Received: 28 November 2019 | Revised: 29 June 2020 | Accepted: 29 June 2020

DOI: 10.1002/jrsm.1433

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Assessing the risk of bias in choice of search sources for environmental meta-analyses

Ko Konno[1] | **Andrew S. Pullin**[2]

[1]School of Natural Sciences, Bangor University, Gwynedd, UK

[2]Centre for Evidence-based Conservation, School of Natural Sciences, Bangor University, Gwynedd, UK

Correspondence
Ko Konno, School of Natural Sciences, Bangor University, Gwynedd LL57 2UW, UK.
Email: afuaad@bangor.ac.uk

Results of meta-analyses are potentially valuable for informing environmental policy and practice decisions. However, selective sampling of primary studies through searches exclusively using widely used bibliographic platform(s) could bias estimates of effect sizes. Such search strategies are common in environmental evidence reviews, and if risk of bias can be detected, this would provide the first empirical evidence that comprehensiveness of searches needs to be improved. We compare the impact of using single and multiple bibliographic platform(s) searches vs more comprehensive searches on estimates of mean effect sizes. We used 137 published meta-analyses, based on multiple source searches, analyzing 9388 studies: 8095 sourced from commercially published articles; and 1293 from grey literature and unpublished data. Single-platform and multiple-platform searches missed studies in 100 and 80 of the meta-analyses, respectively: 52 and 46 meta-analyses provided larger-effect estimates; 32 and 28 meta-analyses provided smaller-effect estimates; eight and four meta-analyses provided opposite direction of estimates; and two each were unable to estimate effects due to missing all studies. Further, we found significant positive log-linear relationships between proportions of studies missed and the deviations of mean effect sizes, suggesting that as the number of studies missed increases, deviation of mean effect size is likely to expand. We also found significant differences in mean effect sizes between indexed and non-indexed studies for 35% of meta-analyses, indicating high risk of bias when the searches were restricted. We conclude that the restricted searches are likely to lead to unrepresentative samples of studies and biased estimates of true effects.

KEYWORDS
availability bias, database bias, language bias, location bias, publication bias

1 | INTRODUCTION

Since the emergence of meta-analysis, a statistical tool for combining the magnitude of effect (effect size) across different studies, reporting the quantitative aggregation of effect has become commonplace,[1-3] and it is now often performed in the "gold standard" evidence synthesis methodology: systematic reviews, in which comprehensive searches for relevant primary studies are conducted, thus enabling more reliable statistical estimate of effect of interest.[4-6] The contributions of meta-analysis are not limited to enabling quantitative assessment of effect of interest and its variation, but also represent a cultural change in the use of scientific evidence, as well as raising the standard of reporting, and therefore meta-analysis is recognized as an essential contributor to scientific progress.[7]

However, earlier research has shown that omission of certain data sources could lead to biased estimate of effect of interest due to revealed biases in statistical results. For example, it is widely recognized that exclusion of grey literature and unpublished data can impact on meta-analytical estimates due to publication bias.[8,9] Also, it has been shown that if relevant non-English-language literature is omitted during literature searches, different meta-analytical inferences could be drawn due to language bias.[10-12] Such empirical evidence suggests that if meta-analytical reviewers intentionally or unintentionally sample primary studies in a selective way during literature searches (eg, collect only studies from commercially published articles or those published in English-language), meta-analyses are expected to provide potentially biased estimates of overall effects. The reliability of statistical estimates therefore largely depends on how literature searching or data collection is conducted, hence the search strategy is of paramount importance to meta-analytical estimates.[5,6]

One of the most commonly applied search strategies in environmental evidence reviews is the sole use of widely recognized bibliographic platform(s),[13] for example, Web of Science ("WoS" hereon; www.webofknowledge.com). Earlier studies showed that the vast majority of environmental evidence reviews relied on limited bibliographic sources,[14,15] presumably because academic institutions have access to such platforms for educational and research purposes, and therefore academic researchers are familiar with them, and if, for example, resources are limited (eg, time, funding) or reviewers are unwilling to carry out supplementary searches (eg, web-based searches, contacting experts and stakeholders), certain bibliographic platform(s) are used as the sole source for searching literature. A major problem in using single bibliographic platforms as search sources is that they contain only selected records from all existing data (ie, records selectively made available to users as a service) with a very strong preference for commercially published articles,[13] and therefore it is possible that meta-analytical estimates relying on such practice will be affected (Figure 1). In other words, studies indexed in such bibliographic platform(s) may not represent the true population of relevance, hence the impacts of availability bias—intentional or unintentional selective sampling of studies that are easily accessible to samplers[16,17]—on overall mean effect sizes may be observed. In this paper, we use as a measure of availability bias whether or not studies are indexed in certain bibliographic platform(s) (ie, search sources). Note bias and random error cannot often be distinguished, thus we assess the risk of availability bias rather than testing presence or absence of availability bias (see Reference 18 for the concept of "risk of bias"). To date, no studies assessed the impacts of search sources on environmental meta-analytical estimates.[19] Since the number of environmental meta-analyses are increasing rapidly,[7,20] and are increasingly used to inform policy and practice, the assessment of risk of bias in search sources is a critical element of review conduct. The Collaboration for Environmental Evidence (CEE; www.environmentalevidence.org), an independent organization that provides guidelines and standards for environmental evidence synthesis, has long been advocating the conduct of searches that are not solely dependent on bibliographic platforms.[6] However, despite the formal establishment and development of environmental evidence synthesis methodology,[6,21] the importance of such comprehensiveness of search strategy remains underappreciated by the scientific community of the environmental sector,[14,15,22] presumably because including such comprehensive search requires more resources and

What is already known?

- The reliability (transparency and repeatability) of search strategies of environmental evidence reviews needs to be improved.

What is new?

- First assessment of impacts of search strategies relying only on widely used bibliographic platform(s) on effect sizes provided in published environmental meta-analyses.
- Restricting searches to a few, widely used, bibliographic platform(s) may lead to provision of biased estimates of effect sizes. Such practice is unlikely to lead to representative samples of primary studies due to missing studies from grey literature, unpublished data and non-English-language publications.

Potential impact for RSM readers outside the authors' field

- Although it is crude, we estimate that about 35% of meta-analyses have high risk of availability bias when restricting searches to widely used bibliographic platform(s). The generalizability of this finding needs to be investigated in other fields.

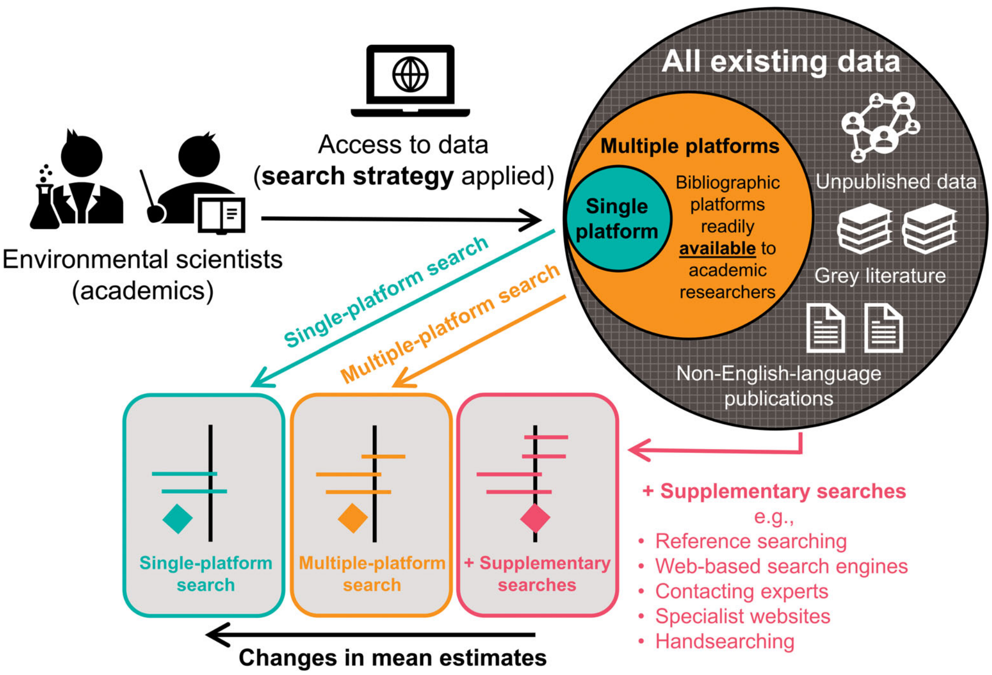

FIGURE 1 Conceptual model. Single-platform and multiple-platform searches may miss studies that are not indexed in the bibliographic platform(s). It is therefore expected if academic researchers rely only on these search sources, the estimates of mean effect sizes will be affected [Colour figure can be viewed at wileyonlinelibrary.com]

time,[23] and/or to date there is no empirical evidence that such comprehensiveness of literature search matters to outcomes of environmental evidence syntheses.

Here we evaluate impacts of searches restricted to widely used bibliographic platform(s) on outcomes of meta-analyses, and assess the risk of availability bias by using published meta-analyses that are based on multiple source searches and comparing overall mean effect sizes between: (a) studies indexed in one single platform (WoS) and all studies included in published meta-analyses using both unweighted meta-analysis ("unweighted single-platform search group" hereon) and weighted meta-analysis ("weighted single-platform search group" hereon); (b) studies indexed in multiple platforms (WoS and six other platforms; see below) and all studies included in published meta-analyses using both unweighted meta-analysis ("unweighted multiple-platform search group" hereon) and weighted meta-analysis ("weighted multiple-platform search group" hereon). We also analyze the effect of proportions of studies missed by the single-platform search and the multiple-platform search on the deviations of overall mean effect sizes for investigating the pattern across the meta-analyses in all the four comparison groups. We then compare empirical models (based on our observations of whether studies are indexed or not) and simulation models (random sampling of studies) to assess whether random sampling of studies reduce deviations of mean effect sizes across the meta-analyses. Furthermore, we directly compare mean effect sizes between studies that are indexed in the platform(s) ("indexed" hereon) and those not indexed ("non-indexed" hereon) to assess the risk of availability bias for individual meta-analyses. Finally, we provide guidance on search strategy development for reducing the risk of bias.

2 | METHODS

2.1 | Overview

An overview of the workflow of this study is provided in Figure 2. We first conducted searches for published meta-analyses that were based on multiple source searches (Section 2.2) followed by eligibility screening (Section 2.3), and checking study index status (in which sources studies are indexed) for both single-platform search and multiple-platform search groups and study sources (Section 2.4). We then quantified effect size differences due to the exclusion of studies that were not indexed in the bibliographic platform(s) (Section 2.5). Further, we assessed relationships between effect size differences and proportions of studies missed by the restricted searches to explore the pattern across the meta-analyses, followed by simulation modeling and subgroup analyses that tested whether the existence of studies from grey literature and unpublished data in the original datasets had an effect on the differences in mean effect sizes (Section 2.6). The simulation models were to assess the effect of randomization of primary studies on deviations of mean effect sizes across the meta-analyses. The subgroup analyses were conducted because we hypothesized that original meta-analyses that involved grey literature and unpublished data would be more affected (ie, result in larger deviations of effect sizes)

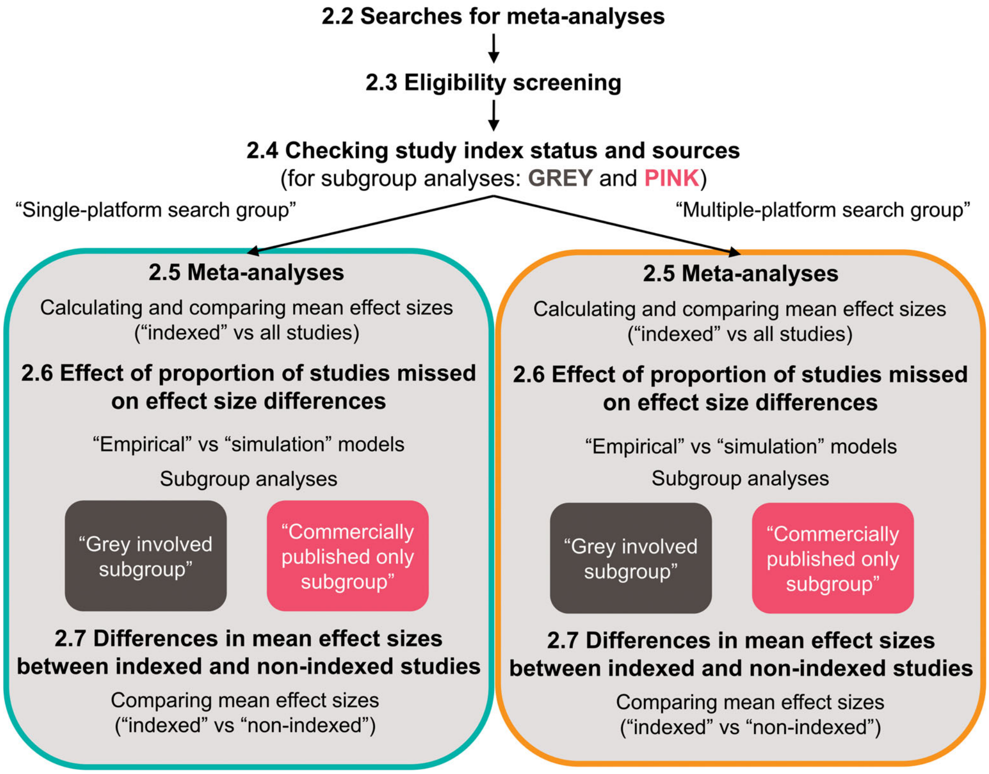

FIGURE 2 Overview of this study. Words in bold represent the subsections. "Single-platform search group" (green) uses Web of Science. "Multiple-platform search group" (orange) uses: Web of Science; CAB Direct; ProQuest; ScienceDirect; Wiley Online Library; JSTOR; and BioOne COMPLETE (Section 2.4). Where original meta-analyses include studies from grey literature and unpublished data, they are classified as "grey involved subgroup". Others are classified as "commercially published only subgroup" (Section 2.4) [Colour figure can be viewed at wileyonlinelibrary.com]

by the restricted searches compared to those that involved commercially published articles only, even when the proportion of studies missed was taken into account. To assess the risk of bias for each individual meta-analysis, we directly compared mean effect sizes between indexed and non-indexed studies for each of the unweighted search groups (Section 2.7).

2.2 | Searches for meta-analyses

We conducted searches in Web of Science Core Collection ("WoSCC" hereon), Scopus (www.scopus.com), CAB Direct (www.cabdirect.org) and used two search engines: Bielefeld Academic Search Engine ("BASE"; www.base-search.net); and Google Scholar via Publish or Perish version 6 (harzing.com/resources/publish-or-perish) to collect potentially relevant meta-analytical articles published between 2014 and 2018 (see Appendix S1 for detailed search strategy and results). The search results from WoSCC were imported into EndNote Basic (endnote.com/product-details/basic), and then the records were converted into RIS files for importing to CADIMA (www.cadima.info).[24] The search results from Scopus, CAB Direct, BASE, and Google Scholar were exported as RIS files for importing to CADIMA. We restricted the year in order to assess up-to-date meta-analyses as environmental evidence synthesis methodology is being rapidly developed[6] and the first textbook of meta-analysis in the environmental field was published in 2013.[5]

2.3 | Eligibility screening

We ran the automatic duplicate removal function (based on title) in CADIMA. We then applied the same eligibility criteria as O'Leary et al 2016[14] that states: (a) "reviews should be undertaken in relation to a specific question or topic of relevance to environmental management and have recommendations for policy or practice"; and (b) "article type should be a review and/or synthesis of primary research". We also applied further eligibility criteria to collect only meta-analytical articles that used effect sizes and conducted multiple source searches: (c) effect size—a meta-analysis must have used one of the following effect sizes: raw mean difference; standardized mean difference; response ratio; odds ratio; risk ratio; risk difference; or Fisher's z transformed r; and (d) search source—a meta-analysis must have been based on search of at least one bibliographic platform (WoS) and another source. Screening was conducted at two stages: title and abstract in CADIMA; and full text using a spreadsheet and collected PDFs. Unobtainable records and reasons for exclusion at full text were recorded. Meta-analytical articles were excluded where they did not provide analyzable datasets (eg, Microsoft Excel Workbook), which enable calculation of effect sizes or contain effect sizes, as supplementary materials or in tables or appendixes. Unconfigurable or unclear datasets were excluded. This procedure of identifying datasets divided papers into meta-analysis level since one paper may produce more than one mean effect sizes to address multiple populations, interventions/exposures and/or outcomes questions that are common in the environmental sector. To ensure that we do not count the same primary studies twice in each meta-analysis, we used data for aggregation of all studies when overall mean effect size was presented, and the data were openly accessible. In cases where overall mean effect size was not presented or data only for subgroup analysis were available, we used the data for subgroup analyses (eg, taxonomic groups that were deemed unsuitable for aggregation in the original papers).

2.4 | Checking study index status and sources

We checked whether studies in the analyzable datasets providing sufficient information were indexed in one single platform: WoS (with a default setting of WoSCC database) and in multiple platforms: WoS (All Databases option which includes WoSCC); CAB Direct (CAB ABSTRACTS, Global Health, and CABI Full Text); ProQuest (Core Databases; search.proquest.com); ScienceDirect (www.sciencedirect.com); Wiley Online Library (onlinelibrary.wiley.com); JSTOR (www.jstor.org); and BioOne COMPLETE (bioone.org) via Bangor University institutional access (see details in Appendix S2). WoS was chosen as the single-platform search because it is the most commonly searched platform in environmental evidence reviews; the other six platforms were chosen as they are also commonly searched in environmental evidence reviews. Records' titles and/or Digital Object Identifiers (DOIs) were used to check the index status of each study in the platforms. Where studies were indexed in a platforms, we treated them as retrieved by searches of that platform because our interest was not in the performance of search strategies actually used, rather the risk of sole reliance on studies that were indexed in the platforms (hence findable).[19] This approach was chosen because applying the original search strings might miss studies indexed in the platforms.[19] To enable subgroup analyses, we also checked study sources and classified those into: (a) commercially published article; and (b) grey literature and unpublished data by applying the Luxembourg definition of grey literature: "manifold document types produced on all levels of government, academics, business, and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by libraries and institutional repositories, but not controlled by commercial publishers".[25] For those studies that had already been classified as unpublished, thesis, dissertation, conference proceedings, organizational or government report, or equivalent in the datasets for meta-analyses, we simply classified as: grey literature and unpublished data. For those studies not indexed in the platforms or not listed as commercially published articles, we searched for the study sources using Google Scholar (scholar.google.co.uk), Google (www.google.co.uk), and ResearchGate (www.researchgate.net). We then classified each eligible meta-analysis (ie, each dataset for producing a mean effect size) into two subgroups: (a) original datasets containing studies from commercially published articles only ("commercially published only subgroup" hereon; Figure 2); and (b) original datasets containing studies from grey literature and unpublished data ("grey involved subgroup" hereon; Figure 2).

2.5 | Meta-analyses

We calculated effect sizes and their variances where datasets did not contain effect sizes but did contain sufficient information such as means, standard deviations and sample sizes for both groups: comparator and treatment. We chose the same effect size metrics as the original meta-analyses, for example, where log response ratios were used, we calculated log response ratios.[26] We otherwise used the effect sizes provided in the original datasets. To investigate weighted effect sizes, we standardized the methods of weighting by focusing on inverse of variance as our interest was not to conclude or infer effects, rather to investigate deviations of mean effect sizes by treating all meta-analyses equally.[27] Where original meta-analyses were conducted under a fixed effect model, we weighted by inverse of within-study variance, and where meta-analyses were conducted under random or mixed-effects model, we weighted by inverse of within- and between-study variances (ie, pooled variance) under a random effects model[27]; we did not conduct meta-analyses under mixed-effects models as our interest was not to examine interaction of assessed intervention or exposure and other factors. We did not conduct weighting where original meta-analyses did not weight studies, or datasets did not provide sufficient information. All meta-analyses were performed using the metafor package[28] in R version 3.5.0.[29]

2.6 | Effect of proportion of studies missed on effect size differences

For each eligible meta-analysis (ie, each dataset for producing a mean effect size) we calculated overall mean effect size differences between: (a) studies indexed in the single platform and all studies included in the original meta-analysis (the unweighted and weighted single-platform search groups); and (b) studies indexed in the multiple platforms and all studies included in the original meta-analysis (the unweighted and weighted multiple-platform search groups) by subtracting overall mean effect size of studies indexed in the platform(s) from overall mean effect size of all studies included in the original meta-analyses. Effect size differences can occur in both positive and negative directions and so where effect size differences were negative, we transformed those to positive values, and then log-transformed all the values for obtaining a log-normal distribution[30]; we did not convert any effect size metrics as there were different meanings and assumptions in each metric of effect size.[27] We then calculated a proportion (%) of studies missed by the single-platform search and the multiple-platform search separately for each meta-analysis by dividing the number of studies missed by the number of all studies included in the original meta-analysis. The proportions of studies missed were then log-transformed to be log-normally distributed. We then fitted a log-linear regression model for each group: (a) unweighted single-platform search group; (b) weighted single-platform search group; (c) unweighted multiple-platform search group; and (iv) weighted multiple-platform search group, with log (Δ *mean effect size*) as a response variable and the log (*proportion of studies missed*, %) as an explanatory variable. We also developed simulation models based on randomly selected studies with the actual proportions of indexed studies for the unweighted single-platform search group and the unweighted multiple-platform search group. For example, when a single-platform search missed 20% of the studies included in an original meta-analysis, we randomly sampled 80% of the studies. We replicated the randomised sampling of studies 1000 times for each meta-analysis (ie, number of meta-analyses × 1000 runs for each of the unweighted single-platform search group (98 000 runs) and the unweighted multiple-platform search group (78 000 runs)). We calculated a mean of the 1000 replicates for each meta-analysis. We then calculated mean effect size differences (as positive values), and then log-transformed the effect size differences as described above. These simulations of randomized sampling of studies were to compare slopes (effect sizes, *b*) and the intercepts (*c*) against the empirical models (ie, actual observations; described above). We tested for statistical significance by comparing the empirical and simulation models. Further, we conducted a subgroup analysis for each group (excluding the simulation models) that compared two log-linear models with and without a binomial fixed factor: *subgroup* (the existence of grey literature and unpublished data in the original datasets; categorization described above; Figure 2) to test whether the factor had an influence on the differences in mean effect sizes. To statistically test significance of a potential confounding factor: *effect size metrics*, we added the fixed factor to the above full log-linear models (ie, log (Δ *mean effect size*) ~ log (*proportion of studies missed*, %) + *subgroup* + *effect size metrics*) and compared to reduced models (without the fixed factor: *effect size metrics*). Statistical tests were conducted in R version 3.5.0.[29]

2.7 | Differences in mean effect sizes between indexed and non-indexed studies

In order to avoid statistical results to be correlated between meta-analyses provided in a same paper (non-independent), we selected one meta-analysis from each

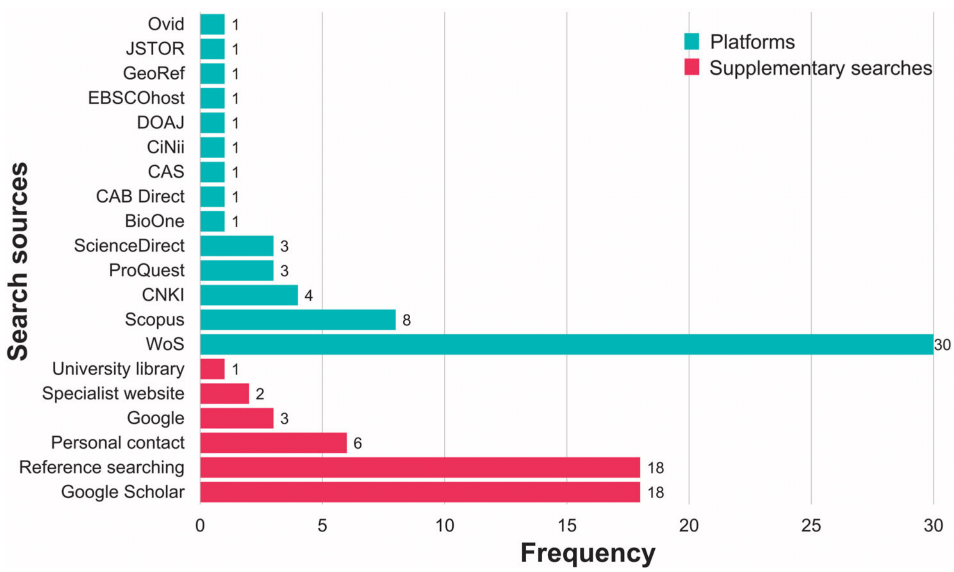

FIGURE 3 Frequency of searched platforms and supplementary searches reported in the eligible metaanalytical articles (n = 30) [Colour figure can be viewed at wileyonlinelibrary.com]

article that enabled the most potentially meaningful comparison of unweighted-mean effect sizes between indexed and non-indexed studies. For each paper, we selected one that obtained the largest sample size in the lower sample size groups (eg, 10 vs 10 would be preferred over 100 vs 5). Where lower sample size groups provided the same sample size, we selected the one that obtained the largest total sample size (eg, 100 vs 10 would be preferred over 10 vs 10). We then compared unweighted-mean effect sizes between indexed and non-indexed studies using Welch two-sample *t*-test for both the unweighted single-platform and the unweighted multiple-platform search groups.[31] As sensitivity analyses, we fitted a log-linear regression model described above (without subgroup analyses) using these subsets of meta-analyses for both the unweighted single-platform search group and the unweighted multiple-platform search group. We further compared the simulation and the empirical models for these subsets of meta-analyses using the same methods described above, and then tested for statistical significance.

3 | RESULTS

3.1 | Description of dataset

Thirty articles from 22 journals out of 28 139 unique records retrieved by the searches for meta-analyses met all the eligibility criteria (flow diagram: Appendix S3; list of eligible articles: Appendix S4; list of articles excluded at full text and unobtainable records: Appendix S5), resulting in 137 eligible meta-analyses (ie, 137 overall mean effect sizes without duplication of primary studies between meta-analyses presented in the 30 articles,[32-61] hence retaining independence of individual effect sizes). These 30 eligible meta-analytical articles conducted diverse searches in addition to the single-platform search in WoS (Figure 3). The most frequently used platform after WoS was Scopus followed by China Knowledge Resource Integrated Database (CNKI), ProQuest, ScienceDirect, BioOne, CAB Direct, Chinese Academy of Sciences (CAS), CiNii, Directory of Open Access Journals (DOAJ), EBSCOhost, GeoRef, JSTOR, and Ovid. The most frequently applied supplementary searches were Google Scholar and reference searching (also known as snowballing) followed by personal contact, Google, specialist website, and university library. These 137 meta-analyses involved 9388 primary studies (number of studies per meta-analysis: median = 23; minimum = 2; maximum = 1490): 8095 studies from commercially published articles; and 1293 studies from grey literature and unpublished data. Effect sizes used in the 137 meta-analyses were log response ratio; Hedges' *d*; nonlog response ratio; Fisher's *z* transformed *r*; log risk ratio; and log odds ratio (Figure 4). We also found that 62 of 78 meta-analytical articles excluded on the basis of limited search sources (ie, those not conducting multiple source searches) used WoS only (79%), indicating that WoS only search was indeed commonly applied in the environmental sector (Appendices S5 and S6).

The single-platform search (WoS) captured all studies in 37 meta-analyses but did not capture all studies in 100 meta-analyses (Figure 5A). In two of the 137 meta-analyses, the single-platform search missed all studies, and therefore 98 overall mean effect sizes could be compared between studies indexed in the single platform and all studies included in the original meta-analyses. Effect sizes used in the 98 meta-analyses were: log response ratio; Hedges' *d*; nonlog response ratio; Fisher's *z* transformed *r*; log risk ratio; and log odds ratio (Figure 4). Sixty-six of the 98 meta-analyses could be

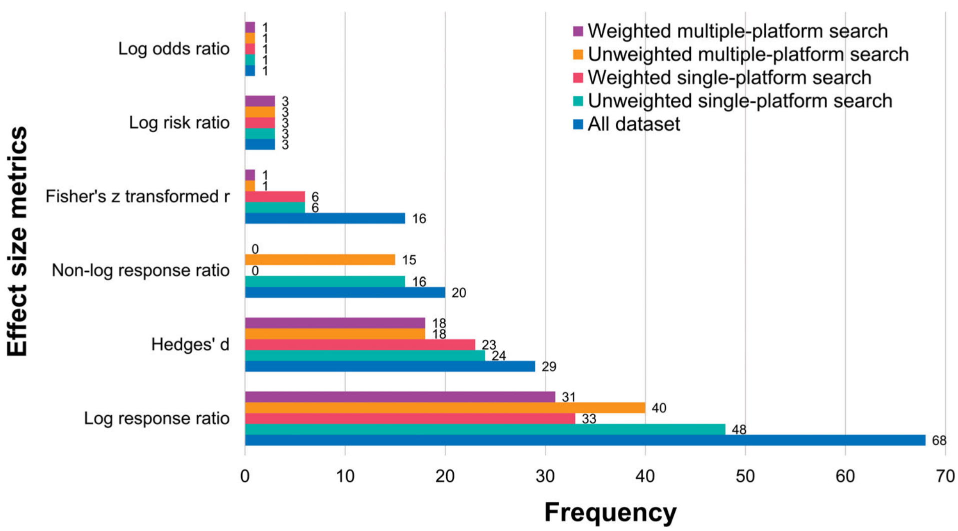

FIGURE 4 Frequency of effect size metrics used in the eligible meta-analyses (all dataset; n = 137), and those that are comparable within groups: unweighted single-platform search group (n = 98); weighted single-platform search group (n = 66); unweighted multiple-platform search group (n = 78); weighted multiple-platform search group (n = 54) [Colour figure can be viewed at wileyonlinelibrary.com]

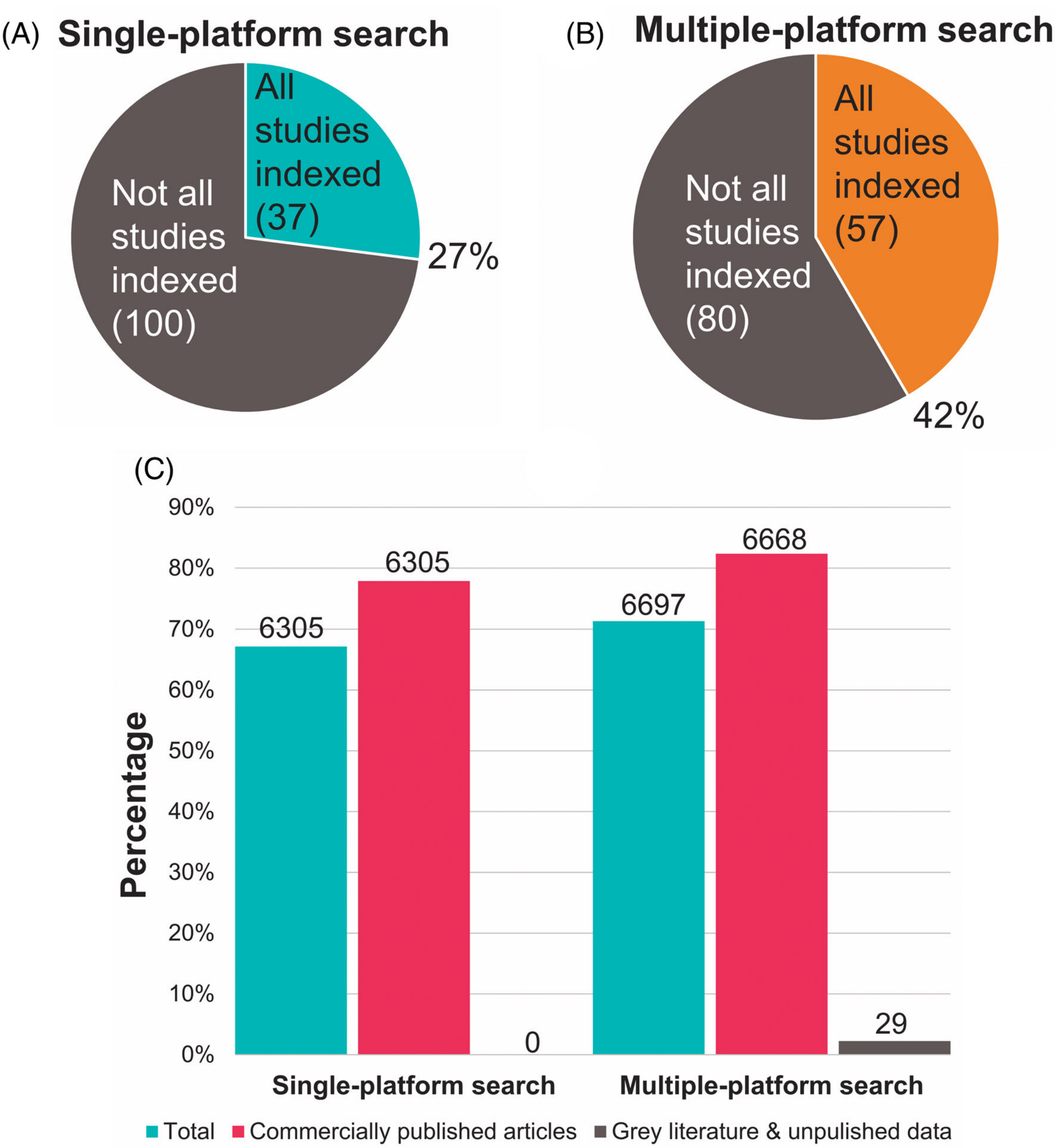

FIGURE 5 Proportions of meta-analyses in which all studies were indexed in: A, the single platform; and B, the multiple platforms. Values in brackets show the number of meta-analyses. C, Proportions of studies indexed in the single platform and the multiple platforms. Values represent the number of indexed primary studies [Colour figure can be viewed at wileyonlinelibrary.com]

weighted (Figure 4). In total, 6305 studies were indexed in the single platform; these indexed studies were all from commercially published articles, and therefore all studies from grey literature and unpublished data were missed by the single-platform search (Figure 5C).

The multiple-platform search captured all studies in 57 meta-analyses, that was 15% higher than the single-platform search (Figure 5B). However, the multiple-platform search did not capture all studies in 80 meta-analyses and missed all studies in two meta-analyses, and therefore 78 meta-analyses could be compared between studies indexed in the multiple platforms and all studies in the original meta-analyses. Effect sizes used in the 78 meta-analyses were: log response ratio; Hedges' d; nonlog response ratio; log risk ratio; log odds ratio; and Fisher's z transformed r (Figure 4). Fifty-four of the 78 meta-analyses could be weighted (Figure 4). In total, 6697 studies were indexed in the multiple platforms: 6668 studies from commercially published articles and 29 studies from grey literature and unpublished data (Figure 5c). The vast majority of studies from commercially published articles that were not indexed in the multiple platforms were those published in non-English-languages (99%; n = 1415).

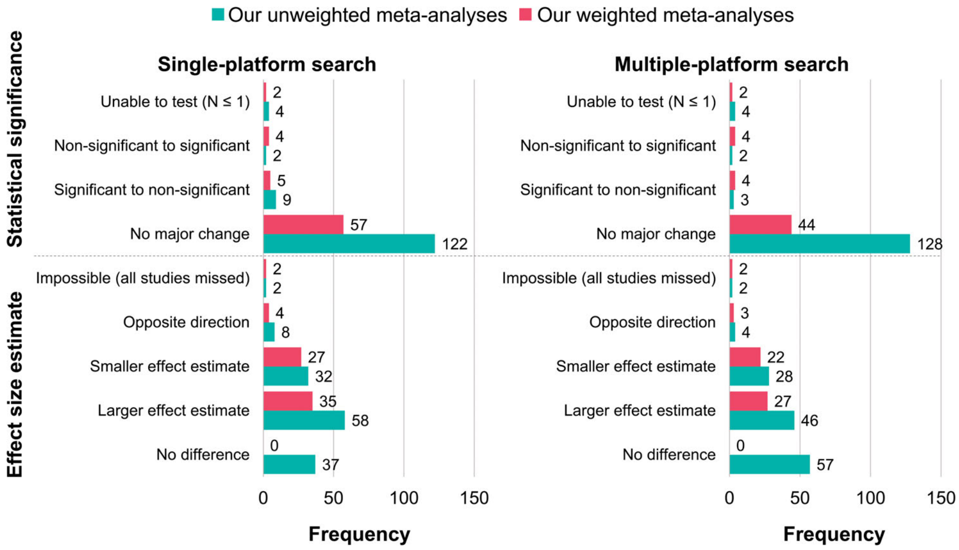

FIGURE 6 Consequences of the single-platform search (left): 137 unweighted meta-analyses and 68 weighted meta-analyses; and the multiple-platform search (right): 137 unweighted meta-analyses and 54 weighted meta-analyses. Weighting were conducted for meta-analyses in which the searches did not capture all studies [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Impacts of the restricted searches on estimates and statistical significance

Within our meta-analyses, as consequences of the single-platform search (WoS), 58 unweighted and 35 weighted provided larger-effect estimates, and 32 unweighted and 27 weighted provided smaller-effect estimates (Figure 6; unweighted: Appendix S7 and weighted: Appendix S8). In eight unweighted and four weighted meta-analyses, the single-platform search even led to opposite direction of effects (Figure 6; unweighted: Appendix S7 and weighted: Appendix S8). There were 20 cases in which the single-platform search had impacts on statistical significance: nine unweighted and five weighted mean effect sizes became nonsignificant from significant; and two unweighted and four weighted mean effect sizes became significant from nonsignificant (Figure 6).

As consequences of the multiple-platform search, 47 unweighted and 28 weighted meta-analyses provided larger-effect estimates, and 28 unweighted and 22 weighted meta-analyses provided smaller-effect estimates (Figure 6; unweighted: Appendix S9 and weighted: Appendix S10). In four unweighted and three weighted meta-analyses, the multiple-platform search even led to opposite direction of effects (Figure 6; unweighted: Appendix S9 and weighted: Appendix S10). There were 13 cases that the multiple-platform search had impacts on statistical significance: three unweighted and four weighted mean effect sizes became nonsignificant from significant; and two unweighted and four weighted mean effect sizes became significant from nonsignificant (Figure 6). Overall, effect size differences were relatively greater for meta-analyses in which the proportions of studies missed were also relatively greater (see below).

3.3 | Effect of proportion of studies missed on effect size differences

Log-linear regression models indicated that there were significant positive linear relationships between the proportions of studies missed and differences in mean effect sizes in all the four groups (Figure 7): (a) unweighted single-platform search group ($F_{1,\ 96} = 35.36$; $P < .001$); (b) weighted single-platform search group ($F_{1,\ 64} = 27.09$; $P < .001$); (c) unweighted multiple-platform search group ($F_{1,\ 76} = 33.37$; $P < .001$); and (d) weighted multiple-platform search group ($F_{1,\ 52} = 27.51$; $P < .001$). Effect sizes (slopes, b) were larger in the single-platform search group than the multiple-platform search group (112% larger in the unweighted; 108% larger in the weighted). These log-linear relationships remained in the simulation models based on randomly selected studies for both the unweighted single-platform search group ($F_{1,\ 96} = 15.19$; $P < .001$) and the unweighted multiple-platform search group ($F_{1,\ 76} = 7.16$; $P < .001$). Their effect sizes (slopes, b) were smaller ($b = 0.63$ in the two simulation models) compared to the empirical models which provided 151% and 135% larger effect sizes (b), respectively (Figure 7; Appendix S11). Deviations of mean effect sizes were significantly greater in the empirical models compared to the simulation models, suggesting that restricting searches did not provide random samples of primary studies (Figure 7; Appendix S11). Further, subgroup analyses through model comparisons indicated that the grey involved subgroup had significantly larger overall effect size differences than the commercially published only subgroup in all the four groups (Figure 7): (a) unweighted single-platform search group ($F = 7.42$; $P = .008$); (b) weighted single-platform search group

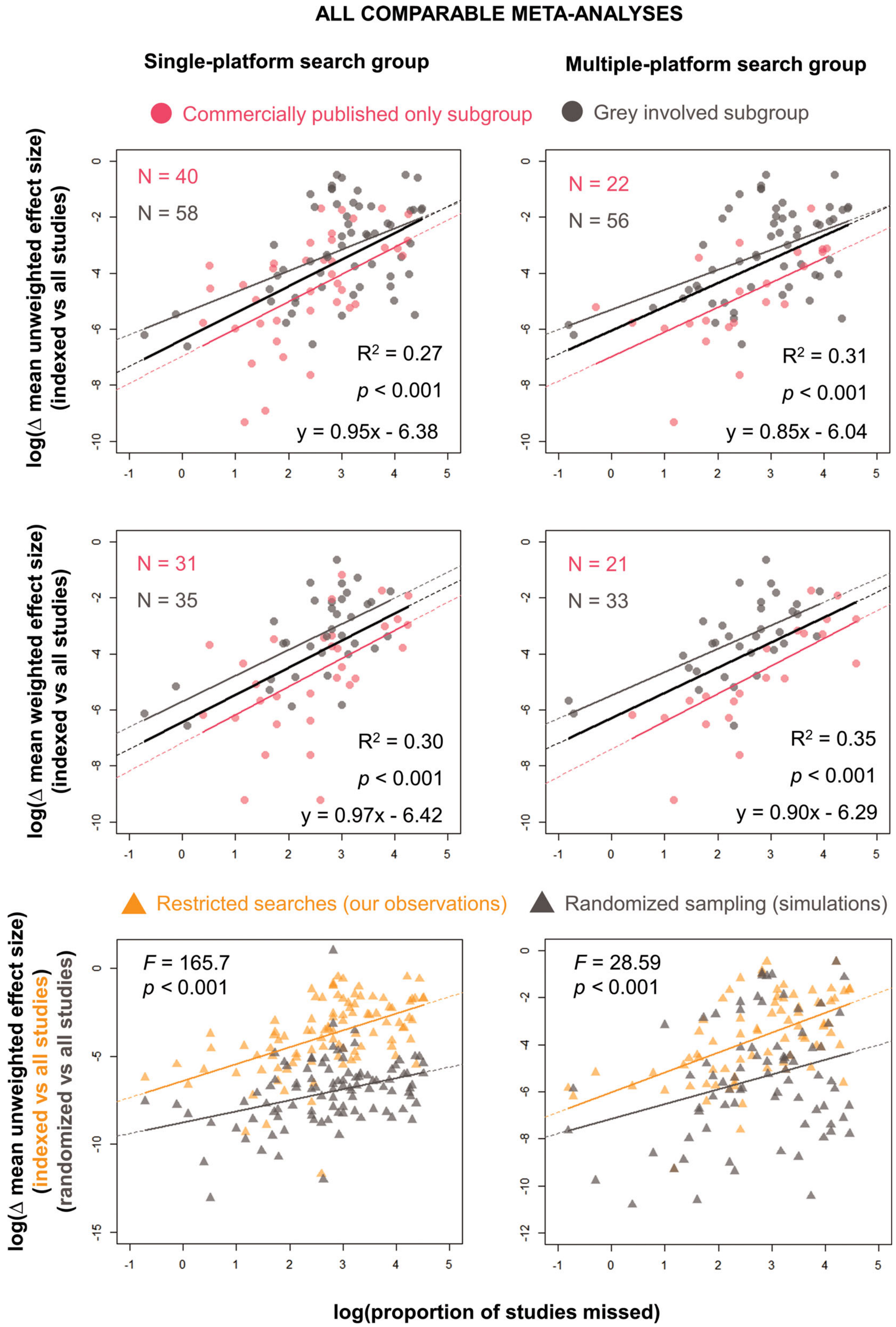

FIGURE 7 Relationships between proportions of studies missed and differences in mean effect sizes using all comparable meta-analyses. Top left: unweighted single-platform search group (n = 98). Middle left: weighted single-platform search group (n = 66). Bottom left: unweighted single-platform search group (n = 98) compared to the simulation model (randomized sampling of studies). Top right: unweighted multipleplatform search group (n = 78). Middle right: weighted multiple-platform search group (n = 54). Bottom right: unweighted multiple-platform search group (n = 78) compared to the simulation model (randomized sampling of studies). The black lines represent the best fits of all data. Dotted parts of the lines are beyond the log-linear models [Colour figure can be viewed at wileyonlinelibrary.com]

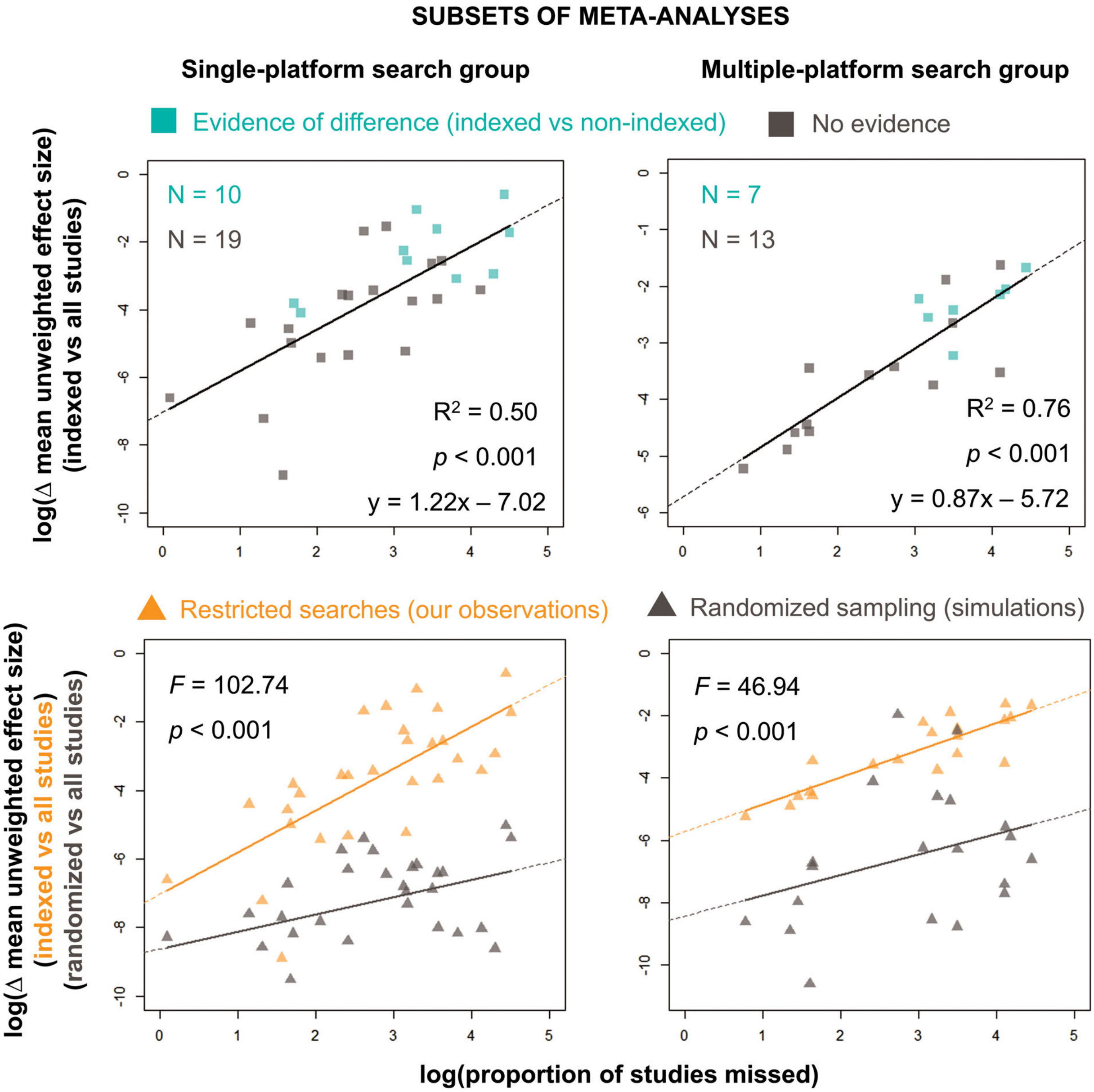

FIGURE 8 Relationships between proportions of studies missed and differences in mean effect sizes using subsets of meta-analyses. Top left: unweighted single-platform search group (n = 29). Bottom left: unweighted single-platform search group (n = 29) compared to the simulation model (randomized sampling of studies). Top right: unweighted multiple-platform search group (n = 20). Bottom right: unweighted multiple-platform search group (n = 20) compared to the simulation model (randomized sampling of studies). The lines represent the best fits. Dotted parts of the lines are beyond the log-linear models [Colour figure can be viewed at wileyonlinelibrary.com]

(F = 12.22; P < .001); (c) unweighted multiple-platform search group (F = 11.60; P = .001); and (d) weighted multiple-platform search group (F = 27.66; P < .001). A potential confounding factor: *effect size metrics* did not have a significant influence on the outcomes of the subgroup analyses in all the four groups: (a) unweighted single-platform search group (F = 1.71; P = .14); (b) weighted single-platform search group (F = 0.91; P = .47); (c) unweighted multiple-platform search group (F = 1.67; P = .15); and (d) weighted multiple-platform search group (F = 1.14; P = .35).

3.4 | Differences in mean effect sizes between indexed and non-indexed studies

For the single-platform search group, differences in unweighted-mean effect sizes between indexed and non-indexed studies could be tested for 29 meta-analyses (from 29 articles). We found evidence of differences in unweighted-mean effect sizes for 10 (34.5%) meta-analyses (Figure 8; Appendix S12). For the multiple-platform search group, differences in unweighted-mean effect sizes could be tested for 20 meta-analyses (from

20 articles). We found evidence of differences in unweighted-mean effect sizes for seven (35%) meta-analyses (Figure 8; Appendix S12). The significant log-linear relationships (described above) remained in these subsets of meta-analyses (single-platform search group: $F_{1,\ 27} = 27.2$; $P < .001$; multiple-platform search group: $F_{1,\ 18} = 56.22$; $P < .001$). The unweighted single-platform search group provided 140% larger effect size (slope, b) compared to the unweighted multiple-platform search group (Figure 8). Further, in the simulation model for the unweighted single-platform search group, the significant relationship remained ($F_{1,\ 27} = 5.70$; $P = .02$). However, it did not remain for the unweighted multiple-platform search group ($F_{1,\ 18} = 2.32$; $P = .15$). Their effect sizes (slopes, b) were smaller compared to the empirical models that provided 244% and 134% larger effect sizes (b), respectively (Figure 8; Appendix S11). Statistically significant differences between the empirical and the simulation models remained for the both groups (single-platform search group: $F = 27.2$; $P < .001$; multiple-platform search group: $F = 56.22$; $P < .001$).

4 | DISCUSSION

We have demonstrated that the restricted search strategies using only the widely used bibliographic platform(s) could affect the estimates of overall mean effect sizes in environmental meta-analyses. A single-platform search is especially risky with an estimation that 11% of meta-analyses could result in different statistical outcomes. The multiple-platform search could reduce the frequency of the adverse events by 4% although major influences could not be eliminated. Missing studies could affect not only the mean effect sizes and statistical significance, but also confidence intervals by narrowing or widening them. This seemingly depends on the missed studies' characteristics; where studies with relatively unique characteristics are missed (eg, studies on species that responded differently from the other species to given exposure), it might lead to narrower confidence intervals due to homogenization while studies with similar characteristics are missed (eg, studies on the same species to given exposure), it might lead to wider confidence intervals due to reduced sample size (Appendices S7-S10). Although it is crude, we estimate that about 35% of meta-analyses have high risk of availability bias (ie, mean effect sizes differ between indexed and non-indexed studies) when restricting searches to these widely used platform(s). These results imply that sole reliance on the widely used bibliographic platform(s) should be avoided to mitigate the risk of bias in search sources.

We also showed that there were significant positive log-linear relationships between the proportions of studies missed and the deviations of mean effect sizes. This suggests that as the number of studies missed increases, deviation of mean effect size is likely to expand. We revealed that slopes (effect sizes, b) were steeper where the single-platform search was conducted compared to the multiple-platform search, as well as to the simulation model (Figures 7 and 8; Appendix S11). One possible explanation for this is that studies retrieved by relatively comprehensive searches or random sampling better represent the true population of relevance compared to studies retrieved by the single-platform search. However, even the multiple-platform search missed a large proportion of studies from grey literature and unpublished data (98% of known studies from grey literature and unpublished data were missed), as well as from non-English-language publications (99% of missed studies from commercially published articles were those published in non-English-languages). The missing of non-English-language publications may be explained by revealed underrepresentation of non-English-language literature in widely used bibliographic platforms.[13] Thus, both the single-platform and the multiple-platform searches are unlikely to lead to random samples of primary studies. Further, the grey involved subgroup (original meta-analyses that involved grey literature and unpublished data) resulted in significantly larger deviations of mean effect sizes than the commercially published only subgroup (original meta-analyses that involved commercially published articles only). This seems to be an effect of publication bias[16,62]; missing studies from grey literature and unpublished data (the aforementioned 98%), on average, may have a stronger impact on meta-analytical estimates than missing studies from commercially published articles that are not indexed due to a systematic difference in reported effects between the two types of study source (Figure 7).

Given the demonstrated risk of bias in search sources, policy and practice decision-makers should be informed by reliable evidence reviews that are based on comprehensive literature searches including sources beyond the widely used bibliographic platforms. Shockingly, such practice is currently not common in environmental evidence reviews as we found that 229 (77%) of 298 excluded meta-analytical articles published between 2014 and 2018 did not report their search strategies or search multiple platforms and conduct supplementary searches (Appendix S6). These findings support earlier recommendations that comprehensiveness, transparency, and repeatability of search strategies need to be improved[14,15] as biased review outcomes might trigger ineffective or even harmful policy and practice.[63] Our findings are an indication that authors, editors and peer-reviewers need to take greater account of risk of bias in search strategies as a whole (including the performance of search strings) when producing environmental evidence reviews. In fact, restricting search strategy has already been cautioned against by organizations such as CEE and the Department for Environment, Food & Rural Affairs of the UK government—even in rapid evidence assessment—acknowledging the risk of bias.[6,64] Thanks to the formal establishment of systematic evidence assessment methodology, existing CEE evidence syntheses and government rapid evidence assessments usually meet the standards of searching additional sources of information and contacting external experts (eg, [65,66]). Our findings imply a need for formal training in developing search strategies for evidence assessment to further advance evidence-based approaches in environmental management and conservation disciplines just as practiced in the health sector.[67-70]

To reduce the risk of bias in search strategies as a whole, evidence reviewers should formally assess the performance of their search strategy prior to the conduct of actual searches.[6,71] This can be done in three steps: (a) creating a test list of relevant study sources that preferably include both commercially published and grey literature; (b) trying out a developed search strategy using bibliographic platforms and supplementary searches (eg, web-based searches such as using Google Scholar[72,73]); and (c) checking whether the search strategy captures all relevant records in the test list. This type of assessment often requires formal training of searches,[5] and therefore we recommend the *CEE Guidelines and Standards for Evidence Synthesis in Environmental Management*[6] and Livoreil et al 2017[74] as the first-step detailed guidance on developing search strategy. Haddaway et al developed ROSES, a set of reporting standards in the environmental sector, which may also help search strategy planning, documentation and reporting.[75]

This study has a few limitations. First, our inferences might depend on the methodology applied in the original meta-analyses as we used available data (ie, coded information). We assumed that consideration was given to study independence (eg, correlation between within-study effect sizes for multiple-treatment studies) prior to the original meta-analyses.[76,77] We also assumed that all primary studies were sufficiently valid although only two articles,[32,33] that provided three meta-analyses, reported the conduct and results of critical appraisals; this might mean that the other collected datasets had already been biased, for example, due to selection bias, detection bias, performance bias, attrition bias that would occur within primary studies.[6,62] Second, we did not control for potential confounders other than effect size metrics. It is theoretically possible that there are associations between study characteristic factors and search sources (ie, certain study characteristics are more likely to be indexed), and these better explain the deviations of effect sizes. However, it is likely that the risk of availability bias will increase when there are associations (ie, certain study characteristics are more likely to be retrieved). Also, even when there are associations, the assessed impacts on meta-analyses would not differ from the results because missed studies remain missed studies. Third, the level of access to databases differs between institutions. This means that studies classified as indexed in the single platform or the multiple platforms in this study may not be accessible from other institutions. Hence, the risk should differ according to the level of access. Also, we did not have full access, and therefore studies classified as not indexed in this study may be accessible from other institutional access (see Appendices S2 and S4 for our access level and dates checked). In order to conduct more comprehensive assessment of risk of bias in search sources, we need more complete reporting of search strategies.

Our findings nevertheless suggest that environmental evidence reviewers should be cautious about availability bias. Conducting search strategies that use only widely used bibliographic platform(s) should preferably be avoided to reduce the risk of bias in search sources.

ACKNOWLEDGMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the editor and two anonymous reviewers for their constructive comments that greatly improved the manuscript.

CONFLICT OF INTEREST

The authors reported no conflict of interest.

DATA AVAILABILITY STATEMENT

Original datasets for meta-analyses were accessed via Bangor University subscription.

ORCID

Ko Konno https://orcid.org/0000-0002-5319-9178

REFERENCES

1. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res*. 1976;5(10):3-8. https://doi.org/10.3102/0013189X005010003.
2. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol*. 1977;32(9):752-760. https://doi.org/10.1037/0003-066X.32.9.752.
3. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. San Diego, CA: Academic Press; 1985.

4. Dicks L, Haddaway N, Hernández-Morcillo M, et al. Knowledge synthesis for environmental decisions: an evaluation of existing methods, and guidance for their selection, use and development. 2017;84. http://www.eklipse-mechanism.eu/apps/Eklipse_data/website/EKLIPSE_D3-1-Report_FINAL_WithCovers_V6.pdf%5Cnhttp://www.eklipse-mechanism.eu/eklipse_outputs.
5. Koricheva J, Gurevitch J, Mengersen KL. *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton, NJ: Princeton University Press; 2013.
6. CEE. *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Version 5.0. 2018. http://www.environmentalevidence.org/information-for-authors
7. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nature*. 2018;555:175-182. https://doi.org/10.1038/nature25753.
8. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet (London, England)*. 2000;356(9237):1228-1231. https://doi.org/10.1016/S0140-6736(00)02786-0.
9. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252-260. https://doi.org/10.1056/NEJMsa065779.
10. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of babel bias? *J Clin Epidemiol*. 1995;48(1):159-163. http://www.ncbi.nlm.nih.gov/pubmed/7853041 Accessed September 29, 2017.
11. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet*. 1997;350(9074):326-329. https://doi.org/10.1016/S0140-6736(97)02419-7.
12. Konno K, Akasaka M, Koshida C, et al. Ignoring non-English-language studies may bias ecological meta-analyses. *Ecol Evol*. 2020. https://doi.org/10.1002/ece3.6368.
13. Mongeon P, Paul-Hus A. The journal coverage of web of science and Scopus: a comparative analysis. *Scientometrics*. 2016;106(1):213-228. https://doi.org/10.1007/s11192-015-1765-5.
14. O'Leary BC, Kvist K, Bayliss HR, et al. The reliability of evidence review methodology in environmental science and conservation. *Environ Sci Policy*. 2016;64:75-82. https://doi.org/10.1016/j.envsci.2016.06.012.
15. Woodcock P, O'Leary BC, Kaiser MJ, Pullin AS. Your evidence or mine? Systematic evaluation of reviews of marine protected area effectiveness. *Fish Fish*. 2017;18(4):668-681. https://doi.org/10.1111/faf.12196.
16. Bayliss HR, Beyer FR. Information retrieval for ecological syntheses. *Res Synth Methods*. 2015;6(2):136-148. https://doi.org/10.1002/jrsm.1120.
17. Rothstein HR, Sutton AJ, Borenstein M. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. West Sussex: John Wiley & Sons 2006.
18. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0. Cochrane. 2019. doi:https://doi.org/10.1002/9781119536604
19. Marshall IJ, Marshall R, Wallace BC, Brassey J, Thomas J. Rapid reviews may produce different results to systematic reviews: a meta-epidemiological study. *J Clin Epidemiol*. 2019;109:30-41. https://doi.org/10.1016/J.JCLINEPI.2018.12.015.
20. Haddaway NR. A call for better reporting of conservation research data for use in meta-analyses. *Conserv Biol*. 2015;29(4):1242-1245. https://doi.org/10.1111/cobi.12449.
21. Pullin AS, Stewart G. Guidelines for systematic review in conservation and environmental management. *Conserv Biol*. 2006;20(6):1647-1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x.
22. Koricheva J, Gurevitch J. Uses and misuses of meta-analysis in plant ecology. Gómez-Aparicio L. *J Ecol*. 2014;102(4):828-844. https://doi.org/10.1111/1365-2745.12224.
23. Haddaway NR, Westgate MJ. Predicting the time needed for environmental systematic reviews and systematic maps. *Conserv Biol*. 2019;33:434-443. https://doi.org/10.1111/cobi.13231.
24. Kohl C, McIntosh EJ, Unger S, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ Evid*. 2018;7(1):8. https://doi.org/10.1186/s13750-018-0115-5.
25. Farace DJ, Frantzen J. *Sixth International Conference on Grey Literature: Work on Grey in Progress*. Amsterdam: TextRelease; 2004:2005.
26. Hedges LV, Gurevitch J, Curtis PS. The meta-analysis of response ratios in experimental ecology. *Ecology*. 1999;80(4):1150. https://doi.org/10.2307/177062.
27. Borenstein M, Hedges LV, Higgins JPT, Rothstein H. *Introduction to Meta-Analysis*. West Sussex: John Wiley & Sons; 2009.
28. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;91(3). https://doi.org/10.1103/PhysRevB.91.121108.
29. R Core Team. R Core Team. *R A Lang Environ Stat Comput*. 2018. http://www.r-project.org/.
30. Harlos C, Edgell TC, Hollander J. No evidence of publication bias in climate change science. *Clim Change*. 2017;140(3–4):375-385. https://doi.org/10.1007/s10584-016-1880-1.
31. Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behav Ecol*. 2006;17:688-690. https://doi.org/10.1093/beheco/ark016.
32. Villemey A, Jeusset A, Vargac M, et al. Can linear transportation infrastructure verges constitute a habitat and/or a corridor for insects in temperate landscapes? A systematic review. *Environ Evid*. 2018;7(1):5. https://doi.org/10.1186/s13750-018-0117-3.
33. Laverick JH, Piango S, Andradi-Brown DA, et al. To what extent do mesophotic coral ecosystems and shallow reefs share species of conservation interest? A systematic review. *Environ Evid*. 2018;7(1):15. https://doi.org/10.1186/s13750-018-0127-1.
34. Batáry P, Kurucz K, Suarez-Rubio M, Chamberlain DE. Non-linearities in bird responses across urbanization gradients: a meta-analysis. *Glob Chang Biol*. 2018;24(3):1046-1054. https://doi.org/10.1111/gcb.13964.
35. Félix GF, Scholberg JMS, Clermont-Dauphin C, Cournac L, Tittonell P. Enhancing agroecosystem productivity with woody perennials in semi-arid West Africa. A meta-analysis. *Agron Sustain Dev*. 2018;38(6):57. https://doi.org/10.1007/s13593-018-0533-3.
36. Eötvös CB, Magura T, Lövei GL. A meta-analysis indicates reduced predation pressure with increasing urbanization.

Landsc Urban Plan. 2018;180:54-59. https://doi.org/10.1016/J.LANDURBPLAN.2018.08.010.

37. Banks PB, Carthey AJR, Bytheway JP. Australian native mammals recognize and respond to alien predators: a meta-analysis. *Proc R Soc B Biol Sci*. 2018;285(1885):20180857. https://doi.org/10.1098/rspb.2018.0857.
38. Darras K, Batáry P, Furnas B, et al. Comparing the sampling performance of sound recorders versus point counts in bird surveys: a meta-analysis. Willis S, ed. *J Appl Ecol*. 2018;55(6):2575-2586. https://doi.org/10.1111/1365-2664.13229.
39. Ferlian O, Eisenhauer N, Aguirrebengoa M, et al. Invasive earthworms erode soil biodiversity: a meta-analysis. *J Anim Ecol*. 2018;87(1):162-172. https://doi.org/10.1111/1365-2656.12746.
40. Gaynor KM, Hojnowski CE, Carter NH, Brashares JS. The influence of human disturbance on wildlife nocturnality. *Science*. 2018;80:1232-1235. https://doi.org/10.1126/science.aar7121.
41. Koshida C, Katayama N. Meta-analysis of the effects of rice-field abandonment on biodiversity in Japan. *Conserv Biol*. 2018;32:1392-1402. https://doi.org/10.1111/cobi.13156.
42. Wang J, Chadwick DR, Cheng Y, Yan X. Global analysis of agricultural soil denitrification in response to fertilizer nitrogen. *Sci Total Environ*. 2018;616:908-917. https://doi.org/10.1016/j.scitotenv.2017.10.229.
43. Xiao Z, Wang X, Koricheva J, et al. Earthworms affect plant growth and resistance against herbivores: a meta-analysis. *Funct Ecol*. 2018;32(1, SI):150-160. https://doi.org/10.1111/1365-2435.12969.
44. Tonietto RK, Larkin DJ. Habitat restoration benefits wild bees: a meta-analysis. Diamond S, ed. *J Appl Ecol*. 2018;55(2):582-590. https://doi.org/10.1111/1365-2664.13012.
45. Tian D, Xiang Y, Wang B, et al. Cropland abandonment enhances soil inorganic nitrogen retention and carbon stock in China: a meta-analysis. *L Degrad Dev*. 2018;29(11):3898-3906. https://doi.org/10.1002/ldr.3137.
46. Cassey P, Delean S, Lockwood JL, Sadowski JS, Blackburn TM. Dissecting the null model for biological invasions: a meta-analysis of the propagule pressure effect. *PLoS Biol*. 2018;16:e2005987. https://doi.org/10.1371/journal.pbio.2005987.
47. Tuck SL, Winqvist C, Mota F, Ahnstrom J, Turnbull LA, Bengtsson J. Land-use intensity and the effects of organic farming on biodiversity: a hierarchical meta-analysis. *J Appl Ecol*. 2014;51(3):746-755. https://doi.org/10.1111/1365-2664.12219.
48. Avery JD, Aagaard K, Burkhalter JC, Robinson OJ. Seabird longline bycatch reduction devices increase target catch while reducing bycatch: a meta-analysis. *J Nat Conserv*. 2017;38:37-45. https://doi.org/10.1016/J.JNC.2017.05.004.
49. Yan W, Zhong Y, Shangguan Z. Contrasting responses of leaf stomatal characteristics to climate change: a considerable challenge to predict carbon and water cycles. *Glob Chang Biol*. 2017;23(9):3781-3793. https://doi.org/10.1111/gcb.13654.
50. Richardson JS, Béraud S. Effects of riparian forest harvest on streams: a meta-analysis, Strecker A, ed. *J Appl Ecol*. 2014;51(6):1712-1721. https://doi.org/10.1111/1365-2664.12332.
51. Gerstner K, Dormann CF, Stein A, Manceur AM, Seppelt R. Effects of land use on plant diversity - a global meta-analysis. Bennett J, ed. *J Appl Ecol*. 2014;51(6):1690-1700. https://doi.org/10.1111/1365-2664.12329.
52. Musitelli F, Romano A, Moller AP, Ambrosini R. Effects of livestock farming on birds of rural areas in Europe. *Biodivers Conserv*. 2016;25(4):615-631. https://doi.org/10.1007/s10531-016-1087-9.
53. Fedrowitz K, Koricheva J, Baker SC, et al. Can retention forestry help conserve biodiversity? A meta-analysis. *J Appl Ecol*. 2014;51(6):1669-1679. https://doi.org/10.1111/1365-2664.12289.
54. Ren Y, Lu Y, Fu B. Quantifying the impacts of grassland restoration on biodiversity and ecosystem services in China: a meta-analysis. *Ecol Eng*. 2016;95:542-550. https://doi.org/10.1016/j.ecoleng.2016.06.082.
55. Thompson ME, Donnelly MA. Effects of secondary Forest succession on amphibians and reptiles: a review and meta-analysis. *Copeia*. 2018;106(1):10-19. https://doi.org/10.1643/CH-17-654.
56. Bai Z, Caspari T, Gonzalez MR, et al. Effects of agricultural management practices on soil quality: a review of long-term experiments for Europe and China. *Agric Ecosyst Environ*. 2018;265:1-7. https://doi.org/10.1016/J.AGEE.2018.05.028.
57. Clements JC, Darrow ES. Eating in an acidifying ocean: a quantitative review of elevated CO2 effects on the feeding rates of calcifying marine invertebrates. *Hydrobiologia*. 2018;820(1):1-21. https://doi.org/10.1007/s10750-018-3665-1.
58. Gutzat F, Dormann CF. Decaying trees improve nesting opportunities for cavity-nesting birds in temperate and boreal forests: a meta-analysis and implications for retention forestry. *Ecol Evol*. 2018;8(16):8616-8626. https://doi.org/10.1002/ece3.4245.
59. Cattano C, Claudet J, Domenici P, Milazzo M. Living in a high CO_2 world: a global meta-analysis shows multiple trait-mediated fish responses to ocean acidification. *Ecol Monogr*. 2018;88(3):320-335. https://doi.org/10.1002/ecm.1297.
60. Fabio ES, Smart LB. Effects of nitrogen fertilization in shrub willow short rotation coppice production—a quantitative review. *GCB Bioenergy*. 2018;10(8):548-564. https://doi.org/10.1111/gcbb.12507.
61. van Eeden LM, Crowther MS, Dickman CR, et al. Managing conflict between large carnivores and livestock. *Conserv Biol*. 2018;32:26-34. https://doi.org/10.1111/cobi.12959.
62. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
63. Pullin AS, Knight TM. Science informing policy—a health warning for the environment. *Environ Evid*. 2012;1(1):15. https://doi.org/10.1186/2047-2382-1-15.
64. Collins A, Miller J, Coughlin D, Kirk S. *The production of quick scoping reviews and rapid evidence assessments: A how to guide*. England: Department for Environment, Food and Rural Affairs, UK Government; 2015. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/560521/Production_of_quick_scoping_reviews_and_rapid_evidence_assessments.pdf.
65. King S, Exley J, Winpenny E, Alves L, Henham M-L, Larkin J. The health risks of bathing in recreational waters: A rapid evidence assessment of water quality and gastrointestinal illness. 2015;4. http://www.ncbi.nlm.nih.gov/pubmed/28083352%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5158261.

66. Gillingham P, Stewart J, Binney H. *The historic peat record: Implications for the restoration of blanket bog (NEER011)*. Worcester: Natural England; 2016.
67. Kareiva P, Marvier M. What is conservation science? *Bioscience*. 2012;62(11):962-969. https://doi.org/10.1525/bio.2012.62.11.5.
68. Cooke SJ, Johansson S, Andersson K, et al. Better evidence, better decisions, better environment: emergent themes from the first environmental evidence conference. *Environ Evid*. 2017;6(1):15. https://doi.org/10.1186/s13750-017-0092-0.
69. Eales J, Haddaway NR, Webb JA. Much at stake: the importance of training and capacity building for stakeholder engagement in evidence synthesis. *Environ Evid*. 2017;6(1):22. https://doi.org/10.1186/s13750-017-0101-3.
70. Pullin AS, Knight T. Time to build capacity for evidence synthesis in environmental management. *Environ Evid*. 2013;2(1):21. https://doi.org/10.1186/2047-2382-2-21.
71. Woodcock P, Pullin AS, Kaiser MJ. Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: a methodology. *Biol Conserv*. 2014;176:54-62. https://doi.org/10.1016/j.biocon.2014.04.020.
72. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google scholar, PubMed, and 26 other resources. *Res Synth Methods*. 2020;11:181-217. https://doi.org/10.1002/jrsm.1378.
73. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. Wray KB, ed. *PLoS One*. 2015;10(9):e0138237. https://doi.org/10.1371/journal.pone.0138237.
74. Livoreil B, Glanville J, Haddaway NR, et al. Systematic searching for environmental evidence using multiple tools and sources. *Environ Evid*. 2017;6(1):23. https://doi.org/10.1186/s13750-017-0099-6.
75. Haddaway NR, Macura B, Whaley P, Pullin AS. ROSES RepOrting standards for systematic evidence syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ Evid*. 2018;7(1):7. https://doi.org/10.1186/s13750-018-0121-7.
76. Valentine JC, Hedges LV, Cooper HM. *The Handbook of Research Synthesis and Meta-Analysis*. New York, NY: Russell Sage Foundation; 2009.
77. Lajeunesse MJ. On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*. 2011;92:2049-2055. https://doi.org/10.1890/11-0423.1.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Konno K, Pullin AS. Assessing the risk of bias in choice of search sources for environmental meta-analyses. *Res Syn Meth*. 2020;11:698–713. https://doi.org/10.1002/jrsm.1433