOWLER, M.S., SKOV, M.W., DOERR, S.H., BEAUMONT, N. & GRIFFIN, J.N. (2017). Livestock grazing alters multiple ecosystem properties and services in salt marshes: a meta-analysis. *Journal Of Applied Ecology* **54**, 1395–1405.

De Boeck, H.J., Bloor, J.M.G., Kreyling, J., Ransijn, J.C.G., Nijs, I., Jentsch, A. & Zeiter, M. (2018). Patterns and drivers of biodiversity-stability relationships under climate extremes. *Journal Of Ecology* **106**, 890–902.

Des Roches, S., Post, D.M., Turley, N.E., Bailey, J.K., Hendry, A.P., Kinnison, M.T., Schweitzer, J.A. & Palkovacs, E.P. (2018). The ecological importance of intraspecific variation. *Nature Ecology & Evolution* **2**, 57–64.

Dunn, R.E., White, C.R. & Green, J.A. (2018). A model to estimate seabird field metabolic rates. *Biology Letters* **14**, 20180190.

Gorné, L.D. & Díaz, S. (2017). A novel meta-analytical approach to improve systematic review of rates and patterns of microevolution. *Ecology and evolution* **7**, 5821–5832.

Jiang, Y., Bolnick, D.I. & Kirkpatrick, M. (2013). Assortative mating in animals. *The American Naturalist* **181**, E125–E138.

Kettenring, K.M. & Adams, C.R. (2011). Lessons learned from invasive plant control experiments: a systematic review and meta-analysis: invasive plant control experiments. *Journal of Applied Ecology* **48**, 970–979.

Kinlock, N.L., Prowant, L., Herstoff, E.M., Foley, C.M., Akin-Fajiye, M., Bender, N., Umarani, M., Ryu, H.Y., Şen, B. & Gurevitch, J. (2018). Explaining global variation in the latitudinal diversity gradient: Meta-analysis confirms known patterns and uncovers new ones. *Global Ecology And Biogeography* **27**, 125–141.

Kornder, N.A., Riegl, B.M. & Figueiredo, J. (2018). Thresholds and drivers of coral calcification responses to climate change. *Global change biology* **24**, 5084–5095.

Li, F.-R., Peng, S.-L., Chen, B.-M. & Hou, Y.-P. (2010). A meta-analysis of the responses of woody and herbaceous plants to elevated ultraviolet-B radiation. *Acta Oecologica* **36**, 1–9.

Merkling, T., Nakagawa, S., Lagisz, M. & Schwanz, L.E. (2018). Maternal testosterone and offspring sex-ratio in birds and mammals: A meta-analysis. *Evolutionary Biology* **45**, 96–104.

Moatt, J.P., Nakagawa, S., Lagisz, M. & Walling, C.A. (2016). The effect of dietary restriction on reproduction: a meta-analytic perspective. *BMC evolutionary biology* **16**, 401.

O'Dea, R.E., Lagisz, M., Hendry, A.P. & Nakagawa, S. (2019). Developmental temperature affects phenotypic means and variability: A meta-analysis of fish data. *Fish and Fisheries* **12**, 523–18.

Ord, T.J., King, L. & Young, A.R. (2011). Contrasting theory with the empirical data of species recognition. *Evolution* **65**, 2572–2591.

Pearson, D.E., Ortega, Y.K., Runyon, J.B. & Butler, J.L. (2016). Secondary invasion: The bane of weed management. *Biological Conservation* **197**, 8–17.

Thomsen, M.S., Altieri, A.H., Angelini, C., Bishop, M.J., Gribben, P.E., Lear, G., He, Q., Schiel, D.R., Silliman, B.R., South, P.M., Watson, D.M., Wernberg, T. & Zotz, G. (2018). Secondary foundation species enhance biodiversity. *Nature Ecology & Evolution* **2**, 634–639.

Uller, T., Nakagawa, S. & English, S. (2013). Weak evidence for anticipatory parental effects in plants and animals. *Journal of Evolutionary Biology* **26**, 2161–2170.

Wood, J.L.A., Yates, M.C. & Fraser, D.J. (2016). Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications* **9**, 640–657.

Estimated and Perceived Reporting Quality of Meta-Analyses in Ecology and Evolution

Survey Methods

Estimating Reporting Quality

To estimate current reporting quality, we sought a representative sample of meta-analyses published in ecology or evolutionary biology journals, which we then used to assess reporting quality. Below we provide detailed descriptions of how the representative sample was selected, and the methods used to assess reporting quality reliably.

Protocol for estimating reporting quality

The study protocol for our assessment of reporting quality is available at https://osf.io/8zkxt (click through to the 'files' page). We registered the study protocol prior to abstract screening (deviations from this protocol are described below, under subheadings beginning with 'Deviation from protocol').

Design of search strategy

The search strategy was influenced by previous papers that sought representative samples of meta-analyses in ecology and/or evolutionary biology (Fraser *et al*., 2018; Koricheva & Gurevitch, 2014; Nakagawa & Santos, 2012; Senior *et al*., 2016). After developing a search string (detailed below), we decided upon the minimum number of papers we wished to include in the representative sample (n = 100), and then used pilot screening to estimate the number of titles and abstracts that would be required to meet this target. We performed two rounds of pilot abstract and full-text screening (screening 40 full-text papers in total). Just under half of screened papers met our inclusion criteria, so we aimed to screen approximately *n* = 300 studies, to ensure enough full-texts would meet our inclusion criteria.

Database search

List of journals

We exported a list of all ‘Ecology’ and ‘Evolutionary Biology’ journals and their ISSN (International Standard Serial Number), classified according to the 2017 rankings published by the ISI InCites Journal Citation Reports. Journals that appeared in both the ecology and evolution lists were classified as ‘Both’. We combined the lists, removed duplicates, and searched for studies published in these journals by including journal ISSNs in the search strings. We also manually searched for each journal’s ISSN to account for missing or erroneous information from the ISI export. Where a journal had both a print and online ISSN (i.e. eISSN), we included both in the search.

Finding studies

On 25 March 2019 we searched the *Scopus* database with the following search string: TITLE-ABS-KEY (“meta-analy*” OR “metaanaly*” OR “meta-regression”) AND PUBYEAR AFT 2009 AND ISSN(“13652486” OR “14610248” OR “01678809” OR “16000706” OR “14668238” OR “00063207” OR “14321939” OR “14712954” OR “15231739” OR “20457758” OR “13652664” OR “09218009” OR “13652745” OR “16161599” OR “1365294X” OR “14209101” OR “19395582” OR “13652435” OR “17264189” OR “2041210X” OR “15585646” OR “13652656” OR “21508925” OR “15375323” OR “17517370” OR “16000587” OR “13652699” OR “1744957X” OR “14657279” OR “15731464” OR “15577015” OR “22120416” OR “14350629” OR “15738477” OR “13652427” OR “14712148” OR “15729710” OR “09258574” OR “03043800” OR “14724642” OR “17524571” OR “16541103” OR “1526100X” OR

“10958312” OR “15749541” OR “23519894” OR “15735052” OR “01695347” OR “14391791” OR “14320762” OR “21619549” OR “16171381” OR “0022541X” OR “15729761” OR “14338319” OR “14429993” OR “09123814” OR “17083087” OR “17545048” OR “01401963” OR “01692046” OR “15371719” OR “2397334X” OR “14726785” OR “15733017” OR “15409309” OR “1432184X” OR “10968644” OR “1654109X” OR “16161564” OR “19360592” OR “13652540” OR “17550998” OR “00314056” OR “1438390X” OR “15507424” OR “00652504” OR “17596653” OR “10959513” OR “14691795” OR “15891623” OR “20511434” OR “11769343” OR “19342845” OR “22132244” OR “22244662” OR “00220981” OR “14657333” OR “19413300” OR “14697831” OR “01106465” OR “13653008” OR “24501395” OR “15731642” OR “10353712” OR “0003455X” OR “15452069” OR “17447429” OR “11645563” OR “03672530” OR “00472484” OR “17529921” OR “08000395” OR “16083334” OR “1076836X” OR “19385455” OR “13652028” OR “17279380” OR “20412851” OR “10301887” OR “02757540” OR “14230445” OR “10960031” OR “15882756” OR “22145753” OR “15735133” OR “14390574” OR “15220613” OR “19339747” OR “13504509” OR “15731561” OR “14321432” OR “08858608” OR “21996881” OR “14753057” OR “10369872” OR “23524855” OR “02497395” OR “15482324” OR “19968175” OR “18741746” OR “10960325” OR “13991183” OR “19436246” OR “18094392” OR “1146609X” OR “24108200” OR “00030031” OR “13488570” OR “15735125” OR “18185487” OR “03051978” OR “00030090” OR “0079032X” OR “1065657X” OR “19954263” OR “1432041X” OR “2073106X” OR “16423593” OR “1476945X” OR “14427001” OR “19399170” OR “11956860” OR “20419139” OR “1525142X” OR “21553874” OR “1876312X” OR “03781844” OR “14455226” OR “17513758” OR “00220930” OR “15525015” OR “1944687X” OR “02705060” OR “14645262” OR “14772019” OR “14390469” OR “1860188X” OR

“17451019” OR “20513933” OR “00280712” OR “19385307” OR “0029344X” OR “16181077” OR “19385331” OR “0340269X” OR “14421984” OR “14322056” OR “18739652” OR “01380338” OR “19385293” OR “18397263” OR “07176317” OR “19385412” OR “00384909” OR “13653113” OR “05643295” OR “02408759” OR “15270904” OR “09096396” OR “14636409” OR “13541013” OR “1461023X” OR “00301299” OR “1466822X” OR “477525” OR “00298549” OR “09628452” OR “08888892” OR “00218901” OR “00220477” OR “01718630” OR “09621083” OR “1010061X” OR “10510761” OR “02698463” OR “17264170” OR “00143820” OR “00218790” OR “00030147” OR “17517362” OR “09067590” OR “03050270” OR “17449561” OR “10452249” OR “13873547” OR “2818182” OR “14329840” OR “02697653” OR “00465070” OR “09603115” OR “13669516” OR “11009233” OR “10612971” OR “00244066” OR “13850237” OR “03405443” OR “09212973” OR “14429985” OR “07374038” OR “09639292” OR “15409295” OR “00953628” OR “2769667” OR “14022001” OR “09483055” OR “19360584” OR “0018067X” OR “1755098X” OR “14383896” OR “10557903” OR “13679430” OR “00713260” OR “15659801” OR “00221503” OR “00224561” OR “02664674” OR “00306053” OR “15052249” OR “10838155” OR “1543592X” OR “623257” OR “10674136” OR “10635157” OR “00840173” OR “01416707” OR “10220119” OR “09377409” OR “07483007” OR “15858553” OR “22145745” OR “03781909” OR “16124642” OR “00980331” OR “00222844” OR “03782697” OR “00322474” OR “03636445” OR “00400262” OR “18741738” OR “00405809” OR “21933081” OR “02775212” OR “00445967” OR “24107220” OR “09187960” OR “13862588” OR “17986540” OR “19954255” OR “0949944X” OR “2833888” OR “1520541X” OR “19344392” OR “1399560X” OR “15525007” OR “00222933” OR “09475745” OR “18601871” OR “17451000” OR “10926194” OR

"14396092" OR "00948373" OR "0913557X" OR "07224060" OR "00973157" OR "0370047X" OR "0716078X" OR "15287092" OR "03076970" OR "03003256")

The search string above was designed to identify studies that mentioned meta-analysis or meta-regression within the title, abstract, or key words, and were published since the beginning of 2010 in our list of ecology and evolution journals. Performing this search on 25 March 2019 yielded 1,668 papers from 134 journals.

We exported all citation, abstract, and serial identifier information from the 1,668 search results as a .csv file. After importing this information into *R* (v. 3.5.1) (R Core Team, 2018) we corrected small formatting differences in journal names, and matched each journal to the subject classification from our list of journals (either 'Ecology', 'Evolution', or 'Both').

Reducing the number of journals and studies

Our aim was to find meta-analyses on questions in the primary fields of ecology or evolutionary biology, and to this end we manually excluded some journals in more applied sub-fields (e.g. ecological economics). We arranged the remaining journals in descending order of the number of times the journal was returned in the search results, and then took the top 10 journals from ecology, the top 10 journals from evolutionary biology, and the top 11 from both (because the 10th and 11th journals had the same number of hits). This left us with the 31 journals that had published the most meta-analyses in their fields (Table S1).

Table S1.

Journals publishing the majority of meta-analyses in ecology and evolutionary biology. Journals classified as ‘Both’ are listed under both ‘Ecology’ and ‘Evolutionary Biology’ in the ISI InCites Journal Citation Reports. ‘N hits’ shows the number of studies returned from the search of *Scopus* (described above), out of the 1,668 total hits. ‘N screened’ shows the number of studies taken as a random sample, for a total of 300 abstracts to screen. ‘5-year IF’ is the five-year impact factor of the journal (this metric was not available for ‘Nature Ecology & Evolution’, as the journal was less than five years old at the time of the search).

ISI classification	**Full journal title**	**5-year IF**	**N hits**	**N screened**
Both	*Proceedings of the Royal Society of London B: Biological Sciences*	5.611	47	17
Both	*Ecology and Evolution*	2.788	43	17
Both	*Molecular Ecology*	6.885	35	17
Both	*Journal of Evolutionary Biology*	2.946	32	17
Both	*Evolution*	4.268	25	17
Both	*American Naturalist*	4.33	20	17
Both	*Biology Letters*	3.556	17	17
Both	*Evolutionary Ecology*	2.223	14	14
Both	*Nature Ecology & Evolution*	Not Available	7	7
Both	*Heredity*	3.92	5	5
Both	*Molecular Ecology Resources*	6.073	5	5
Ecology	*Global Change Biology*	9.791	135	9
Ecology	*Ecology Letters*	11.775	98	9
Ecology	*Oikos*	3.728	65	9
Ecology	*Global Ecology and Biogeography*	7.315	51	9
Ecology	*Biological Conservation*	4.995	49	9
Ecology	*Oecologia*	3.409	47	9
Ecology	*Conservation Biology*	5.755	43	9
Ecology	*Journal of Applied Ecology*	6.16	42	9
Ecology	*Journal of Ecology*	6.525	35	9
Ecology	*Marine Ecology Progress Series*	2.682	35	9
Evolution	*BMC Evolutionary Biology*	3.628	12	12
Evolution	*Evolutionary Applications*	5.063	10	10
Evolution	*Biological Journal of the Linnean Society*	2.322	9	9
Evolution	*Molecular Biology and Evolution*	14.479	7	7
Evolution	*Genome Biology and Evolution*	4.171	4	4
Evolution	*Molecular Phylogenetics and Evolution*	4.294	4	4
Evolution	*Evolutionary Bioinformatics*	1.71	3	3

Evolution	*Evolutionary Biology*	2.335	3	3
Evolution	*Journal of Heredity*	2.475	3	3
Evolution	*Systematic Biology*	14.501	2	2

Taking a sample of studies

To select approximately 300 studies to undergo screening, we took a random sample of up to 17 studies from the journals classified as 'Evolutionary Biology' or 'Both', and up to 9 studies from journals classified as 'Ecology' (because ecology journals contained more meta-analyses). This procedure provided us with 297 studies to undergo screening, of which 90 were classified as Ecology, 57 as Evolutionary Biology, and 150 as 'Both'.

Files for processing Scopus exports

The folder "Systematic Search" at https://osf.io/2xpfg/files/ contains the original .csv export from *Scopus*, the code used to process this file, and the sample of studies exported for abstract screening.

Abstract screening

Two contributors (R.E.O. and M.L.) independently screened all titles and abstracts to select studies for full-text screening (i.e. parallel abstract screening), using Rayyan software and following the decision tree in Fig. S1. This decision tree was first tested on two rounds of pilot screening (each round contained 20 randomly selected studies). Abstracts were excluded when they indicated that a study was unlikely to present results from a meta-analysis that had been conducted on data from multiple studies, in the fields of ecology or evolutionary biology. Meta-analyses based on existing data sets (e.g. results from a long-term experiment) were excluded. Conflicting decisions were resolved by discussion between R.E.O. and M.L. Included abstracts were exported as a RefMan file and imported into a reference manager to download the full texts. In total, 64% of screened abstracts were included for full-text screening ($n = 189$).

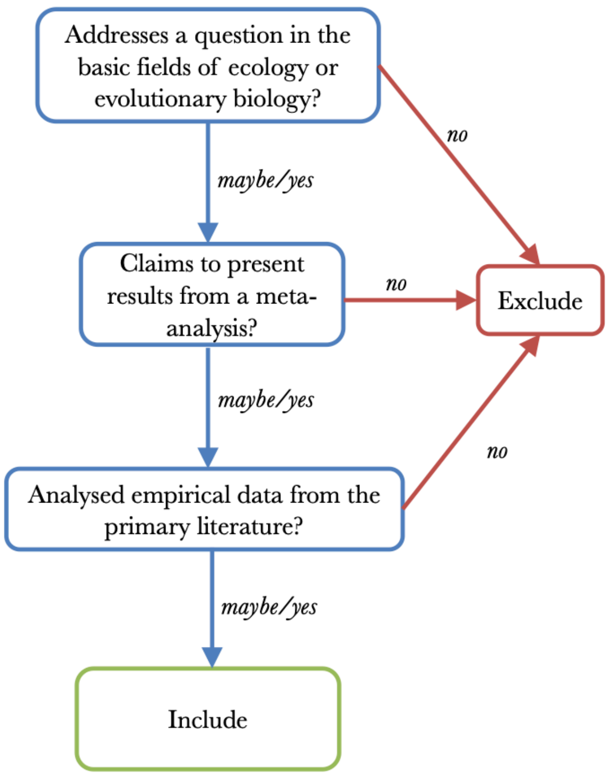


Figure S1.

Decision tree for abstract screening.

Inclusion and exclusion criteria

We assessed reporting standards of meta-analyses published in ecology and evolutionary biology journals that met the following criteria:

Addresses a question in the fields of ecology and evolutionary biology

In pilot screening, we found a number of studies that were predominantly in different fields (e.g. soil chemistry). Because the target audience for PRISMA-EcoEvo is researchers working primarily in the fields of ecology and evolutionary biology, we excluded, at the full-text stage, studies that did not clearly address an ecological or evolutionary question. Uncertainty over this distinction was resolved by discussion among three contributors (R.E.O., M.L., and S.N.).

Claims to present results from a meta-analysis

This criterion excluded studies that merely mentioned meta-analysis without conducting one, such as methods papers, narrative reviews, or commentaries.

Performed a search for data available from primary literature, and collected data from primary literature

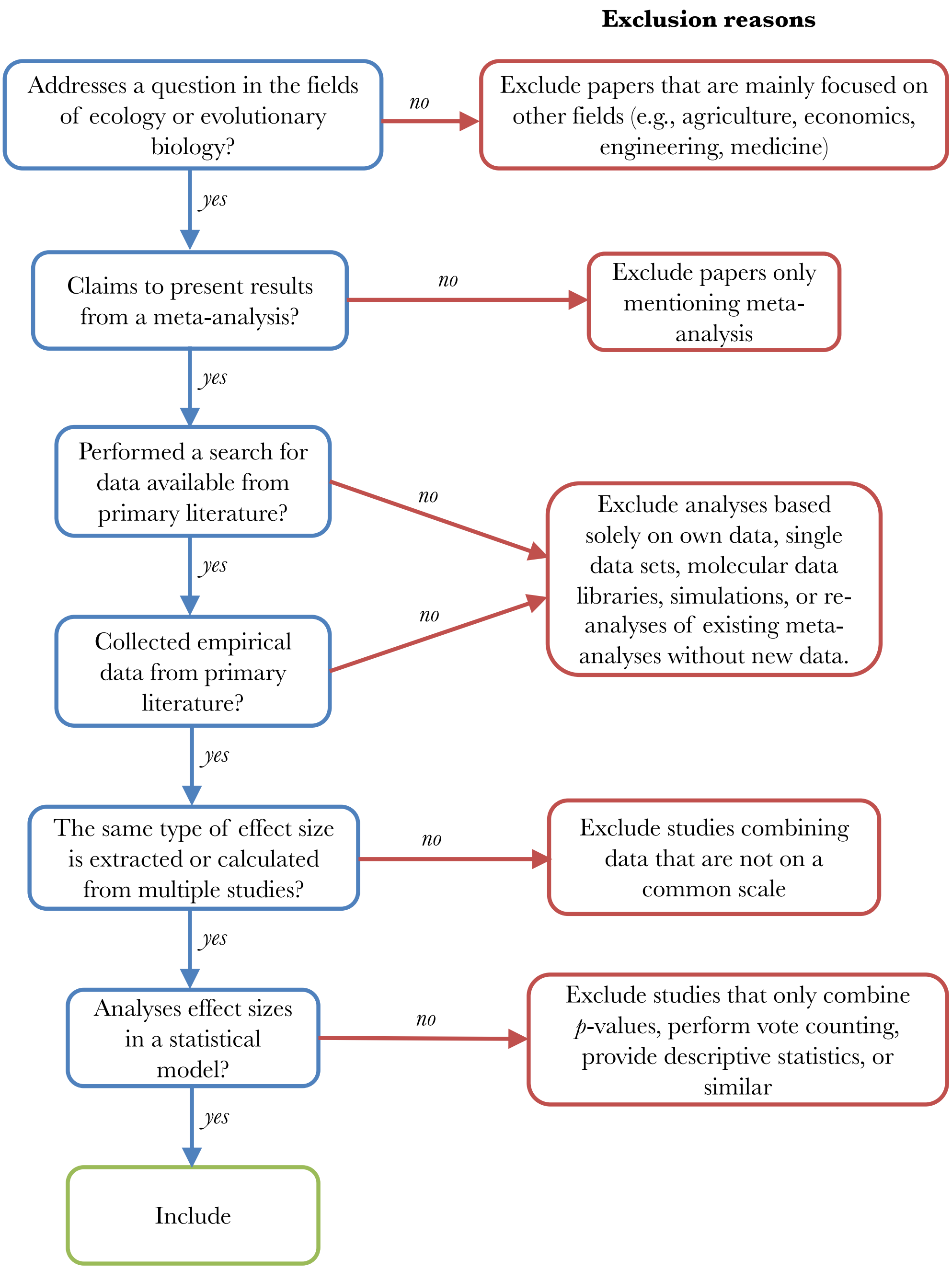
Because many of the items on PRISMA-EcoEvo relate to reporting standards of how literature was found and data collected, we wanted to ensure a sufficient sample size of meta-analyses to assess these items. Consequently, we excluded 44 studies that did not perform a literature search followed by data extraction (e.g. studies that analysed data sets that were downloaded without reference to the original studies, such as gene sequences or gene expression profiles).

The same type of effect size is extracted or calculated from multiple studies, and effect sizes are analysed in a statistical model

The included meta-analyses combined results from multiple primary studies using at least one effect size metric. Included studies could use more than one effect size on different subsets of the data or to address different questions.

Full text screening

Each full-text Portable Document Format (PDF) was given a unique identifier (full-text ID), and assigned to two reviewers (R.E.O. and M.L.) for independent screening. The full-text IDs were randomly assigned by first shuffling the order of the PDF documents using the sample function in *R*, before re-naming each file with a sequential ID. A *Google* Form (containing the decision tree shown in Fig. S2) was used to record the screening decisions, exclusion reasons, whether or not included papers contained supplementary information, and (for papers containing a supplement) the URL of the paper (to expedite downloading of supplements for included papers). Both the decision tree and *Google* Form were tested with two rounds of pilot screening (on 12 and 10 full-texts, respectively). Agreement rate after full-text screening was 80%. Conflicted decisions were discussed between M.L. and R.E.O., after which 17 decisions (9%) were still unresolved; the funding contributor, S.N., made decisions in these cases. After full-text screening, 125 papers were assigned as 'include', although some were marked as 'prefer not to be used' (described below). The sample sizes at different stages of the screening process are shown in Fig. S3.



1178

1179 **Figure S2.**

1180 Decision tree for full-text screening.

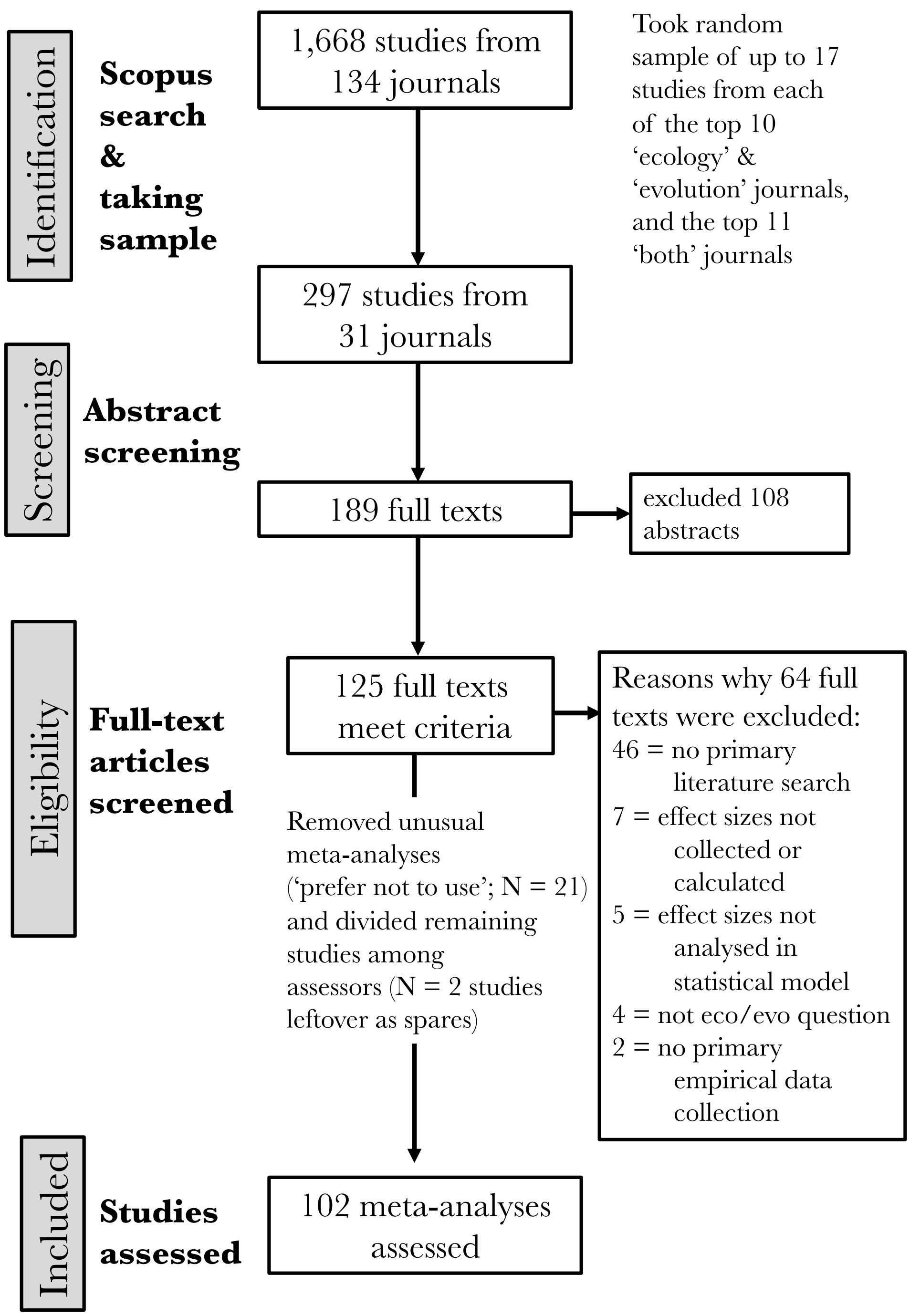


Figure S3.

PRISMA-style flowchart showing the process of collecting and screening studies, to obtain a representative sample of meta-analyses recently published in journals from the fields of ecology and evolutionary biology.

Deviation from protocol: papers assigned 'prefer not to use'

During full-text screening we assigned some included papers the label 'prefer not to use', to indicate papers that might technically meet our inclusion criteria but would be difficult to compare to the majority of the data set (e.g. use of very unusual effect sizes, uncertainty over whether the field was too applied). In total, 21 out of 125 papers received these labels. As we were aiming to screen at least 100 meta-analyses, we elected not to include the papers labelled as 'prefer not to use'. The remaining papers were evenly divided between assessors, so that two papers were 'left over' and not assessed (see Fig. S3).

Assigning papers for assessment

The lead contributor (R.E.O.) divided the included full texts between seven contributors participating in assessment (D.W.A.N., J.K., M.D.J., M.L., R.E.O., S.N., T.H.P.; hereafter known as 'assessors'). First, R.E.O. recorded when an assessor was a co-author on one of the included papers, so that they could be excluded from assessing their own paper. Papers were then evenly divided between assessors, roughly aligned to the assessor's fields of interest or areas of expertise. Each assessor was assigned 14 papers to be assessed individually, with an additional 4 papers to be assessed by multiple people (as part of data extraction training – see below). In total, therefore, 102 papers were assigned for data extraction. From these papers we extracted information necessary to assess the reporting standards of meta-analyses in ecology and evolutionary biology over the past nine years, as well as the current use of reporting checklists and guidelines.

List of assessed publications

References for the 102 assessed publications are provided in the main text reference list.

Training papers

To reduce subjectivity and clarify interpretation differences among the assessors, each assessor underwent training. First, four papers were assessed in person by a small group (R.E.O., M.L., and S.N.). During this time, ambiguity and interpretations were discussed and the assessment form was amended. From these four papers, three were chosen as 'training' material for the other assessors (the three papers were chosen as exemplars of 'easy', 'medium', and 'hard').

For each training paper that an assessor completed, the lead assessor (R.E.O.) created a document comparing the answers provided by the assessor with the 'consensus answers', including notes on the reasons for differences. In instances where the assessor had noticed something that the consensus team missed, the consensus answer was updated.

Deviation from protocol: changes made to Google *Form after training*

After assessor training, R.E.O. identified the questions with the lowest 'agreement rate'. Additional instructions were then added to the *Google* Form to clarify interpretation of these questions, including the addition of 'caution' images to highlight tricky questions.

Data extraction *Google* Form

We created a custom *Google* Form that could be used to assess each paper, to increase efficiency and minimise errors during data extraction. The *Google* Form stored answers for each paper in one row of a .csv file, which was then processed using a customised *R* script. The form contained instructions and examples to help increase consensus across different reviewers. The *Google* Form and *R* script underwent two rounds of piloting with four contributors before data extraction began.

Paper-level information

For each paper we recorded:

- Whether or not a registration existed, whether that registration was public, and the URL for any public registration
- Whether or not the paper presented a PRISMA-style flow diagram and, if so, a link to a screenshot of the figure stored in *Google Drive*
- Whether or not the paper referenced PRISMA, or any reporting guideline (if so, which one), and the wording of this reference
- Where PRISMA was referenced (methods, results, or SI) and how it was referenced (as a reporting guideline, methodological guideline, or flow-chart template; multiple answers could be selected)
- Where the list of all studies included in the meta-analysis was reported, if at all (main reference list, a table in the main text, or the supplementary information)
- Which tests of publication bias were presented in the paper from the following list of options:
    - Funnel plots (including contour-enhanced funnel plots)
    - Normal quantile (QQ) plots (Wang & Bushman, 1998)
    - Correlation-based tests (e.g. rank correlation; Begg & Mazumdar, 1994)
    - Regression-based tests (e.g. Egger *et al*., 1997, and its variants)
    - File drawer numbers or fail-safe N (methods described in Rosenthal, 1979; Orwin, 1983; and Rosenberg, 2005)
    - Trim-and-fill tests (Duval & Tweedie, 2000*a,b*)
    - P-curve, P-uniform or its variants (reviewed in McShane *et al*., 2016; van Aert *et al*., 2016)

- Selection (method) models (e.g. Copas & Li, 1997; Hedges, 1984; Iyengar & Greenhouse, 1988)
- Time-lag bias tests (e.g. regression or correlation on the relationship between effect size and time or cumulative meta-analysis)
- Other (specify)

Item-level information

Each of the 27 PRISMA-EcoEvo items was given their own section within the form, and subdivided into one or more components. For each item, the form asked:

Q questions

These compulsory questions were labelled Q01, …, Q27. They recorded all the components that a paper (or the supplementary information) reported. Three asterisks (***) at the end of the component indicated that the component may potentially be not applicable.

NA questions

The NA questions (NA01, …, NA27) recorded reasons for some components being ‘not applicable’.

SI questions

The SI questions (SI01, ... , SI27) recorded when a component was met solely because of the existence of the supplementary information. Assessors were asked to select which components would not have been selected in the Q questions had the SI not existed. Data and code were *not* considered part of the SI in most circumstances (the only exception being when data were hosted as supplementary information to the online version of the article).

Notes

The notes questions (notes_item01, ... , notes_item27) were used to record information that could be useful if we needed to re-evaluate the paper, or to record when a paper contained a particularly useful example of satisfying an item.

Processing of *Google* Form answers

Custom *R* functions were written to process the .csv file that was exported from the *Google* Form into component-level, item-level, and paper-level information. The script used a file ("Google Form Items.xlsx") containing the weighting of components within an item, and details of not-applicable items. First, the script extracted all paper-level information for each paper. Next, the script recorded whether a component was recorded in the supplementary information, and assessed the component following this basic algorithm:

- Was the component of the item met? yes/no
- Was the component of the item applicable? yes/no
- If the component of the item was not met, and if it was not applicable, then it received a penalty (yes), otherwise it received no penalty (no)

If the component received a penalty, then:

- Was every item of the component weighted equally (according to the weighting information found in "Google Form Items.xlsx")? yes/no
    - ➢ If yes, then the deduction was equal to the inverse of the number of applicable items multiplied by 100 (so that if every applicable item were not met, the penalty equalled 100%).
    - ➢ If no, then:
        - Were some components not applicable?
            - ➢ If no, then the deduction was equal to the inverse of the pre-specified weight

- If yes, then the pre-specified weights were updated so that the weight of the not-applicable item was absorbed by the applicable item with the highest pre-specified weight

Based on the assessment of each component, the following columns were generated:

- A ‘score’ for each component that was equal to 100 minus any deduction that the component received
- Whether or not the whole item was met [either yes (if 100), partial (if > 0 and < 100), no (if 0) or not applicable (if none of the components of the item were applicable)]

Data extraction spreadsheet

The results of the *Google* Form were stored in a database as two related sheets: one for the paper-level information, and one for the component-level information.

The meta-data for these sheets are shown in Tables S2 and S3.

Table S2.

Paper-level meta-data for the variables we extracted from each paper that was assessed.

Column name	Description
unique_extraction	unique ID for the *Google* Form (combination of paper_ID and who_extracted)
paper_ID	ID for the paper that data were extracted from
paper_title	title of the paper that data were extracted from, as a back-up in case there is a typo in the paper_ID
timestamp	date and time the *Google* Form was submitted
who_extracted	reviewer who completed the data extraction
general_paper_comments	any general comments that were made about the paper
online_data_exists	whether or not data for the paper are available online
data_link	link to online data
registration_exists	whether or not a registration for the paper exists
registration_accessible	if a registration exists, whether or not it is readily available
registration_link	link to the online registration
flow_diagram_exists	whether or not the paper presents a PRISMA-style flow diagram
flow_link	link to a screenshot of the PRISMA-style flow diagram on *Google* Drive (password protected)
references_provided_in	where the paper reports a list of all studies included in the meta-analysis, if at all (main reference list, a table in the main text, or the supplementary information)

paper_references_PRISMA	whether the paper references PRISMA
prisma_referenced_where	where PRISMA is referenced (methods, results, or SI)
prisma_referenced_how	how PRISMA is referenced (as a reporting guideline, methodological guideline, or flow-chart template; multiple options are allowed)
prisma_referenced_wording	wording of the PRISMA reference, copy-pasted
other_guidelines_referenced	wording of any other guideline referenced, copy-pasted

Table S3.

Component-level meta-data for the variables we extracted for each item and each paper that was assessed.

Column name	Description
row_ID	unique ID for the row number of the data
unique_extraction	ID for the *Google* Form (combination of paper_ID and who_extracted)
paper_ID	ID for the paper that data were extracted from
deduction	percentage to subtract from 100 if the component was penalised
equal_weight	whether or not the components in the item are weighted equally
item_met	whether or not the component was met (yes or no)
item_notes	any general notes for the item
item_number	item number
item_question	the component of the item, in the form of a question
not_applicable_allowed	whether or not the component could potentially be not applicable
not_applicable_explanation	the reason why the component could be not applicable
not_applicable_met	whether or not the component was not applicable
penalise	whether or not the component was penalised for not being met (yes, no, or not applicable)
reported_in_si	if the component was reported, whether this was due to the existence of supplementary information (yes; otherwise, blank)
score	the percentage score of the overall item
unequal_split	if the components of the item were not weighted equally, this shows the weight each item received (otherwise, blank)
weight	how the item was weighted
whole_item_met	whether or not the whole item was met (yes, partial, no, or not applicable)

Data checks

During data extraction, assessors were encouraged to note if they were unsure of their answers. After submitting the form, assessors saved the link to edit the form in a shared spreadsheet. During data processing, R.E.O. read these comments, and marked ones that required checking. In addition, items were flagged for checking if they met any of the following criteria: (1) the paper was recorded as reporting an item that was reported in < 10% of all papers; (2) a record was made of a publication bias test, but the item ‘assessments of the risk of bias due to missing results (e.g. publication, time-lag, and taxonomic biases)’ was not selected; (3) ‘none’ was selected in addition to another component (because if ‘none’ were selected, then no other component should have been selected). Each assessor received a record of the items they should check, along with the link to edit the submitted answers. Assessors either modified their answers or confirmed that the selected answers were correct. After each assessor had finished this task, the results were re-exported and the final results were processed.

Measured outcomes

Using the processed data, we estimated the following outcomes: (1) for each item, the percentage of papers that met the whole item, partially met the item, did not meet the item, or where the item was not applicable; (2) the frequency with which components of each item were met; (3) the distribution of average scores across the assessed papers; (4) whether the subset of papers that referenced a reporting guideline showed higher standards of reporting than those which did not; (5) which components of items were reported in the supplementary information, and how frequently this occurred; (6) how frequently PRISMA, or other reporting guidelines/checklists, were referenced, and the way in which they were referenced; (7) how frequently PRISMA-style flowcharts were reported, and what these flowcharts looked like; (8) how frequently data, meta-data, and code were made publicly available; and (9) whether there was a correlation between reporting standards of published papers and the impact factor of the journals they were published in.

Surveying the meta-analysis community

Following the assessment of reporting standards, we conducted a survey of the ecology and evolution meta-analysis community to see whether their perception of reporting quality was congruent with our results.

Identifying survey participants

We aimed for a minimum of 100 responses to the survey. We sought responses from members of editorial boards of ecology and evolution journals (i.e. the gatekeepers), and authors of meta-analyses that have been published in these journals after 2009 (i.e. the users). First, we collected the details of editorial board members of ecology and evolution journals that were represented among the 102 papers included in the reporting quality assessment. The included journals were: *American Naturalist*, *Biological Conservation*, *Biological Journal of the Linnean Society*, *Biology Letters*, *BMC Evolutionary Biology*, *Conservation Biology*, *Ecology and Evolution*, *Ecology Letters*, *Evolutionary Applications*, *Evolutionary Biology*, *Evolutionary Ecology*, *Global Change Biology*, *Global Ecology and Biogeography*, *Heredity*, *Journal of Applied Ecology*, *Journal of Ecology*, *Journal of Evolutionary Biology*, *Marine Ecology Progress Series*, *Molecular Ecology*, *Nature Ecology & Evolution*, *Oecologia*, and *Oikos*. On 11 July 2019 we obtained the list of editorial board members, and their affiliations, from each journal's website, and made a list of the email addresses of these editors (either from the website directly, or from *Google* searches; addresses that could not be found were ignored). In total we obtained 1,393 unique email addresses.

Second, we collected the email addresses of corresponding authors from the *Scopus* database on 5 August 2019, using the search string "TITLE-ABS-KEY ("meta-analy*" OR "metaanaly*" OR "meta-regression") AND PUBYEAR AFT 2009 AND ISSN("0003-0147" OR "477525" OR "0024-4066" OR "1744-9561" OR "1471-2148" OR "0888-8892" OR

"2045-7758" OR "1461-023X" OR "0014-3820" OR "1752-4571" OR "0071-3260" OR "0269-7653" OR "1354-1013" OR "1466-822X" OR "0018-067X" OR "0021-8901" OR "0022-0477" OR "1010-061X" OR "0171-8630" OR "0962-1083" OR "2397-334X" OR "0029-8549" OR "0030-1299" OR "0962-8452")". This search yielded 854 results. We then exported a .csv file containing the 'Correspondence address' for each of the papers, imported this file into the software program *R*, and then extracted email addresses (where available). After removing duplicates, we were left with 660 email addresses.

We combined the email addresses from both types of searches (editors and authors), removed duplicates, and removed authors of the PRISMA-EcoEvo project. This left a total of 2,017 prospective participants.

Recruiting survey participants

The funding contributor (S.N.) emailed participants through his university email addresses, inviting them to complete an anonymous online survey. The email specified the aim of the survey and why the potential participant was contacted. Potential participants were contacted through blind carbon copy (Bcc) so that they could not access the emails of other invitees. Two invitations were sent within one month. The first email stated that a reminder email would be sent within 3–4 weeks, and gave participants an opportunity to withdraw their email from the mailing list to avoid receiving the second email.

As an incentive to participate in the survey, participants were given the opportunity to enter a random draw to win one of five copies of the textbook *Handbook of Meta-analysis in Ecology and Evolution* (Koricheva, Gurevitch & Mengersen, 2013) (valued at $152 AUD, with free shipping to their nominated address). To retain the anonymity of survey answers, at the end of the survey participants could enter their email address into a separate Google sheet (thus the email addresses were not linked to the survey answers). The five winners were randomly drawn from this list of email addresses (using the 'sample' function in the base package of *R*, with a pre-specified 'set.seed(336)'), and were notified on 21 October, 2019.

Survey content

We created an anonymous online survey as a *Google* Form. The survey was open to responses between 29 August 2019 and 21 October 2019, and collected the following information:

What experiences had survey respondents had with meta-analyses in ecology and evolution?

We asked survey participants for their career stage, prior experience with meta-analysis (as an author, reviewer, or editor), and whether they plan to author a meta-analysis in the future.

How does the community's perception of reporting standards compare to actual reporting standards?

Participants were asked to evaluate current levels of reporting, on a Likert scale of 1 to 5 (where 1 was 'very poor' and 5 was 'excellent'), in the following categories: (1) Systematic search; (2) Study selection; (3) Model description; (4) Data description; (5) Results description; (6) Bias assessment; (7) Availability of metadata, data, and code. We also asked for general comments about reporting standards. We kept this question succinct and therefore restricted the categories to only a subset of items that were covered by our assessment; the chosen categories were deemed particularly important for understanding how a systematic meta-analysis was conducted.

What is the community's current perception of reporting guidelines?

We asked survey participants whether they had heard of reporting guidelines prior to being invited to complete the survey, and whether they thought guidelines improve reporting standards.

How are reporting guidelines currently being used?

Survey participants with previous experience of reporting guidelines were asked whether they had used reporting guidelines: (1) as an author, to help conduct the study; (2) as an author, to help write the study; (3) as a reviewer/editor, to assess manuscripts; and (4) other (with explanation).

For questions 5–8, survey respondents were asked whether they wanted to view the PRISMA-EcoEvo checklist and answer a longer survey. If they answered yes, then the following additional questions were posed.

Opinions on pre-registration of meta-analyses (a PRISMA item)

Our assessment of reporting standards found that almost no meta-analyses in ecology and evolutionary biology were registered in advance. To help understand why this practice was so uncommon, we asked the respondents: (1) whether they had considered or tried pre-registration for any type of study; (2) why they thought pre-registration was uncommon; and (3) if they had general comments to make about pre-registration of meta-analyses.

Awareness and opinions on assessment of individual study quality in meta-analyses (a PRISMA item)

Compared to medical fields, it is extremely rare for meta-analyses in ecology and evolution to conduct and report an assessment of the quality of the evidence that is aggregated in the review (i.e. ‘critical appraisal’). We asked survey participants whether they thought this practice should be encouraged (‘Yes’, ‘No’, or ‘Don’t understand what this means’), and whether they had comments to make about this practice in general.

Will PRISMA-EcoEvo be useful?

To gauge the level of interest in PRISMA-EcoEvo, we asked survey participants if they would consider using PRISMA-EcoEvo in the future, and in what format it would be most useful to them.

General comments about PRISMA-EcoEvo

Respondents had the opportunity to make free-form, general comments on PRIMSA-EcoEvo.

Processing survey results

After the survey closed, the results were downloaded as a .csv file and processed using a customised *R* script.

Ethics approval

The survey methods were approved by the Human Research Ethics Committee of The University of New South Wales, Sydney. HC No: HC190648.

Results

Assessment results

1. For each item, what percentage of papers met the whole item, partially met the item, did not meet the item, or the item was not applicable?

The scores for each item are shown in the main text, Table 1, and the ranking of average item scores is shown in Fig. S4. The items with the lowest reporting quality (<50% average score) were, starting with the lowest: review registration (also called 'pre-registration'); assessment of individual study quality; study selection (i.e. reporting how studies were screened for inclusion); clarification of *post hoc* analyses; results of the study selection process (e.g. a PRISMA flow-diagram); data collection process; description of steps taken to deal with missing data during analysis; the availability of meta-data, data, and code; finding studies (i.e.

description of the systematic search); and descriptions of publication bias assessments and sensitivity analyses.

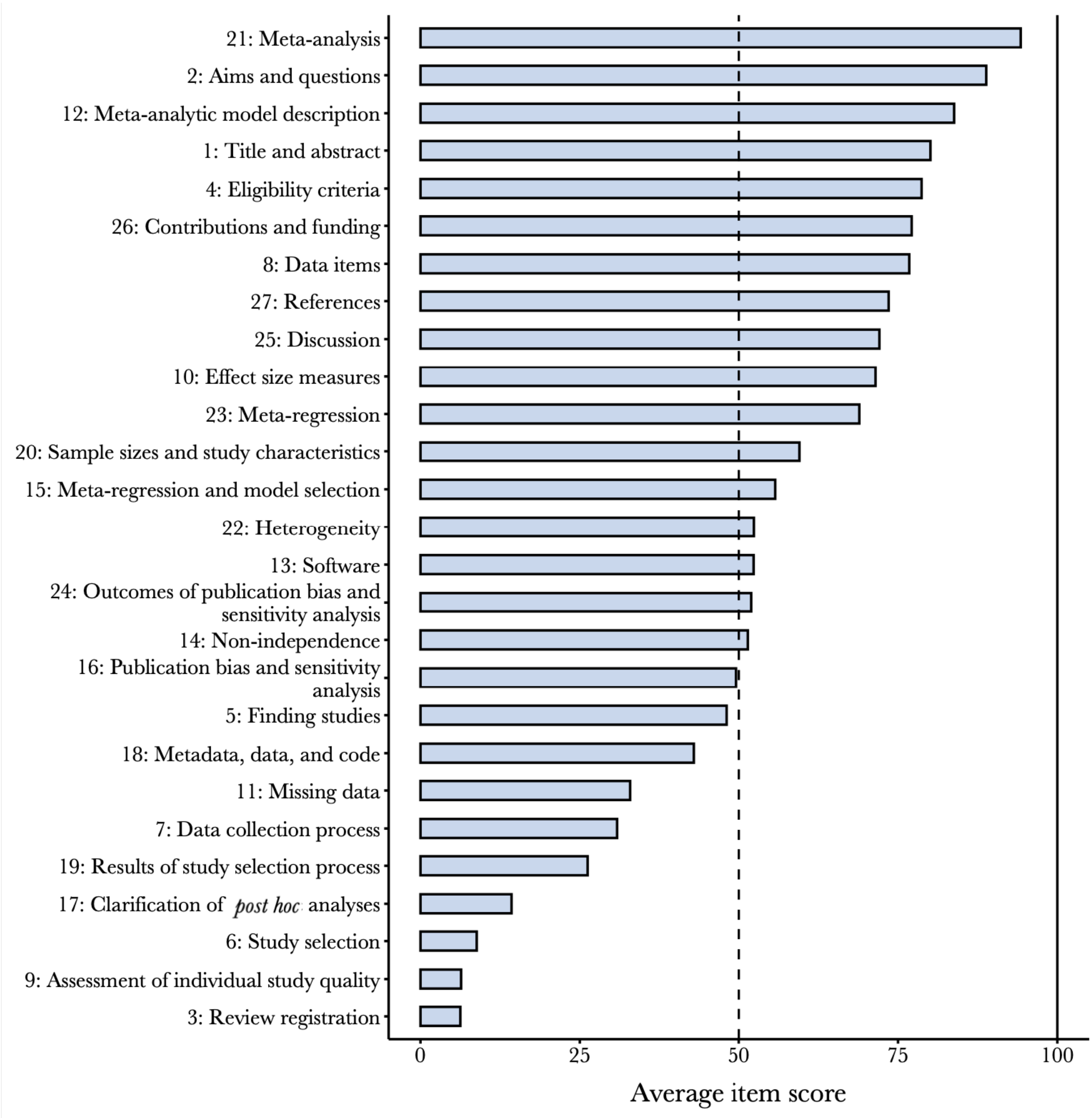


Figure S4.

The average score across the 27 reporting items included in PRISMA-EcoEvo. For each paper, items received a maximum score of 100. Those scores are then averaged across all applicable papers that were assessed (sample sizes shown in the main text, Table 1).

2. The frequency with which components of each item were met.

The percentage of papers meeting the component of each item is shown in Table 1 of the main text. In total, 89 reporting components were assessed, encompassing the full range of reporting (from 0 to 1). The average proportion across the components was 0.53 (SD = 0.32). Three components were never reported in those, albeit rare, instances when they were deemed applicable: disclosing conflicts of interest, acknowledging deviations from registered methods, and justifying deviations from registered methods. The other least-reported components (<10% of applicable papers) were, in increasing order of reporting frequency: the number of extractions that were checked for accuracy by a co-author (n = 1/102 reported); a registration (i.e. link to descriptions of study aims, hypotheses, and methods in a time-stamped and publicly accessible archive; n = 3/102 reported); the number of people involved in screening studies, and how they contributed (n = 3/102 reported); how information about study quality was incorporated into analyses (n = 6/102 reported); whether the quality of studies included in the meta-analyses were assessed (i.e. critical appraisal; n = 7/102 reported); the derivation of effect size and sampling variance equations, when none already existed (n = 2/28 reported); and who collected data (n = 10/102 reported).

The three components that were reported in every paper were in the title or abstract, administrative information, and introduction: (1) providing contact details for the corresponding author (unsurprisingly, as this is a journal requirement); (2) providing a rationale for the study; and (3) identifying the review in the title or abstract as a systematic review, meta-analysis, or both. It is important to note, however, that our systematic search required studies to mention some variation of meta-analysis or meta-regression in the title or abstract, so for this particular component we have an extremely unrepresentative sample. The other most-reported components (reported in greater than 95% of applicable papers) were reporting: sample sizes (number of independent groups, e.g. studies) for data included in meta-analyses (*n* = 87/91 reported); the results of the primary outcome (*n* = 98/102 reported); primary aims and questions (*n* = 98/102 reported); the key data sought for each study (*n* = 98/102 reported); the aims and scope of the review (*n*= 99/102 reported); effect size(s) used (*n* = 99/102 reported); the models used for synthesis of effect sizes (*n* = 99/102 reported); and a discussion of the main findings in terms of their biological/practical relevance (*n* = 100/102 reported).

3. The distribution of average scores across the assessed papers

The histogram of total scores is shown in the main text, Fig. 1. The average score was 55% (*n* = 102; SD = 10%; range = 33–77%).

4. The subset of papers that referenced a reporting guideline showed higher standards of reporting than those which did not

The small subset of papers that referenced a reporting guideline tended to have higher reporting scores than the majority of papers that did not reference a guideline (Guideline subset: mean = 65%; SD = 7%, range = 53–77%, *n* = 16; No guideline subset: mean = 54%, SD = 10%, range = 33–74%, *n* = 86; see Fig. 1 of main text). All papers that referenced a reporting guideline included supplementary information. Papers that referenced a guideline also reported a higher proportion of information in the supplementary information, compared to papers that did not (guideline: mean = 0.15; SD = 0.11; *n* = 16; no guideline: mean = 0.09; SD = 0.09; *n* = 86; Fig. S5).

5. Which components of items were reported in the supplementary information, and how frequently did this occur?

Components of 20/27 items on PRISMA-EcoEvo were reported in the supplementary information at least once, and 90/102 assessed papers made use of a supplement. Sub-items that relied on the supplementary information most heavily were components of: Item 27 (providing a reference list); Item 24 (providing results for the assessments of the risks of bias and the robustness of the review's results); Item 16 (subitem 16.3: describing other analyses of robustness of the results); Item 23 (subitem 23.1: providing estimates of meta-regression slopes and confidence/credible intervals); Item 8 (subitem 8.3: describing main assumptions or simplifications that were made during data extraction); Item 19 (presenting a PRISMA-style flowchart, alongside the number of studies excluded during each stage of screening); and Item 4 (eligibility criteria used during screening). The proportion of components reported in the supplementary information for each paper is shown in Fig. S5. There was a trend for the use of the supplementary information to slightly increase over time (linear regression: slope of z-scaled publication year = 0.02, t-value = 1.92; df = 100).

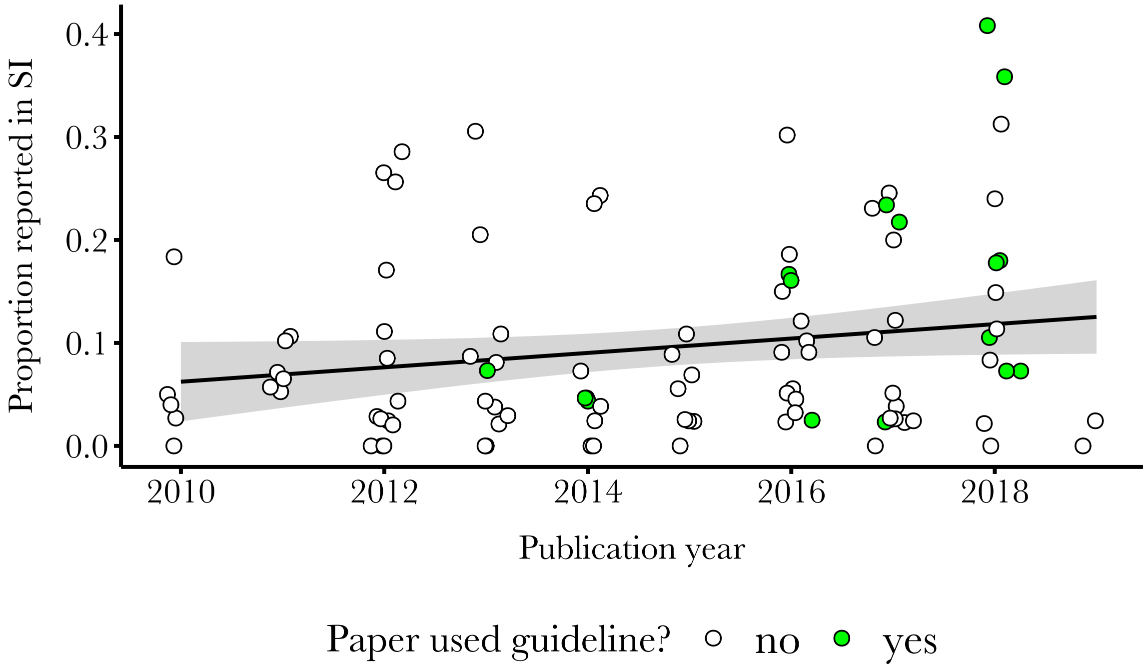


Figure S5.

The proportion of sub-items reported in the supplementary information (SI) across time, for the 102 assessed papers. Green points represent papers that reported using a reporting guideline and/or checklist. The year of publication is shown on the *x*-axis (with random error added to prevent overlapping datapoints), while the *y*-axis shows the proportion of components that were reported in the supplementary information rather than the main text (the components are shown in Table 1 of the main text).

6. How frequently were PRISMA, or other reporting guidelines/checklists, referenced, and in what way were they referenced?

86 papers mentioned no reporting guidelines/checklists, 15 papers referenced PRISMA, and one paper that did not reference PRIMSA referenced a different reporting guideline: Koricheva & Gurevitch (2014). PRISMA was most-often referenced as a flow-chart template (n = 14), followed by a methodological guideline (n = 6), and a reporting guideline (n = 2).

7. How frequently were PRISMA-style flowcharts reported, and what did they look like?

Figures for PRISMA-style flowcharts were reported in 18.6% of assessed papers (*n* = 19). Only one figure followed 'best practice' by splitting the number of identified sources into specific sources, listing sample sizes for different reasons for exclusion at the full-text stage, and listing the number of included studies and effect sizes. Nearly half of the figures used the flowchart template available from PRISMA (*n* = 9/19). The remaining figures varied in style and level of detail. *n* = 9 figures split the number of identified studies into specific sources (e.g. *Scopus* and *Web of Science*), *n* = 5 split identified studies into general sources (e.g. 'database searching' and 'other sources'), and *n* = 5 figures merely listed the number of identified sources. A minority of figures (*n* = 5) provided the number of papers excluded for specific reasons. It was most common for the total number of excluded papers to be listed without reasons (*n* = 13; e.g. one figure that followed the PRISMA flowchart template listed "Full-text articles excluded, with reasons (*n* = 12)", but did not list any reasons). One figure did not distinguish between papers excluded at early and late screening stages. For papers included in the meta-analysis, 5 figures included the number of effect sizes and included studies, while 14 figures only listed the number of included studies.

8. How frequently were data, meta-data, and code made publicly available?

Most assessed papers shared the data required to reproduce results presented in the manuscript (77% of studies). Data were not always accompanied by data descriptions (44% of studies shared metadata), and often excluded additional data beyond what was presented in the paper (39% of studies included additional data, such as the raw data used to calculate effect sizes). The code for analysis (or full descriptions for GUI software) was available for 11% of studies.

9. Was there a correlation between reporting standards of published papers and the impact factor of the journals they were published in?

As shown in Fig. S6, meta-analyses with relatively high standards of reporting were published in journals across the spectrum of impact factors, but the lowest standards of reporting were limited to the lower-impact factor journals. This pattern was reflected by a weak positive correlation ($r = 0.20$) between 5-year impact factor and reporting score, and a slight trend for reporting scores to increase by less than 1% with a one-point increase in the five-year impact factor (linear regression: slope of 5-year impact factor: 0.78, t-value = 2.00; df = 96).

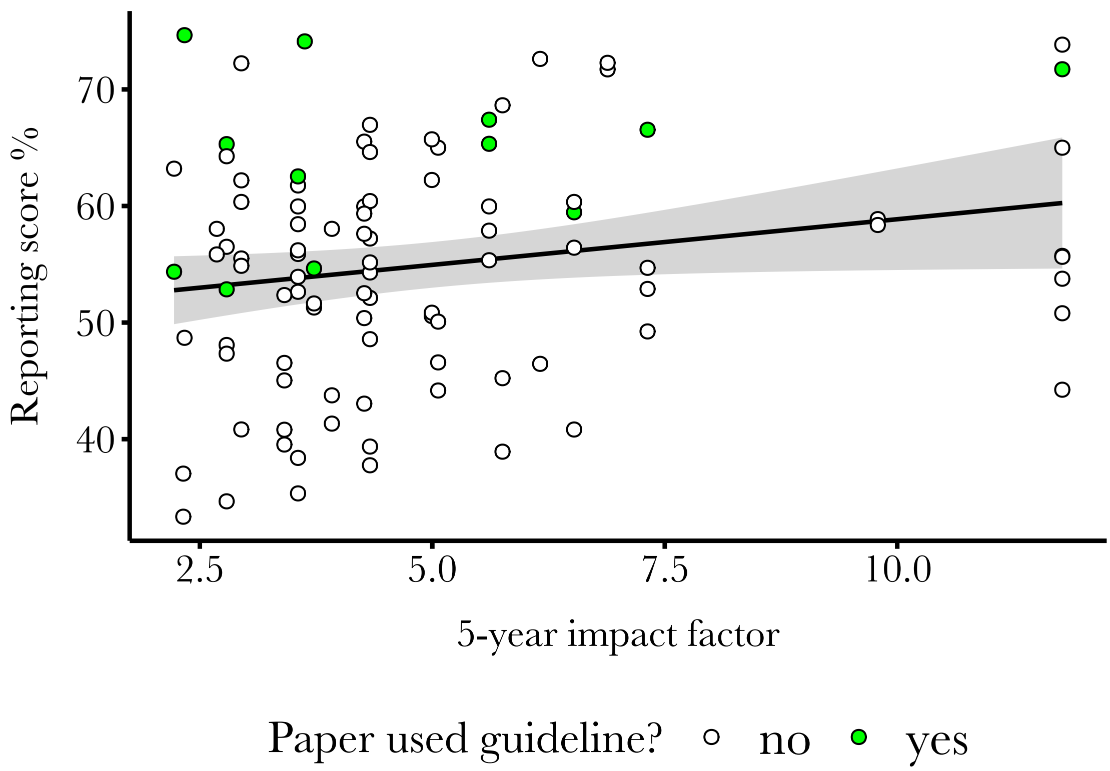


Figure S6.

The relationship between the impact factor of the journal the meta-analysis was published in, and its reporting score, for $n = 98$ meta-analyses (four papers were published in a recent journal that did not yet have a 5-year impact factor). Green points represent papers that reported using a reporting guideline and/or checklist.

Community results

In total, 208 people responded to the survey, of whom 5% were students, 44% were early-career researchers (up to 10 years since PhD), 30% were mid-career (up to 20 years since PhD), 20% were senior researchers (>20 years since PhD), and less than 1% were emeritus. 58% of survey participants identified solely as ecologists, while 36% listed both ecology and evolution as their research fields. Only 6% listed their sole field as evolution. Free-text comments are shown in Table S5.

10. What experiences had survey respondents had with meta-analyses in ecology and evolution?

Of the 208 people who responded to the survey, 87% had previous experience as a meta-analysis author, 55% had been a reviewer, and 29% had handled a meta-analysis as an editor. A smaller subset of respondents had experience with systematic reviews as an author, reviewer, or editor (49%, 32%, and 17% respectively; note that the definition of a 'systematic review' was at the discretion of the survey respondent).

First authors were the most common survey participants (67% have been first author of a meta-analysis, and 36% had been first author of a systematic review). Middle and senior authors were represented at similar rates (36% and 34% for meta-analyses; 19% and 20% for systematic reviews).

Among the smaller subset of responses from people who only identified with the research field of evolution (i.e. not ecology; n = 13), there was a smaller distinction between meta-analysis and systematic reviews; 62% of sole-evolution respondents identified as having authored both a meta-analysis and a systematic review.

11. How does the community's perception of reporting standards compare to actual reporting standards?

Survey respondents were asked to assess current standards of reporting for seven aspects of meta-analyses that were conducted following a systematic search, on a scale of 1 to 5 (where 1 was 'very poor' and 5 was 'excellent'). Descriptive statistics of responses are shown in Table S4, along with the number of extreme assessments (1 or 5). Note that the number of responses is smaller than the number of people who took the survey, because the questions were optional.

To compare this perceived quality from our estimated quality, we multiplied the mean score by 20 to give a percentage mean score (with 'very poor' = 20%, and 'excellent' = 100%), and compared these scores to the average reporting scores found in our earlier assessment (the relevant items are listed in Table S4). Fig. S7 shows the comparison of estimated and perceived scores. The largest discrepancy between perceived and estimated reporting quality was seen for study selection; the perceived reporting quality for study selection was reasonably high, while the actual reporting quality we estimated from the literature was very low. Estimated reporting qualities for the systematic search, bias assessment, and computational reproducibility were also markedly lower than the perception. The perceived reporting scores were not noticeably different between people who chose to access the full descriptions of items on PRISMA-EcoEvo by taking a longer survey, and those who took the shorter survey.

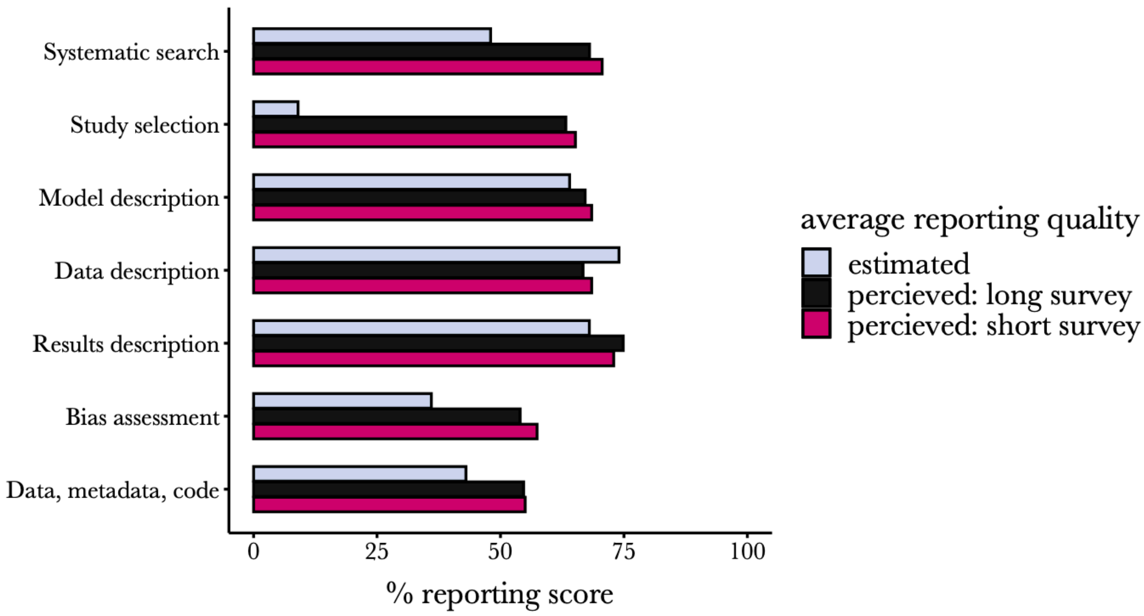


Figure S7.

A comparison of estimated reporting scores from the assessment of reporting quality ('Mean reporting score %' in Table S4; top grey bars) with the average of the community's perceived reporting scores. The *y*-axis labels reflect the survey questions (see column 5 in Table S4). Perceived reporting scores were calculated by multiplying the mean score (from 1 to 5, where 1 was 'very poor' and 5 was 'excellent') by 20 (sample sizes shown in Table S4). Black bars show the scores for those who took the longer survey (which gave the opportunity to read the items on PRISMA-EcoEvo; 57% of responses), whereas the red bars show the scores for those who chose to take the shorter survey (43% of responses).

Table S4.

Mean estimated reporting score (%) for relevant items from PRISMA-EcoEvo with the average score estimated by survey respondents. The survey asked for a rating of reporting quality on a scale of 1 to 5, with 1 being ‘Very poor’ and 5 being ‘Excellent’.

PRISMA-EcoEvo Reporting items	**Mean reporting score %**	**Survey field**	**N**	**Survey question**	**Mean score (1–5)**	**SD score**	**‘Very poor’ %**	**‘Excellent’ %**
Finding studies	48	Ecology	116	Systematic search (finding studies)	3.44	0.75	0	6
		Evol B.	11		3.64	0.67	0	9
		Both	70		3.46	0.83	1	6
		Total	197		3.46	0.77	1	6
Study selection	9	Ecology	115	Study selection (screening studies)	3.22	0.79	0	5
		Evol B.	11		3.27	0.65	0	0
		Both	70		3.17	0.85	1	3
		Total	196		3.20	0.80	1	4
Meta-analytic model description; Non-independence; Meta-regression and model selection	64	Ecology	116	Model description (statistical methods)	3.41	0.90	0	10
		Evol B.	11		3.18	0.87	0	0
		Both	70		3.37	0.84	0	7
		Total	197		3.39	0.87	0	9
Data items; Effect size measures	74	Ecology	116	Data description (what was collected and calculated)	3.41	0.90	0	9
		Evol B.	11		3.36	0.67	0	0
		Both	71		3.31	0.73	0	3
		Total	198		3.37	0.83	0	7
Meta-analysis; Heterogeneity; Meta-regression	68	Ecology	113	Results description	3.75	0.74	0	16
		Evol B.	11		3.64	0.81	0	9
		Both	71		3.63	0.76	0	10
		Total	195		3.70	0.75	0	13
Assessment of individual study quality; Publication bias and sensitivity analysis; Outcomes of publication bias & sensitivity analysis	36	Ecology	116	Bias assessment (e.g. publication bias, study quality)	2.72	0.99	6	6
		Evol B.	11		3.09	0.70	0	0
		Both	71		2.80	0.89	3	4
		Total	198		2.77	0.94	5	5
Metadata, data, code, and materials	43	Ecology	116	Availability of metadata, data, and code	2.72	0.99	5	7
		Evol B.	11		2.91	0.54	0	0
		Both	70		2.76	0.92	6	4
		Total	197		2.74	0.94	5	6

12. What is the community's current perception of reporting guidelines?

A total of 69% of survey participants had heard of reporting guidelines prior to completing the survey, while 62% had previous experience using a reporting guideline. People generally thought the use of a reporting guideline would improve the quality of reporting (87% thought they would improve reporting quality and only one person thought reporting guidelines would not improve reporting quality; the remaining responses were 'Don't know').

13. How are reporting guidelines currently being used?

Among the 129 people who had previously used a reporting guideline, 71% had used them 'as an author, to help conduct the study' (i.e. as a methodological guideline), 66% had used them 'as an author, to help write the study' (i.e. as a reporting guideline), and 36% had used them 'as a reviewer/editor, to assess manuscripts').

14. Opinions on pre-registration of meta-analyses (a PRISMA item)

A total of 116 people selected a reason why meta-analyses are rarely registered in advance. 'Lack of awareness' was the most common reason (66%), with 'Too time consuming' and 'Too few incentives' both being selected 48% of the time.

15. Awareness and opinions on assessment of individual study quality in meta-analyses (a PRISMA item)

When asked if 'critical appraisal' should be encouraged in ecology and evolution meta-analyses, 72% of the 116 respondents said 'Yes', and 17% said 'Don't understand what this means'.

16. Will PRISMA-EcoEvo be useful?

When asked 'Would you or your students use PRISMA-EcoEvo?', 69% of 113 respondents said 'Yes', 29% said 'Maybe', and 2% said 'No'. Among respondents who identified with both Ecology and Evolution (45 people), 76% said they would use PRISMA-EcoEvo. 54% of respondents selected a static table being a useful format, while an interactive online form was selected by 35% of respondents (10% listed 'Other' formats).

17. General comments about PRISMA-EcoEvo, registration, and critical appraisal

Seventy-six respondents provided general comments in the form of free text (longer than 5 characters in length), which are shown below in Table S5. To adhere to ethics approval requirements, we have redacted potential identifying comments (deletions and insertions are indicated with square brackets). Obvious spelling errors have also been corrected.

Comments on reporting quality and PRISMA-EcoEvo

There were 38 comments on reporting quality and/or PRISMA-EcoEvo. The comments regarding PRISMA-EcoEvo were generally positive, with the view that reporting quality is highly variable and would benefit from systematic guidance (e.g. s29: "Current reporting quality is very low. Many people doing meta-analyses don't seem aware of existing reporting guidelines"; s42: "The use of these reporting metrics is really bi-modal. Some papers do this extremely well while others fail to do this at all"; s52: "would be great if journals had published guidelines referred to on their website that one could access before submission"; s61: "Many studies seem still not to be using such a systematic approach. It clearly is a good idea"). One respondent was a notable exception, and expressed their opinion that reporting guidelines could worsen research quality rather than improve it (e.g. s16: "If you want better quality research, teach people to think. Making them follow detailed instructions has the opposite effect."). Multiple respondents expressed frustration with meta-analyses that do not cite the primary studies included in the analysis (i.e. sub-item 27.2; Table 1 in the main text), and make it difficult to obtain the data (e.g. s36: "I am frustrated that many authors claim data used in meta-analysis is "already published" and hence they are under no obligation to publish it"; s45: "Most meta-analyses do not include the reference in a format that allows the source paper to be properly cited (e.g. list of paper cited in electronic appendix). This is problematic.”). Two respondents noted practical differences between ecology and evolution and medicine (s26: “It is hard to meet PRISMA guidelines as a single author, since many require multiple reviewers of included manuscripts”; s27: “Decent, but likely far below medicine, for example.”). One respondent expressed the view that reporting quality was a bigger issue for empirical studies than meta-analyses, which was also raised as a barrier to assessing individual study quality (s43: “The main issue I have is not doing meta-analysis but the reporting standards of empirical studies that I tried to use in a meta-analysis”).

Comments on registration

Registration of meta-analyses received comments from 34 respondents, most of whom were sceptical. Some respondents were unfamiliar with registering meta-analysis methods in advance (e.g. s09: “Why one should do this?”; s17: “Honestly I was not aware that pre-registration (as explained in the PRISMA xlsx file) is a necessary step in meta-analysis”; s45: “I am not sure what it really means”), and others saw registration as a protection against being scooped by other researchers (s12: “Can pre-registration really stop other people from publishing a paper on my ideas before me, if they are working faster???”; s76: “useful to make sure there is no duplication”).

Multiple respondents were unconvinced by the benefits of registration (e.g. s36: “Not clear to me that it is needed”; s43: “I’m not sure it really helps”; s10: “I think the scope for p-hacking a meta-analysis by fine-tuning its methods is pretty minor”), or thought registration was better suited to empirical studies (e.g. s47: “I think it would be more difficult than for typical empirical studies”). Two respondents expressed general reservations about registration (s38: “In general, I do not agree with the notion of pre-registration as it interferes with the potential for adjusting to surprises during the study and to evolution of your ideas”; s52: “I’m not a huge fan of preregistration. Half the fun of any study is in exploring the data”).

Two barriers for registration identified in the comments were a lack of reward for the required effort (e.g. s08: “no incentives according to current personal evaluations”), and concerns over constraining scientific creativity (e.g. s58: “Needs to be done sensibly – i.e. allowing for new ideas & hypotheses as the study progresses”). It was identified that registration would be difficult for people with little experience (e.g. s51: “would be very challenging for researchers performing their first or second meta-analysis”; s60 “The process can be very iterative, meaning that you don’t quite know what you’re going to do until you see the papers and data that it is possible to acquire. For researchers with a lot of experience or expertise this might not be an issue, but researchers often do a meta-analysis when entering a new field because it’s a good way to get on top of the literature and identify questions. In these cases preregistration can be inhibitive.”).

There was tentative interest among some respondents for registration (e.g. s24: “if implemented in an ecology-specific way (as detailed in the PRISMA-EcoEvo document), this could work”; s70: “It might be a good idea. I will consider it in the future.”). Three people thought registration was already being conducted, without necessarily being made public (s25: “we have in our lab a detailed process to describe and register electronic documents describing each project”; s41: “I don’t understand how it could be done without pre-registration (though have not heard use of this term to describe outlining goals ahead of time)”; s59: “This place is typically filled by the time investment in research proposals?”). Notably, responses from the three people who claimed prior registration experience were all positive (s35: “it was a very useful exercise, and increased the transparency of the whole process”; s44: “I find it very convenient”; s53: “I found it very useful to have to think about my hypotheses, data collection and statistical analyses ahead of the study. It made the whole process much easier, even if it seemed like I 'lost' time at the start").

Comments on assessment of individual study quality

The topic of assessing individual study quality (i.e. critical appraisal) received the most feedback, with 51 free-text comments. The predominant response was that the quality of ecology and evolution studies is hard to assess, and subjective judgements could introduce, rather than protect from, biases (e.g. s20: "Hard. Quality can be many different things. … You should be very precise when defining quality and selection criteria"; s34: "In my experience individual study quality is very hard to assess in an objective way- some evidence-based guidelines would be very useful"; s37: "Studies on rare species or rare events often have a few data points, but the value of these studies should not be evaluated by statistical approach."). A smaller number of people thought bias assessments were already being carried out (e.g. s16: "The idea that you would not assess this is appalling"; s21: "I would expect it is done but not reported on in the methods"; s45: "Well I thought everyone was checking the quality of the study included in meta-analysis").

Table S5.

Free-text comments given in response to the survey questions “Do you have comments about reporting quality of meta-analyses or PRIMSA-EcoEvo?”, “Do you have general comments to make about pre-registration of meta-analyses?”, and “Comments about critical appraisal/assessment of individual study quality?”. Each row represents a different respondent to the survey. Only responses of length 6 characters or greater are shown. Some information has been redacted to retain anonymity.

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s01	–	–	It is obviously difficult to do this if there are many papers involved. It is difficult enough evaluating the quality of any one study.
s02	–	Most studies evolved as data are collected so preregistration is impractical	–
s03	error checking and taxonomic correction are seldom reported, but are critical steps.	–	I think some degree of filtering is important - but formally assessing every one of hundreds or thousands of source papers is prohibitive. For some variables (easy things like how tall is a plant), it is less important than for other more variably measured traits.
s04	relatively new in field, some unfamiliarity	I think you should not consider meta-analysis and systematic review separately. The former is a sub-case of the latter. E.g. below – there is no such thing as "systematic meta-analysis".	See comment above. Critical appraisal, in my view, is part of the selection criteria

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s05	I think the general reporting quality has room to improve	We need to remove barriers such as time-consuming processes for greater adoption of pre-registration	This creates room for people to subjectively choose which studies they want to include in their analysis and which they want to leave out.
s06	The items list in the title, abstract and introduction sections are not necessary, which could make meta-analysis papers to be too boring to read.	–	–
s07	Fulfilling all these requirements may sometimes be impossible, if data come from very heterogeneous sources and refer to very many taxa and/or studies	–	Critical appraisal is necessary, but always introduces some 'subjective' component. This is not 'bad' but authors as well as readers must be aware of that.
s08	–	no incentives according to current personal evaluations (no IF)	–
s09	–	Why one should do this?	–
s10	–	I think the scope for p-hacking a meta-analysis by fine-tuning its methods is pretty minor. The danger comes in things like how you extract the results from the literature: there are usually multiple results per paper you *could* pick, and some support your hoped-for result more than others.	Yes in principle, though in practice there are usually relatively few primarily studies to pick from, and most of them are somewhat flawed (e.g. good luck including only 'blind' experiments!)
s11	Reasons for search terms often poorly justified	What are the advantages?	Comparing results with study quality as a weighting factor would be interesting

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s12	–	Can pre-registration really stop other people from publishing a paper on my ideas before me, if they are working faster???	–
s13	–	I think this would be more urgent for experimental studies than meta-analyses, but in generally a good idea. However, a change in attitude is needed and scientists should appreciate exploratory studies better. These studies are essential and are currently often disguised as planned experiments. Development of pre-registration systems should be accompanied with good outlets for exploratory work.	It is a bit arbitrary of course (although more objective quality assessments exist; e.g. sample sizes, standard errors etc.), but generally a good idea. I think researchers are generally good in categorising studies into quality categories, particularly if this is done by multiple researchers independently.
s14	–	–	Study quality should (in ecology) always be factored into calculations, using the sample size and standard deviations of reported results. Further, studies should be screened at least briefly for major flaws in the approaches that could bias results.
s15	Many studies do not describe how they have dealt with missing error values associated to the study (e.g. SE, SD, CI) and that is an important caveat.	–	–

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s16	Listen, making a meta-analysis into a controlled series of steps that could practically be carried out by a computer is not the answer to improving research quality. Any intelligent person who has judgement and suitable knowledge will make appropriate choices about inclusion of data and possible biases. Sometimes, in fact, bias is just a negative way of saying expert judgement. For example, I may judge that due to the methods used in Experiment A, it is actually not comparable to Experiments B and C. This is because I have done similar research and I realize that there is a critical but subtle difference between them that in my judgement means that A is not actually measuring the same thing (this happens all the time in ecology). If I can't include such judgements in my process, then ipso facto all my results will be meaningless (in my judgement). If you want better quality research, teach people to think. Making them follow detailed instructions has the opposite effect. You will quickly erode people's ability to think critically about meta-analysis, as well as remove any basis for them to have a critical discussion about methods during peer review, as it will all be reduced to "did you follow the instructions or not?" Analysis, which is about how we think and make sense of things, can never be reduced to a lab procedure or a computer programme. I hate this direction that science is taking. I understand you are probably thinking about an issue like replicability of the method. Guess what, anything that involves (1) expert judgement (2) a huge subsampled data-set, can never be 100% replicable. What we need here are opinion pieces explaining the different logical basis for having an argument about what those non-replications mean. Lets make science smarter, not stupider.	The only reason to preregister is to prove you didn't "cheat" by having a new idea about what you should be doing halfway through. The stakes are also too low, frankly, for me to care whether other people initially thought they would include experiments of type X but then changed their mind. Fine. That's normal.	If you're just dumping results into the study without appraising if they make sense and actually mean what they say they mean, your results will be based on a bunch of incomparable things and will be shite. I always do this. Studies can be valid on their own and claim to address issue X but not, for your purposes, do so in the way you need them to. This VERY COMMON. The idea that you would not assess this is appalling.

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s17	–	Honestly I was not aware that pre-registration (as explained in the PRISMA xlsx file) is a necessary step in meta-analysis	I am not sure I understand what critical appraisal means here. If critical appraisal means to list data sources analysed in the meta-analysis and their main characteristics in order to determine study quality, then I would say that this practice needs to be encouraged because it helps the authors of the meta-analysis to have an idea of the quality of primary studies they are dealing with. I do this for my meta-analyses as well, but I do not normally report study quality in the paper or SI because I think it is not necessary for the reader.
s18	–	–	Although I see the merits, I can imagine it being subjective, and prohibitively time-consuming.
s19	–	–	It shouldn't be related to the impact factor of a journal
s20	–	–	Hard. Quality can be many different things. It can be confused with relevance or up-to-dateness. You should be very precise when defining quality and selection criteria.
s21	–	–	I would expect it is done but not reported on in the methods

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s22	–	–	It's difficult to standardise critical appraisal beyond appropriate sample sizes etc.
s23	–	–	I am not sure how to operationalize it in studies including hundreds of individual experiments.
s24	–	I think if implemented in an ecology-specific way (as detailed in the PRISMA-EcoEvo document), this could work. In a sort of related case, the number of studies with published data has greatly increased only in the last few years. This seems to me like a consequence of editors/journals requiring published data.	I don't know of any cases of quality assessment in ecological meta-analysis, but I could see it being relevant. Not sure exactly how one would define quality in ecology, this may be even more discipline specific than ecology (e.g., quality in a controlled experiment with plants may differ from other subfields).
s25	–	I had not considered it before though we have in our lab a detailed process to describe and register electronic documents describing each project	–
s26	It is hard to meet PRISMA guidelines as a single author, since many require multiple reviewers of included manuscripts.	–	–
s27	Decent, but likely far below medicine, for example.	–	Potential to introduce additional bias
s28	In my view, most papers today are not clear enough in respect with items number 3, 4, 6, and 18 of PRISMA-EcoEvo.	–	–

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s29	Current reporting quality is very low. Many people doing meta-analyses don’t seem aware of existing reporting guidelines.	–	In some fields this an be done in a reasonably straight forward way based on the type of study design. In ecology and evolution there is so much variation in the way studies are designed and implemented that this would need to be done on a case-by-base basis. The result would be very subjective. I suggest weighting by sample size rather than conducting quality assessments
s30	Cite properly the studies included in the MA (not merely in an appendix)	–	–
s31	–	I think it will fall to journals to require pre-registration if we as a field want to see this implemented on a large scale.	I'm ambivalent about 'critical appraisal' because this could introduce bias for the final selection of studies. Unless there is a set of standard objective criteria to follow, I don't know that I would be comfortable with 'critical appraisal' of studies.
s32	Many authors do not understand the models (e.g., fixed v. random effects) and/or results	I view this as analogous to experimental design and power considerations which is often neglected in research	Appraisal also must be viewed in the context of study bias and number of studies incorporated into the analysis
s33	–	–	But this will increase the time taken to do an analysis, needs to be balanced and appropriate
s34	–	–	In my experience individual study quality is very hard to assess in an objective way- some evidence-based guidelines would be very useful

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s35	Reporting quality should be improved as much as possible, for the syntheses to be valuable for future studies. Standardizing reporting will be extremely helpful.	I have used in for a systematic map protocol published in Environmental Evidence -- it was a very useful exercise, and increased the transparency of the whole process.	–
s36	Filling out this survey because I am frustrated that many authors claim data used in meta-analysis is "already published" and hence they are under no obligation to publish it. No requirement, even from top journals (Nature, Science) to do so. And yet, it's impossible to verify whether data have been correctly extracted without seeing the meta-analysis dataset. Currently fighting with an author of a paper in Science because his methods are insufficient for me to understand what data he has extracted from our papers. He has sent me the effect size he got but I have no idea how he got it. I can point to many other similar examples. I've been sent papers to review that don't even give any identification of the studies that they extracted data from.	Not clear to me that it is needed.	My field tests for methodological effects (eg differences between different experimental facilities, different pot sizes). Not sure what else could be done to assess study quality?
s37	–	–	It depends. Studies on rare species or rare events often have a few data points, but the value of these studies should not be evaluated by statistical approach.

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s38	Very thorough.	In general, I do not agree with the notion of pre-registration as it interferes with the potential for adjusting to surprises during the study and to evolution of your ideas. Rather, I would prefer a change in the way that we write papers so that we can describe our process of discovery, rather than writing a paper as if the outcome were pre-ordained. Such a change would be much truer to how science actually happens.	Critical appraisal can be very useful, as long as it is done so that bias can be minimized and assessed. A clear description of exclusion criteria and a citation for each excluded paper, along with the reason for exclusion, should be part of the reported supplementary material of the study.
s39	–	–	I do not know how to include this critical appraisal in the meta-analysis, I use to include some methodological moderators in the analysis to account for different types of studies however this is not strictly an assessment of quality.
s40	–	–	Unless the study is an outlier driving the results it seems unnecessary to single out individual papers

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s41	–	I don't understand how it could be done without pre-registration (though have not heard use of this term to describe outlining goals ahead of time)	If it was published and meets criteria of inclusion, I am not sure if it is the role of a M-A to analyse the study quality? Wouldn't/shouldn't meeting both publication standards and pre-identified inclusion standards for the M-A at hand already make the study quality meet a threshold of acceptability?
s42	The use of these reporting metrics is really bi-modal. Some papers do this extremely well while others fail to do this at all.	–	–
s43	The main issue I have is not doing meta-analysis but the reporting standards of empirical studies that I tried to use in a meta-analysis. Often key data like N or SD/SE are missing, and some times any data that can be used for doing a meta-analysis is absent. More rarely, data presented or conclusions inferred don't match the data given in figures or supplementary information	There is not much incentive to do so. If it's for curbing bias, I'm not sure it really helps given that if journals force authors to provide a PRISMA diagram/information prior to peer review, that will solve issues of bias. So what's the point of pre-registering?	As mentioned before, there is some really poor reporting practises for empirical studies in ecology/evolution. To the point, I'm always frustrated over the proportion of studies I look at closely when doing a meta-analysis about how badly statistical inference is being made and how frequently data is not available in its raw form or presented appropriately in figures/tables. I see this for papers published in top journals!
s44	–	I find it very convenient	–
s45	Most meta-analyses do not include the reference in a format that allows the source paper to be properly cited (e.g. list of paper cited in electronic appendix). This is problematic.	I am not sure what it really means	Well I thought everyone was checking the quality of the study included in meta-analysis...

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s46	In general, the quality is mid-low, but I think it is improving with time	–	–
s47	–	I think it would be more difficult than for typical empirical studies.	My answer above is more of "it depends." If the theory and methods for incorporating quality-based weights was more developed or accessible, then using "critical appraisal" would be fine. If it is not done in a systematic manner though, by definition critical appraisal does not seem appropriate for a systematic review.
s48	–	–	I think that Ecology has so diverse study designs, that I fear that a study quality assessment could be a subjective criteria to exclude potential useful studies (which could be better included at the analyses with a moderator coding its difference from the others).
s49	–	–	Sample size is often used. But study quality, not so much i suppose.
s50	–	–	This is a fairly vague phrase - would need a concrete definition.

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s51	–	This seems like a potentially large additional workload, and would be very challenging for researchers performing their first or second meta-analysis. In the meta-analyses I've performed, data quality was worse than I had hoped, and I had to adapt my questions somewhat based on available information. Eligibility criteria in particular evolved over time as we saw the range of issues with existing studies.	I have discussed this with colleagues, and a barrier to critical appraisal of individual studies is that it feels subjective and it isn't clear how to implement this in the meta-analysis - should I weight studies based on study quality? If so, how much? It can also be difficult to systematically compare study quality when reporting quality is so variable. The closest I've come to this is experimenting with weighting by sample size in my analysis.
s52	totally varies depending on authors and reviewers. would be great if journals had published guidelines referred to on their website that one could access before submission	I'm not a huge fan of preregistration. half the fun of any study is in exploring the data	It depends. Lots of meta-analyses have some standards for including a study, which usually include some basic minimum design and data standards. I think lots of meta-analyses also report some measure of study quality (e.g. this many studies assessed lifetime fitness vs other variables). In my experience applying standards that are too high reduces the sample size to useless, but I think authors should always discuss how the limitations of the literature affect their results

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s53	–	I found it very useful to have to think about my hypotheses, data collection and statistical analyses ahead of the study. It made the whole process much easier, even if it seemed like I 'lost' time at the starts	–
s54	highly variable	–	often “critical appraisal” of individual studies is subjective. Methods are designed to deal with spurious results objectively and statistically.
s55	I have not read that many meta-analytical paper to be able to rate the current quality responsibly	–	–
s56	In the field of cooperative breeding studies, there is very little attention paid to the quality of the data being reported or how it was collected.	–	–
s57	–	–	Difficult to base this on objective criteria
s58	–	Needs to be done sensibly - i.e. allowing for new ideas & hypotheses as the study progresses	–
s59	The fear and disincentives to report non-significant results is a difficult hurdle for the peer process.	This place is typically filled by the time investment in research proposals?	With some interest in the development of FAIR principles, it is evident that semantics imbedded in peer articles of many studies mislead readers - intentionally or not. This is a difficult phenomenon to control for.

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s60	–	The process can be very iterative, meaning that you don't quite know what you're going to do until you see the papers and data that it is possible to acquire. For researchers with a lot of experience or expertise this might not be an issue, but researchers often do a meta-analysis when entering a new field because it's a good way to get on top of the literature and identify questions. In these cases preregistration can be inhibitive.	I said no but actually I'm undecided. I can see pros and cons to both. If studies are done and reported properly in the first place it shouldn't be necessary. Any variation in quality will just add a bit of error. Also I feel we'd have to be careful how it's done because it could be very subjective and therefore introduce unwanted bias. But given we know that there is a lot of variation in the quality of studies then somehow incorporating this into methods and analyses could be helpful. It could also go a long way to improving awareness about what a good study is
s61	Many studies seem still not to be using such a systematic approach. It clearly is a good idea.	–	–
s62	It seems to me that the reporting quality is very variable. I have seen MA that even do not report on the number of studies screened, or even included in the analysis. On the other side of the spectrum, some MA have excellent reporting standards (still, a minority). And it seems that this varies between eco evo fields	–	I have done this for my first MA. I have used a three-levels of trust into the study, but it was challenging to construct these subjectively and consistently. I guess it would be extremely useful if Eco Evo would have some kind of a list of evaluation points to quickly determine the quality of individual studies (I would be very interested to participate in this)

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s63	I've seen a number of studies that are actually meta-analyses (but perhaps called systematic reviews or some other name), but never actually mention the word meta-analysis in the paper. As a result of this mislabelling, the editor perhaps does not recognize that the study at hand is a meta-analysis, the result being that these studies often do not include appropriate bias assessment methods perhaps because they are not sent to reviewers familiar with bias assessment methods. I think this could be remedied with familiarizing editors (and scientists in general) about what separates a meta-analysis from systematic reviews.	–	–
s64	It's a very good one.	–	–
s65	I feel it is very mixed		
s66	–	–	I found very difficult to calculate this in EcoEvo.
s67	For the models, it would be good to recommend that authors report the model structure explicitly. As written, your guidelines suggest to mention whether the model is a “random effects” or “hierarchical” model, which is good. But I see many meta-analysis that just say that and don't say what the model structure was. A better description is to state, in sentence or equation form, what the response variable was, what the data model was, what the predictors were, and whether there were interactions.	I don't think the actual scientific approach is orderly enough to produce pre-registration in many cases. There's art to science. Clarity in procedures and questions often takes years, particularly for grad students, at which point it's too late to “pre-register”. I don't mean that science proceeds in a biased way per se, just in a messier way than is reasonable for pre-registration.	–

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s68	–	–	I feel like most meta-analyses in our field now account for sample size or standard error in the actual analyses. If you mean that we should account for more than that, then I think I would agree and answer “yes”. But if not, I'm sure I agree with the premise.
s69	The quality of meta-analysis is staggered, we have high quality reportings, with critical assessment of individual studies and standardised data analysis methods, and could give us more systematic conclusions for this research area. However, we also have low quality parts, with poor systematic search and data description, and the conclusion of reporting is unbelievable.	If we recoding our study plan in advance, we could have systematic understanding on the research area, know more about the development of this field, understand what is the most cutting-edge research and what is lacking in this field.	Sometimes, an individual study could influence the results or conclusion of a meta-analysis, a critical assessment of individual study quality is needed.
s70	Having read the PRISMA-EcoEvo, my ranking is based on what I think I did so far, or what I have seen as a co-author.	It might be a good idea. I will consider it in the future.	It might be difficult to develop unambiguous tools to qualitatively estimate primary study quality. I thought that weighting effect sizes aimed at accounting for differences in primary data quality. If you ask, either it is not enough, either I am wrong.
s71	It should be mandatory to publish.	–	–
s72	I think availability of data and metadata could definitely improve.	–	–

Table S5 continued

Respondent ID	Comments on reporting quality and/or PRISMA-EcoEvo	Comments on registration of meta-analyses	Comments on critical appraisal of included studies
s73	Although I read a fair few meta- and systematic analyses I have not perhaps read enough to be confident in assigning a rating to each of these metrics. In my experience, some studies do all of these things very well while others report these things in an unsatisfactory way. For the latter, those studies could be improved in their reporting across the board.	–	If clear and unbiased criteria can be set that enhances the confidence in the quality of studies included in a given meta-analysis then that would be a positive thing - but I think that considerable challenges might await decisions regarding quantifying quality in terms of individual studies.
s74	–	–	Only if criteria for critical appraisal are very clear and not prone to bias
s75	It has gotten better over past decade as journals require data publication and authors become more aware of reporting standards	I co-authored 10+ meta-analyses since 2007 but was not aware of this option while analysing, writing (last published 2017)	This is challenging. Data science is making it easier to assimilate data without critical appraisal, or with code alone to check data quality.
s76	–	it would be useful to make sure there is no duplication	should be based on extremely simple decision rules

References: Surveys

VAN AERT, R.C.M., WICHERTS, J.M. & VAN ASSEN, M.A.L.M. (2016). Conducting meta-analyses based on p values: reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science* **11**, 713–729.

BEGG, C.B. & MAZUMDAR, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**, 1088–1101.

COPAS, J.B. & LI, H.G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society: Series B-Methodological* **59**, 55–95.

DUVAL, S. & TWEEDIE, R. (2000*a*). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* **95**, 89–98.

DUVAL, S. & TWEEDIE, R. (2000*b*). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463.

EGGER, M., SMITH, G.D., SCHNEIDER, M. & MINDER, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629–634.

FRASER, H., PARKER, T., NAKAGAWA, S., BARNETT, A. & FIDLER, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303.

HEDGES, L.V. (1984). Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics* **9**, 61–85.

IYENGAR, S. & GREENHOUSE, J.B. (1988). Selection models and the file drawer problem. *Statistical Science* **3**, 109–117.

KORICHEVA, J., GUREVITCH, J & MENGERSEN, K. (2013). *Handbook of Meta-analysis in Ecology and Evolution*. Princeton University Press, Princeton and Oxford.

Koricheva, J. & Gurevitch, J. (2014). Uses and misuses of meta-analysis in plant ecology. *Journal of Ecology* **102**, 828–844.

McShane, B.B., Böckenholt, U. & Hansen, K.T. (2016). Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science* **11**, 730–749.

Nakagawa, S. & Santos, E.S.A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* **26**, 1253–1274.

Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics* **8**, 157–159.

R Core Team. (2018). R: A language and environment for statistical computing.

Rosenberg, M.S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution* **59**, 464–468.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**, 638–641.

Senior, A.M., Grueber, C.E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E.S.A. & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology* **97**, 3293–3299.

Wang, M.C. & Bushman, B.J. (1998). Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods* **3**, 46–54.