The perils of misinterpreting and misusing "publication bias" in meta-analyses: an education review

For correspondence:
jneves@fade.up.pt

José Afonso[1], Rodrigo Ramirez-Campillo[2], Filipe Manuel Clemente[3,4,5], Fionn Cléirigh Büttner[6,7], and Renato Andrade[8,9,10]

1 – Centre of Research, Education, Innovation, and Intervention in Sport (CIFI$_2$D), Faculty of Sport, University of Porto, Portugal. jneves@fade.up.pt

2 – Exercise and Rehabilitation Sciences Laboratory. School of Physical Therapy. Faculty of Rehabilitation Sciences. Universidad Andres Bello. Santiago, Chile. rodrigo.ramirez@unab.cl

All authors have read and approved this version of the manuscript. This article was last modified on May 19, 2023.

3 – Escola Superior de Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Rua Escola Industrial Comercial de Nun'Álvares, 4900-347 Viana do Castelo, Portugal. Filipeclemente@esdl.lpvc.pt

4 – Instituto de Telecomunicações, Delegação da Covilhã, Covilhã, Portugal. Filipeclemente@esdl.lpvc.pt

5 - Research Center in Sports Performance, Recreation, Innovation and Technology (SPRINT), 4960-320 Melgaço, Portugal. Filipeclemente@esdl.lpvc.pt

6 – Pragmatic Clinical Trials Unit, Centre for Evaluation & Methods, Wolfson Centre of Population Health, Queen Mary University of London, London, United Kingdom. fionn.cleirigh-buttner@ucdconnect.ie

7 – Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, United Kingdom.

8 - Porto Biomechanics Laboratory (LABIOMEP), University of Porto, Portugal. randrade@espregueira.com

9 – Clínica Espregueira - FIFA Medical Centre of Excellence, Porto, Portugal. randrade@espregueira.com

10 – Dom Henrique Research Centre, Porto, Portugal. randrade@espregueira.com

ABSTRACT

Publication bias refers to a systematic deviation from the truth in the results of a meta-analysis due to the higher likelihood for published studies to be included in meta-analyses than unpublished studies. Publication bias can lead to misleading recommendations for decision- and policy-making. In this education review, we introduce, explain, and provide solutions to the pervasive misuses and misinterpretations of publication bias that afflict evidence syntheses in sport and exercise medicine. Publication bias is more routinely assessed by visually inspecting funnel plot asymmetry, although it has been consistently deemed unreliable, leading to the development of statistical tests to assess publication bias. However, statistical tests of publication bias (i) cannot rule out alternative explanations for funnel plot asymmetry (*e.g.*, between-study heterogeneity, choice of metric, chance), and (ii) are grossly underpowered, even when using an arbitrary minimum threshold of ≥10 studies. We performed a cross-sectional, meta-research investigation of how publication bias was assessed in systematic reviews with meta-analysis published in the top two sport and exercise medicine journals throughout 2021. This analysis highlights that publication bias is frequently misused and misinterpreted, even in top tier journals. Due to conceptual and methodological problems when assessing and interpreting publication bias, preventive strategies (e.g., pre-registration, disclosing protocol deviations, and reporting all study findings regardless of direction or magnitude) offer the best and most efficient solution to mitigate the misuse and

misinterpretation of publication bias. Because true publication bias is very difficult to determine, we recommend that future publications use the term "risk of publication bias".

INTRODUCTION

Systematic reviews containing meta-analyses descriptively combine and quantitatively aggregate the results of several individual studies (if sufficiently homogeneous) to summarize bodies of research evidence [1-11]. However, the results of systematic reviews can be biased when there is selective reporting of outcomes or statistical analyses within individual research studies depending on the direction, magnitude, and perhaps most profoundly, the statistical significance of their results [2]. Outcome reporting bias occurs when authors report only a subset – typically statistically significant – of all measured outcomes in the published manuscript [4, 5, 12-23]. Reporting bias may be alternatively interpreted as an umbrella term encompassing several biases including publication bias [23-28].

Studies with statistically significant results are more likely to be published in peer-reviewed journals than studies with statistically non-significant results [3, 4, 9, 13, 14, 16, 17, 29-38]. Publication bias refers to the differential choice to publish studies depending on the nature and directionality of their results (Box 1) [3-5, 14, 16-18, 21, 23, 25, 29, 34, 38-41]. Authors regularly report study findings and submit for publication in peer-reviewed journals *only* when study findings are intriguing, impressive, and pass the threshold of statistical significance. Similarly, journal editors regularly accept submitted manuscripts for publication at peer-reviewed journals when study results will interest clinician- and scientist-readers, garner attention from media outlets, and enhance journal Impact Factor – all of which are heavily influenced by whether study results are statistically significant [3-5, 14, 16-18, 21, 23, 25, 29, 34, 38-41].

Box 1. What is publication bias?
Statistically significant studies are more likely to be published, are more easily identified in the scientific literature, and are thus are more likely to appear in meta-analyses compared with statistically non-significant studies. Because studies with larger effect sizes are more likely to be statistically significant than studies with smaller effect sizes (given the same sample size), studies included in meta-analyses tend to have systematically larger effect sizes than those that are not identified for inclusion in meta-analyses. Hence, publication bias refers to a systematic deviation

from the truth in the results of a meta-analysis due to the higher likelihood for published studies to be included in meta-analyses than unpublished studies. Assessing publication bias using visual plots or statistical tests cannot conclusively determine whether included study effect sizes and potential "missing" studies overestimate the true summary effect size in a meta-analysis. Therefore, 'risk of publication bias' or 'potential publication bias' – rather than simply 'publication bias' – is what is being assessed to judge the risk on whether publication bias is present.

This preferential publication of statistically significant studies regularly misrepresents the true distribution of individual study effect sizes and can cause misleading recommendations for decision-making and informing policy [12, 39, 42]. Published studies, which usually have larger effect sizes than unpublished studies of the same sample size, are also more easily located in the scientific literature and thus more likely to be identified by and included in meta-analyses [2]. Although outcome or analytical (non-)reporting biases can be assessed if published studies have pre-registered/published protocols [4, 14-17, 19, 21, 26, 28, 42, 43], detecting publication bias is more complicated [14].

When the literature is dominated by small studies, as is common in the fields of Sport and Exercise Medicine and Sports Science [44] (hereby referred to as sport and exercise science or SES), the results of meta-analyses are often contradicted by subsequent studies with larger sample sizes [30, 42]. Smaller studies tend to follow less rigorous methods and exhibit larger effect sizes [3, 5, 9, 27, 30, 42]. Regardless of the direction of the effects, the predominance of small studies, which experience greater effect size changes in response to systematic error (*i.e.*, biases such as flexible analytical techniques) and random error (such as dramatic sampling variation) may bias the interpretation of meta-analytic results [26, 45]. Larger studies, despite not being immune from biases of their own that can negatively influence effect size accuracy [34, 46, 47], identify a study effect size with greater precision (*i.e.*, with smaller variability).

Assessing for potential publication bias in meta-analyses judges whether a summary effect size might be biased. There are many methods to assess the potential presence of publication bias in a meta-analysis. Both graphical plots and statistical tests exist to consider (i) whether there is any evidence of publication bias, (ii) whether the summary effect size in a meta-analysis might be at least partly due to publication bias, and (iii) what influence publication bias might yield on the summary effect size [48]. This education review aims to educate the SES research communities on methods to assess potential publication bias in

meta-analyses, their use, and how the interpretation of assessment findings can influence review conclusions.

IS THERE POSSIBLE EVIDENCE OF PUBLICATION BIAS - WHAT DOES FUNNEL PLOT ASYMMETRY REALLY MEAN?

The presence of potential publication bias in a meta-analysis is usually assessed by analyzing funnel plot asymmetry [1, 2, 4, 5, 8, 9, 12, 18, 27, 29, 30, 39, 40, 42, 49, 50]. Funnel plots are scatterplots [5, 8, 12, 25, 27, 29, 51] that plot some measure of study size (*e.g.*, sample size or standard error) against a measure of study effect size [1, 4, 9, 29, 30, 40, 42, 52]. Several metrics can be used to represent study effect size, such as standardized binary effect size metrics (*e.g.,* odds ratio [OR] or risk ratio [RR]) [10, 12] or standardized continuous effect size metrics (*e.g.*, Hedges' *g*, partial eta-squared). The estimated precision of a study's effect size increases as the study sample size increases [29, 30, 42]. Funnel plots may be contour-enhanced, presenting one or more areas of statistical significance (*e.g.*, p<0.1, p<0.05 and/or p<0.01 with the 90%, 95% and/or 99% confidence intervals [CIs], respectively), and are particularly useful to identify outliers and guide sensitivity analyses [5, 17, 35, 40, 49, 52]. Funnel plots can also display a second, superimposed vertical line that signals the null effect size (*e.g.*, lnOR = 1 or Hedges' *g* = ~0) to easily identify studies with an effect size close to the null effect, which are likely to be non-significant [27].

Smaller studies produce effect sizes that vary widely, have larger standard errors, and are dispersed near the base of the funnel plot. In contrast, larger studies produce more consistent effect sizes and have a narrower distribution of effect sizes near the top of the funnel plot. This results in a smaller standard error and an inverted funnel shape [4, 5, 9, 12, 17, 24, 27, 29, 30, 39, 42, 49, 52]. Funnel plots can be created with a weighted analysis to reduce small-study effects (*i.e.*, smaller studies have more extreme effect sizes than larger studies) [31, 42]. However, larger studies with fewer outcome events will be less powerful than small studies with more outcome events so studies with distinct sample sizes can present similar standard errors [10].

Visually inspecting funnel plot asymmetry (Figure 1) is highly subjective and potentially misleading [5, 17, 30, 39, 42, 45, 50, 51]. When 41 medical researchers visually observed funnel plots with 10 studies each, accurate funnel plot asymmetry was identified for only 52.5% of plots [32]. Many statistical methods have been devised to complement funnel plot interpretation (Box 2). Although some of these tests (*e.g.,* Egger's regression) are more prone to assessing small-study effects [42], others (*e.g.,* robust Bayesian meta-analysis) attempt to assess a broader concept of publication bias [38]. Choosing the most appropriate method depends on assumptions that are often unverified by reviewers [51], and there is an ongoing debate over which tests and metrics should be used to generate the funnel plot [8, 10, 17, 18, 22-24, 33, 36, 40, 51]. Moreover, funnel plot interpretation is prone to "cherry-picking" [38] and statistical tests are each associated with different limitations. A thorough discussion of each test and the corresponding limitations falls outside the scope of our review.
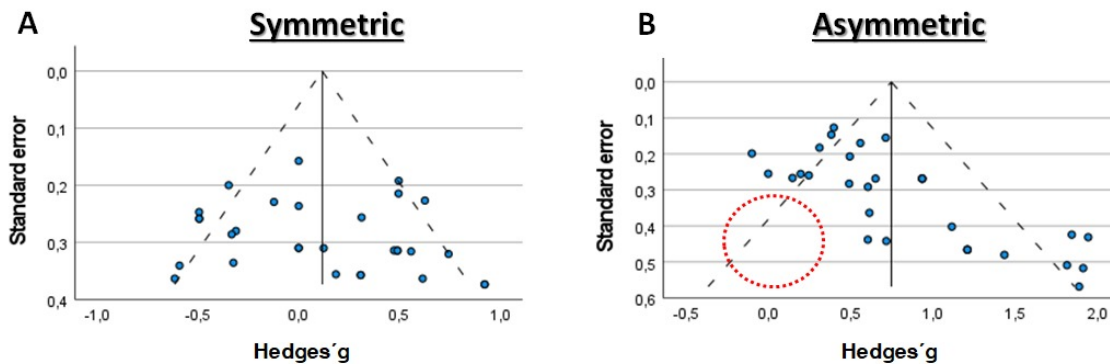

Figure 1 – Illustration on visual inspection of funnel plot to assess the risk of publication bias. The blue dots represent individual studies. *Figure 1a* depicts a symmetrical funnel plot, usually interpreted as reflecting absence of publication bias, while *Figure 1b* presents an asymmetrical funnel plot, typically interpreted as reflecting presence of publication bias. The red dotted circumference denotes an empty space where studies were expected to be present.

Box 2. Examples of plots and statistical tests to assess risk of publication bias

Is there possible evidence of publication bias?
- Forest plots
- Funnel plots
- Doi plot [53]

What is the degree of funnel plot (a)symmetry?
- Egger's weighted regression [42]
- Begg's and Mazumdar's rank correlation test [54]
- Macaskill's funnel plot regression [31]
- Peters' Regression Test [55]
- Debray's regression using estimates of asymptotic precision as study weights [45]
- Luis Furuya-Kanamori (LFK) index [53]
- Meta-regression residuals and inverse sample size [33]
- Imbalance and asymmetry distance [50]

Is the summary effect size at least partly due to publication bias?
- Rosenthal's and Orwin's Fail-safe N [37]

How much might publication bias influence the summary effect size?
- Duval and Tweedie's Trim-and-Fill method [39]
- Copas' selection model [3]
- Precision-Effect Test and Precision-Effect Estimate with Standard Errors (PET-PEESE) [22]
- Rücker's Limit Meta-Analysis Method [56]

Is the p-values distribution affecting the meta-analysis true effect and how large is this effect?
- p-curve [57-59]
- p-uniform method [60, 61]

Are p-values affecting the probability of selection of study for publication and how can it affect the meta-analysis true effect?
- Three-parameter selection model (3PSM) [60, 61]
- Fixed Weights Selection Model [60-62]

Is there evidence bias from selective (non)inclusion of results?
- Potential bias index [63, 64]
- Risk Of Bias due to Missing Evidence (ROB-ME) tool [2]

How much might both small-study effects and missing studies influence the summary effect size?
- Robust Bayesian Meta-Analysis (RoBMA) [38]

A common misconception is that funnel plots will be symmetrical when publication bias is absent and asymmetrical when publication bias is present [5, 8, 9, 29, 30, 39, 42]. For example, smaller (and potentially underpowered) studies with statistically non-significant results are likely to remain unpublished, resulting in an asymmetrical funnel plot [17, 30]. When funnel plot asymmetry is visible, statistical methods aim to provide insight into how potential publication bias may influence the summary effect size. For example, the Trim-and-Fill method (Figure 2) "balances" the funnel plot by re-iteratively removing the most extreme small studies from the positive side of the funnel plot, which generates a less biased summary effect size. To ensure a valid estimate of precision for the summary effect size, "trimmed" effect sizes are returned to the plot alongside imputed missing study effect sizes in the lower "statistically non-significant" quadrant of the funnel plot [24, 35, 39, 49, 65]. This procedure is problematic because the imputation of new "missing" effect sizes represents a daring assumption that such studies with those effect sizes might exist [34] and should be reserved only for sensitivity analyses [35, 39]. Indeed, it has been repeatedly demonstrated that if a funnel plot is asymmetrical and skewed, publication bias *may* be present [29], but it cannot be considered conclusive evidence of publication bias [17, 30, 50].
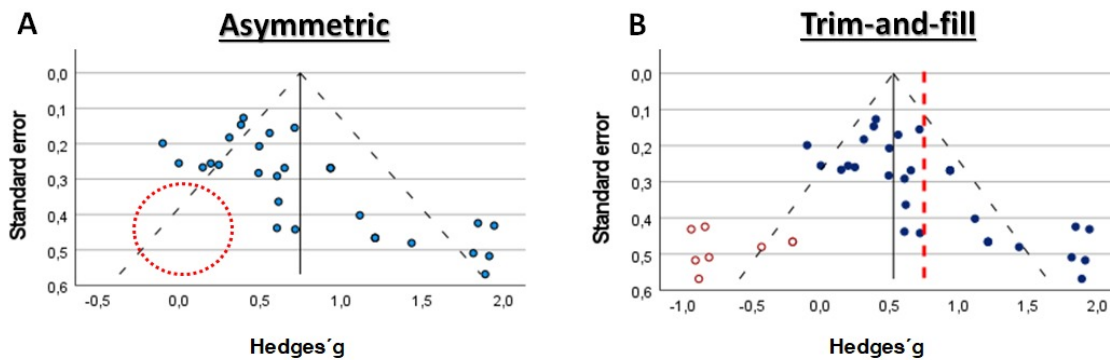

Figure 2 – Illustration on the trim-and-fill statistical test to correct the funnel plot for publication bias. The blue dots represent individual studies. *Figure 2a* shows the absence of studies on the left side of the funnel plot (red dotted circle). *Figure 2b* shows the trim-and-fill correction by adding imputed studies (red circles) and correcting the pooled Hedge's G (previous pooled effect sized in red dotted line).

There are alternative (and often co-existing) sources of funnel plot asymmetry such as between-study variation (*i.e.*, true heterogeneity), choice of effect size metric, statistical artefacts, chance, and even fraud [1, 4, 5, 8, 10, 12, 16-18, 21, 23-25, 27, 30-33, 35, 36, 40, 42, 46, 47, 49-51, 66]. A common source of true heterogeneity emerges from the type of samples in included studies, especially when studying patients at higher risk of an outcome: smaller studies may focus on this subset of the population, while larger studies may include a more diverse sample [9, 26, 30]. Small studies produce results that differ systematically (e.g., by having larger effect sizes) from results produced by larger studies (with smaller effect sizes), generating perceived funnel plot asymmetry [10, 26, 51]. Sensitivity analyses, subgroup analyses, and meta-regression models may be used to explain between-study heterogeneity [2, 5, 8, 11, 17, 30, 45]. Similarly, network meta-regression can be implemented to assess small-study bias in network meta-analysis [14]. Despite the availability and use of these techniques, many sources of true heterogeneity remain unknown [27, 32].

Any estimate of precision is subject to (i) random error due to sampling variation [23, 25, 31, 40, 67], (ii) method of outcome aggregation (*e.g.,* continuous versus binary) [68], (iii) choice of specific outcome metric (*e.g.,* risk ratio versus risk difference) [68], and (iv) funnel plot axis, which combined or in isolation may explain funnel plot asymmetry [8, 10, 17, 24, 33, 36, 42, 51, 67]. For example, the method of outcome aggregation can have a relevant effect because continuous outcomes tend to display high levels of true heterogeneity [2, 33]. The true, underlying effect size in the population, the distribution of study sample sizes, and whether one- or two-tailed tests are employed can influence funnel plot asymmetry [23, 31].

In a review of 198 published meta-analyses that examined changes in the findings of funnel plots based on definitions of precision (e.g., standard error or sample size) and choices of outcome metric (e.g., risk difference and relative risk), Tang, Liu [1] identified three general scenarios: (i) funnel plots that were always symmetrical; (ii) funnel plots that were always asymmetrical (although in some cases the *direction* of the asymmetry changed); and (iii) funnel plots that could be either symmetrical or asymmetrical depending on the definition of precision used and the outcome metric selected. Since there is no consensus on which definitions of precision and outcome metrics should be used, interpreting funnel plot asymmetry warrants caution [1].

ARE CURRENT METHODS POWERLESS?

Meta-analyses often include a limited number of studies, and many of these included studies possess small sample sizes. Consequently, statistical tests used to detect the risk of publication bias are frequently underpowered [5, 9, 18, 23, 25, 27, 31, 32, 38, 42, 45, 50, 51]. Statistical tests such as the Begg's and Mazumdar's rank correlation test [54] assess the risk of publication bias and regularly apply a statistical significance threshold of 10% (α=0.1) with 90% confidence intervals [1, 9, 30, 42]. With p<0.1, a false positive rate of 10% may be attributed to chance alone [69]. False positive findings also arise in the presence of large treatment effects, few events-per-study, or when all studies have similar sample sizes [30], and can occur regardless of the definition of precision or choice of effect measure [1]. On the other hand, tests that assess the risk of publication bias have low statistical power to detect the risk of publication bias even when it is truly present [12, 27, 39, 54]. This means there may be publication bias despite symmetrical distribution [25, 27, 32, 36, 50, 51] – and this has been observed even when 21 studies were available [31], which is above the usual *arbitrary* threshold of 10 studies that is deemed necessary to exceed in order to assess for potential publication bias [2, 5, 14, 25, 27, 35, 50, 51]. When substantial heterogeneity is present, this number should be considerably greater than 10 studies [27, 50].

Simulation analyses of rank- and regression-based methods to assess the risk of publication bias in meta-analyses demonstrated that both methods have reduced statistical power, especially when including fewer than 10 [30], or even ~20, studies [31]. Proponents of the rank correlation test, which assesses the association between study standard errors and effect estimates, proposed that this method requires a minimum of 25 studies to achieve moderate statistical power [54]. However, other authors have recommended at least 30 studies to achieve moderate statistical power [51]. Simulation studies further suggest that even when including more than 50 studies, analyses can be underpowered [45, 50]. When effect sizes are small, statistical tests can remain underpowered even when as many as 63 studies are available [31]. Despite these test-specific limitations relating to statistical power, meta-analysts regularly interpret funnel plots containing fewer than 10 studies [32, 50, 51], often at the request of uninformed reviewers. Although statistical tests are superior to visual inspection when assessing the risk of publication bias, they are underpowered given the number of studies commonly included in meta-analyses [31].

HOW IS THE RISK OF PUBLICATION BIAS BEING INTERPRETED IN SES?

It is largely unknown how the risk of publication bias is assessed – whether through visually inspecting funnel plots or interpreting relevant statistical tests - within meta-analyses published in SES. Therefore, we extracted all systematic reviews containing at least one meta-analysis published in 2021 in the *British Journal of Sports Medicine* and in *Sports Medicine* (i.e., the SES journals with the highest impact factors – 13.800 and 11.136, respectively – at the time of searching). We initially searched systematic reviews through PubMed and double-checked with direct searches on each journal's websites.

We identified 75 systematic reviews with at least one meta-analysis (24 in the *British Journal of Sports Medicine* and 51 in Sports Medicine; a complete list is provided in Electronic Supplementary Material 1). Of these, 24 reviews (32.0%) did not address the risk of publication bias. In these studies, it was unclear whether the authors had no intention to assess the risk of publication bias, if assessing the risk of publication bias was impossible due to between-study heterogeneity, or if there was an insufficient number of included studies and the review authors decided to overlook reporting this component in the manuscript. The remaining 51 systematic reviews reported an intention to assess the risk of publication bias. Fifteen systematic reviews (29.4%) used conservative terms such as "potential publication bias", "risk of bias across studies", "risk of publication bias", "risk of small study bias", or simply objective terms such as "funnel plot asymmetry". However, most systematic reviews (70.6%) used more definitive terms such as "publication bias", "small-study effects", "small-study bias", or similar expressions. Using more cautious and conservative language involving the words "risk", "potential" or "expected", is recommended.

Of 51 systematic reviews that planned to assess the risk of publication bias, five reviews did not perform this assessment due to a perceived insufficient number of included studies. Therefore, our analysis focuses on the remaining 46 systematic reviews containing at least one meta-analysis. Systematic reviews frequently used one (37.0%, k=17) or two (45.7%, k=21) method(s) to assess the risk of publication bias (Figure 3). Visually inspecting funnel plots alone (without any statistical test) was performed to assess the risk of publication bias in 13 of the 46 systematic reviews (28.3%) – this is particularly worrisome given the aforementioned limitations of such an isolated approach. Egger's test was the most applied statistical test (56.5%, k=26), whether in isolation (6.5%, k=3), combined with other statistical tests (17.4%, k=8 – most commonly with trim-and-fill), or most often in addition to visual inspection of funnel plots (32.6%, k=15). The trim-and-fill method complemented other methods in twelve systematic reviews (26.1%). Other statistical tests were less frequently used.
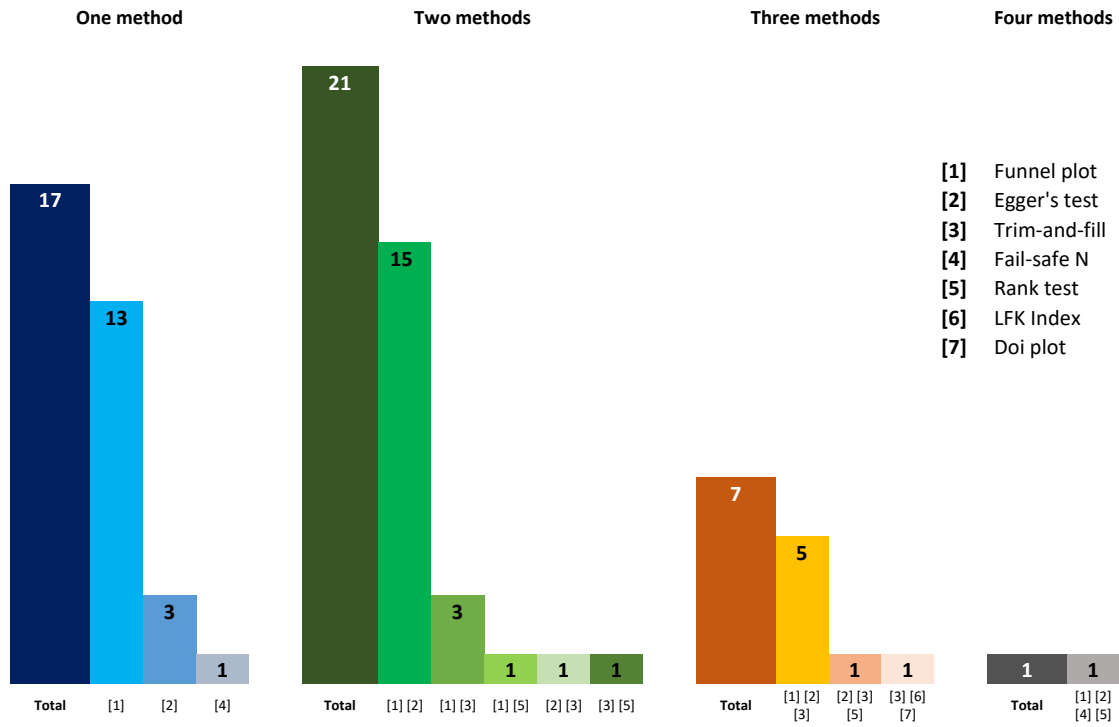
One method **Two methods** **Three methods** **Four methods**

[1] Funnel plot
[2] Egger's test
[3] Trim-and-fill
[4] Fail-safe N
[5] Rank test
[6] LFK Index
[7] Doi plot

One method: Total 17, [1] 13, [2] 3, [4] 1

Two methods: Total 21, [1] [2] 15, [1] [3] 3, [1] [5] 1, [2] [3] 1, [3] [5] 1

Three methods: Total 7, [1] [2] [3] 5, [2] [3] [5] 1, [3] [6] [7] 1

Four methods: Total 1, [1] [2] [4] [5] 1

Figure 3 – Summary of the number of methods used to assess the risk of publication bias and most used tests. LFK – Luis Furuya-Kanamori.

Only twelve of the 41 systematic reviews (26.1%) reported that a minimum of 10 studies was required to assess the risk of publication bias. Two of these systematic reviews stipulated this criterion for Egger's test but not for visual inspection of funnel plots (Table 1). Although approximately half of the systematic reviews (52.2%, k=24) assessed risk of publication bias using the "10-study minimum" rule of thumb, fourteen systematic reviews (30.4%) did not define a minimum number of studies to assess the risk of publication bias. Eighteen systematic reviews (39.1%) assessed risk of publication bias even when fewer than 10 studies were included in each meta-analysis. Two systematic reviews (4.3%) did not apply a statistical test but interpreted funnel plots with fewer than 10 studies, and two systematic reviews did not report sufficient information to determine how the risk of publication bias was assessed. Of the 46 systematic reviews published in 2021 in the two highest-ranked journals in the Sports Sciences section of Thomson Reuters Journal Citation Reports that addressed the risk of publication bias, nearly half (47.8%, k=22) assessed for risk of publication bias when fewer than 10 studies were included in eligible meta-analyses, which contradicts existing recommendations [2, 5, 14, 25, 27, 30, 35, 42, 50, 51].

Table 1 – Summary of studies assessing the risk of publication bias by pre-defining and using the "10-study minimum" heuristic.

The "10-study minimum" heuristic	k (%)
Pre-defined a minimum of 10 studies?	
Yes	10 (21.7%)
No	34 (73.9%)
Only for Egger's test	2 (4.3%)
Implemented the criterion of ≥10 studies?	
Yes	24 (52.2%)
No	18 (39.1%)
Only for Egger's test	2 (4.3%)
Unclear	2 (4.3%)

Overall, systematic review authors overlooked methodological concerns about the potential risk of publication bias, even when considering only very recent systematic reviews (containing at least one meta-analysis) published in the two highest Impact Factor SES journals. The results of the descriptive meta-research study performed within this education review suggest a misuse of currently available methods to assess the risk of publication bias. This is in line with recent observations regarding common errors in meta-analyses in the field [70]. Greater efforts should be implemented to properly educate researchers about when and how to assess the risk of publication bias reliably.

DO STATISTICAL TESTS AND FUNNEL PLOTS TRULY ASSESS THE RISK OF PUBLICATION BIAS?

When meta-analyses report on the risk of publication bias, small-study bias is being assessed [8, 17, 33, 35, 45]. Assessments of small-study effects require at least 10 studies with varied sample sizes (and at least a medium pooled sample size) [24, 27, 30, 31, 33, 51]. Even the term "small-study bias" can be misleading because heterogeneity (e.g., in interventions or sub-populations) may explain funnel plot asymmetry [32]. In summary, funnel plot asymmetry is not a measure of the risk of publication bias; using it for that sole purpose can lead to inaccurate conclusions about the presence or absence of publication bias [17, 27, 32, 34, 50, 51]. Therefore, can we assess the risk of publication bias in meta-analyses without relying on funnel plot asymmetry?

The Cochrane Bias Methods group has developed instruments that aim to assess bias due to missing evidence in meta-analysis [2, 63] and network meta-analysis [14]. The Risk of Bias due to Missing Evidence (ROB-ME) tool first assesses the risk of reporting bias by confronting published papers with pre-registered protocols that are included in systematic reviews, and then infers the risk of publication bias by scrutinizing systematic reviews' search strategy and patterns of reported results [2, 63]. The Risk of Bias due to Missing Evidence in Network meta-analysis (ROB-MEN) tool [14] starts by assessing within-study bias due to unavailable results (i.e., (non-)reporting bias); then, it assesses between-study bias due to unpublished studies (i.e., publication bias). Like instruments that appraise the risk of bias in individual studies, ROB-ME and ROB-MEN tools openly involve rulers' subjective, personal judgments [2, 14, 63], albeit using standardized and objective criteria to inform such judgments.

Key outcomes or treatment comparisons are regularly expected in research on a specific injury or condition [11, 17, 20]. For example, in rehabilitation from sports injuries, time to return to play and rate of reinjuries are commonly reported. However, it is possible that due to their informational value, authors of original studies included in systematic reviews were not interested in a subset of outcomes and therefore did not analyze them. In the previously mentioned example, not all injury rehabilitation studies report the reinjury rate. Likewise, there may be no consensus about what outcomes that should be prioritized in other research areas. For example, in the fields of SES, there is currently no consensus on how to assess fitness in children and adolescents [71, 72] or return-to-sport after a lateral ankle sprain, respectively [73].

ROB-MEN also assesses *suspected* risk of publication bias in network meta-analysis based on (i) failure to search grey literature or track study registrations for unpublished studies, (ii) novelty of a research field/topic, and (iii) previous evidence of publication bias for a given outcome or treatment comparison [14]. These assessments are complex, and the creators of ROB-MEN openly state that assessing the risk of reporting bias is more easily quantified than the risk of publication bias [14]. Indeed, "correcting for this bias is not possible without making *untestable* assumptions (p. 247)" [3], and publication bias cannot be definitively ruled out in most meta-analyses [4]. Conversely, publication bias cannot be conclusively ruled-in either, as statements of missing studies are mere assumptions [24].

The certainty of the evidence of findings in a systematic review is most commonly assessed using the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) [2]. Using this framework, the certainty of the evidence is downgraded by one level upon suspected publication bias [34]. GRADE openly states that funnel plot asymmetry is not evidence of publication bias. GRADE advises users to suspect publication bias when the evidence derives from only a few studies and/or if most studies have underlying commercial interests [34]. Although we understand the relevance of such advice, the existence of few studies may reflect the novelty of the research field instead of indicating publication bias. As many meta-analyses contain fewer than 10 studies, assessing the risk of publication bias is unreliable; thus, few systematic reviews perform this assessment. Considering the concerns associated with assessing risk of publication bias, there is a risk (albeit small) of erroneously downgrading the certainty of evidence when publication bias is not present [33]. Based on the limitations outlined throughout our educational review and the often misused and misinterpreted risk of publication bias, we question the validity of the 'publication bias' GRADE domain in many of the published meta-analyses, especially when less than 10 studies are available for the meta-analysis.

NOT DISCRIMINATING BASED ON THE SIZE OF A STUDY'S P-VALUE - HOW CAN WE REDUCE THE RISK OF PUBLICATION BIAS?

Assessing "true" publication bias is frequently not possible but a few simple solutions can aid in mitigating the risk of publication bias.

For authors of systematic reviews:

(i)   Pre-registration: Pre-register your systematic review and ensure that the review protocol is detailed. Later, in the published manuscript, disclose all deviations from the pre-registered protocol and the reasons for these alterations.

(ii)   Search strategy restrictions on date: Avoid date limits in search strategy filters, except in well-justified cases (*e.g.,* if the intervention of interest did not exist before a certain date).

(iii)   Search strategy restrictions on language: Avoid limiting a search by language of publication, even if most published papers on a topic are in English. Current automated translation technologies and the accessibility of native human translators means that restricting literature searches by the language of publication is less necessary and justifiable.

(iv)   Consider many shades of grey: Grey literature[1] (*i.e.*, conference proceedings, PhD theses, pharmaceutical study reports) should be consulted in addition to published, peer-reviewed literature [74]. Although some may argue that more reliable conclusions would be derived from peer-reviewed randomized studies, moderator analyses can be considered to compare between data derived from grey and non-grey literature.

(v)   Corresponding authors as information sources: When a study lacks summary data/information that are required for inclusion in a systematic review, reviewers should contact the corresponding authors of the study to obtain the missing information/summary data before deciding whether to exclude the study from the meta-analysis. When data are presented only in figures, free software packages allow reliable extraction of relevant data, such as WebPlotDigitizer [75] (https://automeris.io/WebPlotDigitizer/).

(vi)   An ongoing story: Living reviews continually identify and incorporate new evidence at regular intervals and can provide an ongoing interpretation of a body of evidence. By regularly updating literature searches, living reviews can circumvent problems such as the time-lag bias [5, 17, 24, 25]. However, living reviews require platforms where updates can be easily uploaded and made available. Currently, the most practical solution is publishing the initial version in a peer-reviewed journal and provide links to open-access websites hosting future updates (*e.g.,* Open Science Framework [OSF] at https://osf.io).

---

[1] Abstracts usually focus solely on a few primary outcomes [16]. Therefore, if conference abstracts are included in the eligibility criteria, the reviewers should strive to contact the authors and obtain more complete information on study outcomes.

For authors of original studies:

(vii)   Pre-registration: Pre-register/publish a study protocol before data collection. This allows systematic reviewers to identify the protocols of subsequently published studies and the protocols of studies that have yet to be published despite a sufficient timeline to allow for publication. This identification can allow for a more reliable assessment of potential publication bias. Identifying study protocols also allows review authors to compare included publications with their pre-registered protocols (to assess and detect reporting bias) [66]. Authors should explicitly state discrepancies between the registered protocol and the final, published manuscript, and any explanations for those deviations.

(viii)   Report everything: Report all study findings (including statistically non-significant results). If the manuscript is too long, consider providing additional information in supplementary material or a link to external open-access platforms (*e.g.*, OSF).

(ix)   Not everything needs to be novel – there is a lot of value in certainty from replication: Perform replication studies, which help to better assess how robust certain observed effects are (*i.e.,* if they are reproducible and replicable).

(x)   Sharing is caring: Consider making original, individual participant data available to systematic reviewers and other researchers to increase transparency and allow data sharing for inclusion in individual participant data meta-analysis.

WRAPPING UP WITH CONCLUDING REMARKS

The risk of publication bias may arise from many contributing factors and can often be challenging to evaluate and interpret. Many researchers, including those within the SES field, are prone to (i) misunderstanding the concept of publication bias in meta-analyses, (ii) misusing funnel plots and statistical tests to assess potential publication bias, and (iii) misinterpreting subsequent results. However, it is possible to optimize meta-analytic methods and refine user interpretations to conduct better research that aids clinical decision- and policy-making. Funnel plot asymmetry should not be conflated with publication bias because (i) there can be publication bias despite a symmetrical funnel plot, (ii) there can be no publication bias despite an asymmetrical funnel plot, (iii) existing statistical tests that assess the risk of publication bias are underpowered, and (iv) even the minimum threshold of 10 studies to assess the risk of publication bias can be insufficient.

There are many methods to assess the risk of publication bias but ultimately all involve considerable subjectivity. Even when suspecting a high risk of publication bias (for example, upon substantial funnel plot asymmetry from a meta-analysis containing many studies), it might be premature to dismiss the results of sound methodological studies [47, 51]. Therefore, both the researchers and the readers should refrain from conclusively stating that there is, or there is not, risk of publication bias. In line with previous recommendations [4], we endorse that systematic review authors should dilute stronger statements about the risk of publication bias, as they can be misleading. Based on the limitations outlined in our education review, we also recommend eschewing the GRADE domain of "publication bias" when judging the certainty of the evidence, especially when less than 10 studies are available for the meta-analysis.

We encourage authors of systematic reviews to strive for the most rigorous methodological standards of systematic review conduct – pre-registration, restricting search strategies by time and language, scouring grey literature, contacting corresponding authors for additional study-level data, and performing living systematic reviews – to enable the most reliable results and recommendations for practice and policy. Similarly, authors of original research studies must consider open science principles – pre-registering study intentions before data collection, reporting all outcomes, analyses, and results, performing replication studies, and making study data openly available for use by independent researchers. Only by harnessing open science principles when undertaking original research and including all relevant original research using rigorous systematic review methods, can publication bias and its burden be truly estimated.

As key messages, we highlight that (i) "publication bias" should be replaced with "risk of publication bias" and that (ii) the existing methods cannot rule out other sources of funnel plot asymmetry, and so definite statements regarding the presence or absence of risk of publication bias should be avoided.

Contributions

JA conceived the original idea and the first draft. JA, RRC, FMC, FCB and RA extensively worked on subsequent versions of the manuscript. All authors read and approved the final version of the manuscript.

Substantial contributions to conception and design: JA, RRC, FMC, FCB, RA
Acquisition of data (applicable only to meta-review section): JA, RA
Analysis and interpretation of data (applicable only to meta-review section): JA, RA
Drafting the article or revising it critically for important intellectual content: JA, RRC, FMC, FCB, RA
Final approval of the version to be published: JA, RRC, FMC, FCB, RA

Acknowledgements

Funding information

Data and Supplementary Material Accessibility

Not applicable.

REFERENCES

1. Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 2000;53(5):477-84. doi: https://doi.org/10.1016/s0895-4356(99)00204-8.

2. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions (2nd Ed.). Chichester (UK): John Wiley & Sons; 2019.

3. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. Biostatistics. 2000;1(3):247-62. doi: https://doi.org/10.1093/biostatistics/1.3.247.

4. Sinha MK, Montori VM. Reporting bias and other biases affecting systematic reviews and meta-analyses: a methodological commentary. Expert Review of Pharmacoeconomics & Outcomes Research. 2006;6(5):603-11. doi: https://doi.org/10.1586/14737167.6.5.603.

5. Godavitarne C, Robertson A, Ricketts DM, Rogers BA. Understanding and interpreting funnel plots for the clinician. Br J Hosp Med. 2018;79(10):578-83. doi: https://doi.org/10.12968/hmed.2018.79.10.578.

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

7. Tod D. Conducting systematic reviews in sport, exercise, and physical activity. Palgrave Macmillan, Springer Nature Switzerland AG; 2019.

8. Zwetsloot PP, Van Der Naald M, Sena ES, Howells DW, IntHout J, De Groot JA, et al. Standardized mean differences cause funnel plot distortion in publication bias assessments. Elife. 2017;6:e24260. doi: https://doi.org/10.7554/eLife.24260.

9. Biljana M, Jelena M, Branislav J, Milorad R. Bias in meta-analysis and funnel plot asymmetry. Stud Health Technol Inform. 1999;68:323-8.

10. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol. 2001;54(10):1046-55. doi: https://doi.org/10.1016/s0895-4356(01)00377-8.

11. Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials. 2010;11(1):52. doi: https://doi.org/10.1186/1745-6215-11-52.

12. Liu JL. The role of the funnel plot in detecting publication and related biases in meta-analysis. Evid Based Dent. 2011;12(4):121-2. doi: https://doi.org/10.1038/sj.ebd.6400831.

13. Howland RH. What you see depends on where you're looking and how you look at it: publication bias and outcome reporting bias. J Psychosoc Nurs Ment Health Serv. 2011;49(8):13-5. doi: https://doi.org/10.3928/02793695-20110705-06.

14. Chiocchia V, Nikolakopoulou A, Higgins JPT, Page MJ, Papakonstantinou T, Cipriani A, et al. ROB-MEN: a tool to assess risk of bias due to missing evidence in network meta-analysis. BMC Med. 2021;19(1):Article 304. doi: https://doi.org/10.1186/s12916-021-02166-3.

15. Dal-Ré R, Caplan AL. Journal editors impasse with outcome reporting bias. European Journal of Clinical Investigation. 2015;45(9):895-8. doi: https://doi.org/10.1111/eci.12484.

16. Schmid CH. Outcome Reporting Bias: A Pervasive Problem in Published Meta-analyses. Am J Kidney Dis. 2017;69(2):172-4. doi: https://doi.org/10.1053/j.ajkd.2016.11.003.

17. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Res Synth Methods. 2021;12(2):248-59. doi: https://doi.org/10.1002/jrsm.1468.

18. Pustejovsky JE, Rodgers MA. Testing for funnel plot asymmetry of standardized mean differences. Res Synth Methods. 2019;10(1):57-71. doi: https://doi.org/10.1002/jrsm.1332.

19. Ioannidis JP, Caplan AL, Dal-Ré R. Outcome reporting bias in clinical trials: why monitoring matters. BMJ. 2017;356:j408. doi: https://doi.org/10.1136/bmj.j408.

20. Kirkham JJ, Altman DG, Chan A-W, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. BMJ. 2018;362:k3802. doi: https://doi.org/10.1136/bmj.k3802.

21. Pirracchio R, Resche-Rigon M, Chevret S, Journois D. Do simple screening statistical tools help to detect reporting bias? Annals of Intensive Care. 2013;3(1):29. doi: https://doi.org/10.1186/2110-5820-3-29.

22. Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. Res Synth Methods. 2014;5(1):60-78. doi: https://doi.org/10.1002/jrsm.1095.

23. Stanley TD. Limitations of PET-PEESE and Other Meta-Analysis Methods. Social Psychological and Personality Science. 2017;8(5):581-91. doi: https://doi.org/10.1177/1948550617693062.

24. Sedgwick P, Marston L. How to read a funnel plot in a meta-analysis. BMJ. 2015;351:h4718. doi: https://doi.org/10.1136/bmj.h4718.

25. Sedgwick P. Meta-analysis: testing for reporting bias. BMJ. 2015;350:g7857. doi: https://doi.org/10.1136/bmj.g7857.

26. Spineli LM, Pandis N. Reporting bias: Notion, many faces and implications. Am J Orthod Dentofacial Orthop. 2021;159(1):136-8. doi: https://doi.org/10.1016/j.ajodo.2020.10.010.

27. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002. doi: https://doi.org/10.1136/bmj.d4002.

28. McGauran N, Wieseler B, Kreis J, Schüler Y-B, Kölsch H, Kaiser T. Reporting bias in medical research - a narrative review. Trials. 2010;11(1):37. doi: https://doi.org/10.1186/1745-6215-11-37.

29. Egger M, Smith GD. Bias in location and selection of studies. BMJ. 1998;316(7124):61-6. doi: https://doi.org/10.1136/bmj.316.7124.61.

30. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000;53(11):1119-29. doi: https://doi.org/10.1016/s0895-4356(00)00242-0.

31. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med. 2001;20(4):641-54. doi: https://doi.org/10.1002/sim.698.

32. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58(9):894-901. doi: https://doi.org/10.1016/j.jclinepi.2005.01.006.

33. Doleman B, Freeman SC, Lund JN, Williams JP, Sutton AJ. Funnel plots may show asymmetry in the absence of publication bias with continuous outcomes dependent on baseline risk: presentation of a new publication bias test. Res Synth Methods. 2020;11(4):522-34. doi: https://doi.org/10.1002/jrsm.1414.

34. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. J Clin Epidemiol. 2011;64(12):1277-82. doi: https://doi.org/10.1016/j.jclinepi.2011.01.011.

35. Mavridis D, Salanti G. How to assess publication bias: funnel plot, trim-and-fill method and selection models. Evid Based Ment Health. 2014;17(1):30. doi: https://doi.org/10.1136/eb-2013-101699.

36. Hunter JP, Saratzis A, Sutton AJ, Boucher RH, Sayers RD, Bown MJ. In meta-analyses of proportion studies, funnel plots were found to be an inaccurate method of assessing publication bias. J Clin Epidemiol. 2014;67(8):897-903. doi: https://doi.org/10.1016/j.jclinepi.2014.03.003.

37. Rosenthal R. The file drawer problem and tolerance for null results. Psychological Bulletin. 1979;86(3):638-41. doi: https://doi.org/10.1037/0033-2909.86.3.638.

38. Bartoš F, Maier M, Quintana DS, Wagenmakers E. Adjusting for Publication Bias in JASP and R: Selection Models, PET-PEESE, and Robust Bayesian Meta-Analysis. Advances in Methods and Practices in Psychological Science. Ahead of Print. doi: https://doi.org/10.1177/25152459221109259.

39. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-63. doi: https://doi.org/10.1111/j.0006-341x.2000.00455.x.

40. Lin L. Graphical augmentations to sample-size-based funnel plot in meta-analysis. Res Synth Methods. 2019;10(3):376-88. doi: https://doi.org/10.1002/jrsm.1340.

41. Ioannidis JP, Munafò MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cogn Sci. 2014;18(5):235-41. doi: https://doi.org/10.1016/j.tics.2014.02.010.

42. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34. doi: https://doi.org/10.1136/bmj.315.7109.629.

43. van der Steen J, ter Riet G, van den Bogert C, Bouter L. Causes of reporting bias: a theoretical framework [version 2; peer review: 2 approved]. F1000Res. 2019;8(280). doi: https://doi.org/10.12688/f1000research.18310.2.

44. Abt G, Boreham C, Davison G, Jackson R, Nevill A, Wallace E, et al. Power, precision, and sample size estimation in sport and exercise science research. J Sports Sci. 2020;38(17):1933-5. doi: https://doi.org/10.1080/02640414.2020.1776002.

45. Debray TPA, Moons KGM, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. Res Synth Methods. 2018;9(1):41-50. doi: https://doi.org/10.1002/jrsm.1266.

46. Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity. BMJ. 1998;316(7129):469; author reply 70-1.

47. Vandenbroucke JP. Bias in meta-analysis detected by a simple, graphical test. Experts' views are still needed. BMJ. 1998;316(7129):469-70; author reply 70-1.

48. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Publication bias. In: Borenstein LV, Hedges JPT, Higgins JP, Rothstein HR, editors. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009. p. 277-92.

49. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008;61(10):991-6. doi: https://doi.org/10.1016/j.jclinepi.2007.11.010.

50. Simmonds M. Quantifying the risk of error when interpreting funnel plots. Syst Rev. 2015;4:24. doi: https://doi.org/10.1186/s13643-015-0004-8.

51. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ. 2006;333(7568):597-600. doi: https://doi.org/10.1136/bmj.333.7568.597.

52. van der Willik EM, van Zwet EW, Hoekstra T, van Ittersum FJ, Hemmelder MH, Zoccali C, et al. Funnel plots of patient-reported outcomes to evaluate health-care quality: Basic principles, pitfalls and considerations. Nephrology (Carlton). 2021;26(2):95-104. doi: https://doi.org/10.1111/nep.13761.

53. Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. JBI Evidence Implementation. 2018;16(4):195-203. doi: https://doi.org/10.1097/XEB.0000000000000141.

54. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. Biometrics. 1994;50(4):1088-101. doi: https://doi.org/10.2307/2533446.

55. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. Jama. 2006;295(6):676-80. doi: https://doi.org/10.1001/jama.295.6.676.

56. Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics. 2011;12(1):122-42. doi: https://doi.org/10.1093/biostatistics/kxq046.

57. Simonsohn U, Nelson LD, Simmons JP. p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. Perspect Psychol Sci. 2014;9(6):666-81. doi: https://doi.org/10.1177/1745691614553988.

58. Simonsohn U, Nelson LD, Simmons JP. P-curve: a key to the file-drawer. J Exp Psychol Gen. 2014;143(2):534-47. doi: https://doi.org/10.1037/a0033242.

59. Simonsohn U, Simmons JP, Nelson LD. Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). J Exp Psychol Gen. 2015;144(6):1146-52. doi: https://doi.org/10.1037/xge0000104.

60. van Aert RCM, Wicherts JM, van Assen MALM. Conducting Meta-Analyses Based on p Values: Reservations and Recommendations for Applying p-Uniform and p-Curve. Perspectives on Psychological Science. 2016;11(5):713-29. doi: https://doi.org/10.1177/1745691616650874.

61. van Assen MALM, van Aert RCM, Wicherts JM. Meta-analysis using effect size distributions of only statistically significant studies. Psychol Methods. 2015;20(3):293-309. doi: https://doi.org/10.1037/met0000025.

62. Vevea JL, Woods CM. Publication bias in research synthesis: sensitivity analysis using a priori weight functions. Psychol Methods. 2005;10(4):428-43. doi: https://doi.org/10.1037/1082-989x.10.4.428.

63. Page MJ, Bero L, Kroeger CM, Dai Z, McDonald S, Forbes A, et al. Investigation of Risk Of Bias due to Unreported and SelecTively included results in meta-analyses of nutrition research: the ROBUST study protocol. F1000Res. 2019;8:1760. doi: https://doi.org/10.12688/f1000research.20726.2.

64. Page MJ, McKenzie JE, Green SE, Forbes AB. An empirical investigation of the potential impact of selective inclusion of results in systematic reviews of interventions: study protocol. Syst Rev. 2013;2(1):21. doi: https://doi.org/10.1186/2046-4053-2-21.

65. Kaneda H. Asymmetry in funnel plot. Am J Cardiol. 2011;108(2):330-1. doi: https://doi.org/10.1016/j.amjcard.2011.04.004.

66. Büttner F, Toomey E, McClean S, Roe M, Delahunt E. Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. Br J Sports Med. 2020;54(22):1365. doi: https://doi.org/10.1136/bjsports-2019-101863.

67. Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. BMJ. 1998;316(7129):470; author reply -1.

68. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. J Clin Epidemiol. 2017;86:39-50. doi: https://doi.org/10.1016/j.jclinepi.2017.05.007.

69. Seagroatt V, Stratton I. Bias in meta-analysis detected by a simple, graphical test. Test had 10% false positive rate. BMJ. 1998;316(7129):470; author reply -1.

70. Kadlec D, Sainani KL, Nimphius S. With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. Sports Med. 2022. doi: 10.1007/s40279-022-01766-0.

71. Ruiz JR, Ortega FB, Gutierrez A, Meusel D, Sjöström M, Castillo MJ. Health-related fitness assessment in childhood and adolescence: a European approach based on the AVENA, EYHS and HELENA studies. Journal of Public Health. 2006;14(5):269-77. doi: https://doi.org/10.1007/s10389-006-0059-z.

72. Ruiz JR, Castro-Piñero J, España-Romero V, Artero EG, Ortega FB, Cuenca MM, et al. Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. Br J Sports Med. 2011;45(6):518-24. doi: https://doi.org/10.1136/bjsm.2010.075341.

73. Wikstrom EA, Mueller C, Cain MS. Lack of Consensus on Return-to-Sport Criteria Following Lateral Ankle Sprain: A Systematic Review of Expert Opinions. J Sport Rehabil. 2020;29(2):231-7. doi: https://doi.org/10.1123/jsr.2019-0038.

74. Winters M, Weir A. Grey matters; on the importance of publication bias in systematic reviews. Br J Sports Med. 2017;51(6):488. doi: https://doi.org/10.1136/bjsports-2016-096679.

75. Rohatgi A. WebPlotDigitizer, version 4.5. Pacifica, California2021.