The use and misuse of methods for publication bias assessment and adjustment in meta-analyses of psychotherapeutic interventions: A systematic survey of the literature

Espen Aas-Hansen	Susanne Røst	Granit Syla
Stud.psychol.	Stud.psychol.	Stud.psychol.
The Arctic University of Norway, Tromsø	The Arctic University of Norway, Tromsø	The Arctic University of Norway, Tromsø

PSY-2901: Hovedoppgave. 14.12.2021, University of Tromsø

Preface

As all three authors will graduate as clinical psychologists in 2022, we are eager about what our knowledge and practice is based on. Therefore, publication bias was an appropriate subject, as it could pose a threat to the validity of scientific findings.

Our aim with this dissertation is to inform and guide other researchers who are about to conduct a study, and in particular psychologists who seek to inform their clinical practice by means of reading meta-analytic reviews of psychotherapeutic interventions.

All authors contributed to the practical aspects of our study, which included searching for and reading literature, abstract screening, coding and reanalyzing. All authors also contributed to the writing of preface, abstract, introduction, method, results and discussion.

We followed author guidelines provided by BMC Research Integrity and Peer Review: https://researchintegrityjournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript.

We also wish to thank our supervisor Torstein Låg for guidance during the process of writing this dissertation.

Abstract

Publication bias poses a threat to the validity of meta-analytic reviews, as it can lead to summary effect size estimates becoming inflated. Meta-analysts are advised to utilize multiple methods for detecting and controlling for publication bias. Our study aims to examine which and how many methods meta-analysts of psychotherapeutic interventions for depression, anxiety and PTSD utilize to identify and correct for publication bias, and to which extent they detect it. Additionally, we aim to provide some indication of the degree to which publication bias has (or has not) influenced meta-analytic estimates in this field, by reanalyzing meta-analyses for which study level data are available. 86 meta-analyses were included in our sample, and 37 meta-analyses also met the eligibility criteria for reanalysis. Findings demonstrate that 66 of 86 (76,7%) included meta-analyses utilized at least 1 publication bias method. 32 of 86 (37%) of the included meta-analyses utilized at least three publication bias methods. None of the included meta-analyses utilized a selection model approach. The funnel plot asymmetry tests varied from detecting publication bias on ~20% (Egger's regression) to ~65% (trim-and-fill). The results from reanalyzes of study-level data indicates some inflation of effect size estimates, although the adjusted results generally do not considerably change the overall conclusions of these meta-analyses. Our overall findings indicate some degree of publication bias, that could go undetected because some meta-analysts do not sufficiently adhere to recommendations regarding publication bias methods.

Keywords: publication bias, psychotherapeutic interventions, survey, psychotherapy, publication bias methods, reanalysis

Publication bias, the tendency for the availability of research to depend on the results (1), is a concern because it can result in erroneous estimates of intervention effects in meta-analyses (2). Despite the great variety of available techniques for detecting and adjusting for publication bias in meta-analytic datasets, it is a challenging problem to overcome. The purpose of this study was to investigate how meta-analysts of psychotherapy interventions for depression, anxiety and post-traumatic stress disorder (PTSD) manage publication bias. We assessed this by examining the number and type of publication bias methods employed, as well as whether they are utilized in accordance with guidelines derived from the literature on publication bias. Additionally, where adequate study-level statistics were reported, we reanalyzed their data, applying several statistical methods to detect and correct for publication bias, thus following the general advice of using multiple methods.

Defining publication bias

The term *publication bias* has been defined in a variety of ways. According to Hussain et al., publication bias was first defined by Begg in 1985 as the phenomenon where studies with stronger observed efficacy are more likely to be reported than studies with an average effect or no effect (3). Additionally, the Cochrane Handbook defines publication bias as the practice of publishing or withholding research findings based on the nature and direction of the results (4). According to Vevea et al., publication bias is when the unpublished literature systematically differs from the published literature, because selectivity may contribute to the decision of what gets published (1). Similarly, Rothstein et al, refers to the term as whenever the research that appears in the published literature is systematically unrepresentative of the population of completed studies (5). The definitions by Begg and Cochrane Handbook refer to publication bias as something that *causes* differences between the published and unpublished literature, namely the suppression of studies with certain kinds of results. Instead of referring to publication bias as the cause of these differences, Vevea et al. and Rothstein et al. refers to publication bias as the differences themselves. Therefore, the latter definitions are broader, and do not limit what the cause of publication bias might be. Even though the definitions differ, they either refer to something *causing* differences between the published and unpublished literature or refer to the differences themselves. The result either way is that the literature becomes biased.

The differences between the definitions may reflect that the subject matter is broad and encompasses different levels of analysis. Several terms are related to publication bias. These terms refer to phenomena that may comprise some aspects of publication bias. Examples include outcome bias, language bias, availability bias and other biases (5). These terms often highlight specific mechanisms or specific aspects that underlie publication bias. *Dissemination bias* has been suggested as an alternative comprehensive framework to encompass all these related biases (6). However, our impression is that publication bias is a more frequently used phrase, and we will adopt this term throughout this survey. Specifically, we will use the term broadly, to refer to the phenomenon of bias in the availability of research results in published literature, and not the mechanisms that lead to bias. Instead, we will examine some of the mechanisms that lead to publication bias in the following section.

Why does publication bias occur?

It is important to recognize that numerous factors can contribute to differences between the published and the unpublished literature. In this section, we will briefly describe mechanisms which potentially can contribute to publication bias. It is important to note, that the issue is complex, and that a full review of the causes of publication bias is beyond the scope of this survey. By exploring some of the mechanisms in greater detail, we may gain a better understanding of the purpose for the statistical publication bias methods that we will examine and apply in our study. Furthermore, the mechanism may also help to clarify what publication bias is.

A quite prominent cause of publication bias is the suppression of nonsignificant or uninteresting research findings. For instance, studies with significant results, that demonstrates the effectiveness of an intervention, have a higher probability of being available than studies with non-significant findings (7). More specifically, a study's statistical significance and effect magnitude may influence the likelihood of whether a study is published (8). As a result, systematic differences in published and non-published studies occur. This indicates that the decision of researchers to submit or not submit their work may give rise to publication bias (9).

There could be multiple reasons why researchers do not submit their papers. One is that researchers do not perceive their findings to be important enough. Easterbrook found that there was an increased likelihood for studies to be published where the researchers rated the importance of the results highly (10). Similarly, Dickersin’s findings support the role of the researcher's interest. She demonstrated that one of the major reasons authors stated for not publishing a study, was lack of interest (11). Moreover, Koricheva found that studies with significant findings are more likely to be published in journals with high impact factor and may therefore be more accessible to other researchers (12). In addition, she found no evidence of higher rejection rates in high impact journals, and concluded that the researchers did not submit nonsignificant studies as much to the high impact journals (12). This could be due to not deeming the results interesting enough to submit their papers to the high impact journals. Koricheva’s findings also illustrates that the main reason given by researchers for not submitting their papers, was lack of time (12). A possible explanation for this, is that if researchers are interested in their study, or find it important, then they might have been motivated to submit their paper even if they are pressed for time.

Another possible cause of lower publication rates for nonsignificant studies or studies perceived as unimportant, are the incentives surrounding academic publishing. Researchers' may perceive their careers to be contingent on presenting significant results, and they may feel under pressure to do so (13). Research in the United States indicates that researchers at institutions are more likely to publish results that support their own hypotheses (14). This could be explained by that high-stress conditions or environments attract better researchers, capable of developing better hypotheses. Alternatively, academic incentive structures may push researchers to engage in questionable research practices, which include p-hacking and reporting bias (13).

Economical sponsorship of a study may contribute to the likelihood of publishing results (9). For instance, pharmaceutical companies may discourage publication of drug-studies that show no effects (15, 16). One study found that research funded by pharmaceutical companies was less likely to be presented or published compared to research funded by other sources (17). Of studies sponsored by the pharmaceutical industry, up to 89% favored a new therapy or drug, compared to 61% of studies funded by other sources (8). These findings indicate that the funding of drug studies by pharmaceutical industry may be associated with outcomes that favored the sponsor's interests (18). Researchers may feel pressured to present their findings as more favorable, which could lead to bias in study design, outcomes and reporting.

Researchers' design choices can also lead to publication bias. The phenomenon of “p-hacking” is an example of this. P-hacking refers to researchers' choice to selectively report significant p-values. In practice, researchers may do several analyses, and then only report analyses with statistically significant results. In addition, researchers may also quit collecting data as soon as they achieved significant results, and not adhere to the predetermined sample size. On the other hand, excluding or including variables based on the effect the variables would have on the p-value is also a form of p-hacking (13). All of these practices may contribute to an overrepresentation of positive results, and to effect sizes becoming inflated. It is debatable to what extent the researchers' choices are deliberate or unintentional, when data, data collection, or statistical tests are manipulated to generate a statistically significant result (19-21). Regardless, p-hacking will lead to the published results being different from the unpublished results.

Small sample sizes in studies can exacerbate the problems caused by publication bias. Small sample sizes lead to a lack of statistical power, denoting that there is a lower probability for obtaining statistically significant results, even if there is a true effect in the population sampled. Significance may therefore only be obtained when chance amplifies the true effect. Although low sample size/low power will generate fewer significant results, the combination with suppression of unsignificant results will produce publication bias (9). Put differently, because sample statistics of studies with small samples tend to vary a lot around the true population parameter by chance, and the significant results get reported more often, it may appear that there is a large true effect, when it could be only an artifact of low power and bias. Contrary, in studies with better power, results would vary less, and it would therefore be less room for publication bias. Similarly, Ioannidis argues with a statistical approach that most research findings may be due to chance, and that the overall findings from some research fields may only be a measure of different kinds of bias (22).

Finally, another mechanism that could cause publication bias, is that journal editors may be biased in favor of positive findings. The research on this subject is somewhat unclear, and several studies indicates that editors are unbiased (23-25). However, Senn argues that evidence against editor's bias is flawed, and it is plausible that editors are biased (26). Regardless, editorial rejection is not stated as a common reason by researchers for why their studies remain unpublished. However, researchers may withhold the publication process because they anticipate rejection if they have unimportant results (25). Thus, even if bias in editors' reviews might be small or absent, the assumption from authors that it is real can still cause bias.

Publication bias in meta-analyses

To fully grasp the ramifications of publication bias in meta-analyses, it is necessary to examine what meta-analyses are and how they function. Meta-analysis is a research method that enable us to synthesize the findings of several primary studies. When a meta-analysis is conducted in a systematic review, researchers can examine if the results from the primary studies are robust across the various populations studied. Furthermore, if certain conditions are met, it enables researchers to estimate the weighted average effect more precisely than inspecting the primary studies alone (2). In fact, an important main objective of many meta-analyses seems to be indicating the magnitude of effects in interventions. Systematic reviews with a meta-analysis are highly valued in many research fields, due to their ability to produce robust and precise results. Decision makers can benefit from meta-analyses, by gaining a clearer understanding of the average effect and variability, potentially leading to better informed decision making regarding for example policies (27). As meta-analyses are capable of synthesizing large bodies of research, they also help us infer conclusions about the effectiveness of interventions, which could then guide evidence-based clinical psychology practice.

According to the Criteria for Evaluating Treatment Guidelines, “The evaluation of treatment efficacy places greatest emphasis on evidence derived from sophisticated empirical methodologies, including quasi experiments and randomized controlled experiments or their logical equivalents” (28 p.1054). As a result, consumers of research may assume that meta-analyses based on randomized controlled trials (RCT’s), and quasi-experimental trials are a fool proof method for generating exact and valid data. Several criteria, however, should be satisfied before one can rely on meta-analyses, even when they are based on RCT's and well-controlled quasi-experimental trials. As mentioned earlier, John P. A. Ioannidis claims that most published research findings are false. He also states that the probability that research findings represent a true effect will depend on bias, statistical power and the ratio of true to not-true relationships in the field. He argues with a hypothetical example that findings from confirmatory meta-analysis with good quality RCTs could have around 15 % chance of being false. If the meta-analysis instead is based on small inconclusive studies, the risk that the finding is false could increase to around 59 %. Because meta-analyses generally have high statistical power, bias and the ratio of true to non-true relationships in the field are likely to be the most critical factors for the validity of meta-analytic results (22).

Consequences of publication bias

Meta-analysis has been hailed as a more trustworthy method of evaluating research material than, for instance, standard narrative reviews (29). However, regardless of how thorough and extensive a meta-analytic review is in other respects, the validity of a meta-analytic review is threatened if the sample of articles conducted for reviews are biased (5).

When studies with statistically significant outcomes are more likely to be published than studies with non-significant or unfavorable results, publication bias arises (30). With such an impact on meta-analyses, this is not a hypothetical issue (5). Prior research demonstrates that publication bias affects all kind of research (10) and it has been observed in many scientific disciplines, including psychology (31). A study by Fanelli indicates that over 90% of published studies report statistically significant results in psychology (32). This finding is interesting, given that publication bias is related to statistically significant results.

Schäfer and Schwars suggest that publication bias cause a dramatic inflation in published effects (33). As the primary source of knowledge for clinical and health policy decisions are published papers and the literature, exaggerated effect sizes in meta-analysis could impact clinical and health policy decision making (34). Consequently, biased literature can lead to damaging the integrity of knowledge (35).

Taking everything into account, inflation of meta-analytic effect size estimates seems to be a main consequence of publication bias. In turn, the consequences of exaggerated effect sizes could be many. We wonder if and how the psychotherapeutic interventions are influenced by publication bias. Inflated effect size estimates in meta-analysis may also signify an overestimation of treatment effects in clinical settings (36). The consequences of overestimated treatment effects could possibly impact the patient negatively (10, 30, 37). Prior research by Chalmers in 2001 and Rennie in 1997, demonstrated how publication bias could be related to patient damage in medicine, as cited by Hussain, Hassali & Babar (3). Likewise, the American Psychological Association (APA) also mentions in its document (Criteria for Evaluating Treatment Guidelines) that several psychological methods that initially appeared to be beneficial and generally accepted were later discovered to be ineffective or even detrimental to patients (28). It is important to highlight that APA's document makes no explicit reference to publication bias as being the reason. However, one could speculate that publication bias could have been a contributor to this consequence. For instance, prior research demonstrates that publication bias has been found in many intervention fields, such as depression treatment (38). Driessen et al., for example, concluded in their analysis that the efficiency of psychological interventions for adult depression has been overestimated in the published literature, just as it has been for pharmacotherapy (39).

Psychotherapy practice that is evidence-based is founded on research, including meta-analyses. This will be problematic, if publication bias jeopardizes the validity of research, and inflates meta-analytic estimates. Fortunately, there are several ways in which meta-analysts can uncover and possibly adjust for the influence of publication bias.

Statistical methods for dealing with publication bias

Even so, dealing with publication bias in meta-analyses is no simple matter, and probably because it is a complex issue, there is no single, easy-to follow recipe meta-analysts can follow to guide their efforts. However, there are several adequate sources on the subject. For example, Vevea et al. provides an informal introduction to the subject with multiple recommendations for meta-analysists (1). Additionally, several other sources provide recommendations on publication bias methods based on research. In the following sections, our goal is to summarize some of these recommendations.

An important piece of advice that is stressed by experts is that methods for dealing with publication bias are flawed, and therefore meta-analysts should not rely on only 1 method. Publication bias methods are based on assumptions that may be too simple to capture the complexity of real meta-analytic datasets. For example, it is unlikely that publication bias can be completely captured by the relationship between standard error and the effect size, as Eggers's regression test measures (1). Furthermore, the consensus between different publication bias methods is weak or moderate (40). Marks-Angelin & Chen also reports that no single publication bias method is ideal for every condition, and several different variables can affect which method that should be applied (41). Hence the reasons mentioned above, an important advice for addressing publication bias in meta-analyses is to conduct multiple methods instead of relying on the results of a single method (1, 41-43).

Like the advice to utilize several methods, it is also recommended to follow a sensitivity approach with publication bias methods (1, 44-46). This implies utilizing the different publication bias methods to test if the summary results are sensitive to different selection mechanisms that may cause publication bias (46).

While there are numerous publication bias methods which in conjunction can help control publication bias, not all of them are equally effective. Carter et al. state that while correcting for bias, only methods that can expect to perform reasonably well should be used (44). Consequently, it is important for meta-analysts to have some conception of how the different publication bias methods perform, along with strengths, limitations and assumptions of the methods. Therefore, in the following paragraphs, we will briefly summarize advice and information given for some of the most popular and cited publication bias methods.

The funnel plot

The funnel plot (47) is a very popular method for identifying publication bias (48, 49). The funnel plot is essentially a scatterplot of each study's effect size against some measure of precision (usually the inverse of the study's standard error of the effect size). The meta-analyst(s) often visually inspects the funnel plot for asymmetry. If the funnel plot looks asymmetrical, then that is indicative of publication bias. However, causes other than publication bias could also lead to funnel plot asymmetry (1). Other problems with visual inspection of the funnel plot, is subjectivity (1, 2), and that it can be misleading, especially in the presence of large between-study heterogeneity (41). It can still be applied as an exploratory tool (2, 41). Also, contour enhanced funnel plots can help with the subjective nature of the visual inspection (1).

Egger's regression

Egger's regression is a quantitative method for evaluating funnel plot asymmetry (50). It has many variants, but they are similar enough to often be discussed as a group, as in Vevea et al. (1). One of the weaknesses of Egger's regression is low power for detecting publication bias when the study sample is small, hence it is recommended to avoid utilizing Egger's regression when there are fewer than 10 studies in the meta-analysis (51). Egger's regression is also less effective when the studies have little variation in study size and standard error (1, 51). If the meta-analysis has systematic heterogeneity, the researcher can and should control for this when utilizing Egger's regression, either by including study level covariates in the regression model or splitting the data in different subgroups to accommodate for the heterogeneity (1). Egger's regression cannot, however, accommodate for random heterogeneity or estimate variance (1).

When conducting a meta-analysis with odds ratio, the normal variant of Egger's regression is advised against, since odds ratios and their standard error naturally correlates, confounding the regression (1, 40, 52). This is, albeit to a smaller degree, is also true for standardized mean differences, and for other statistical tests of funnel plot asymmetry (49). This latter fact is a very recent development, however, and is unlikely to have influenced the arsenal of the average meta-analyst.

Although Egger's regression has its weaknesses, Vevea et al. states that it is still a useful method (1). Furthermore, van Aert et al. utilized it in their reanalysis of meta-analyses (53).

Rank correlation test

The rank correlation test is another statistical method that tests for funnel plot asymmetry (54). It is generally recognized that it has less power than other publication bias methods, including Egger's regression test (40, 55). The same recommendations as for Egger's test regarding sample sizes, also applies to rank correlation test (51).

Trim-and-fill

The trim-and-fill-method (56, 57), is a non-parametric method to adjust for publication bias. Trim-and-fill, along with Egger's regression and rank correlation, is a method based on funnel plot asymmetry. It trims the most extreme effect sizes in the funnel plot and fills them into the other side of the funnel plot to artificially create funnel plot *symmetry*. This enables researchers to redo the meta-analysis to obtain the new meta-analytic estimate of the mean effect size (53). Since it is based on funnel plot asymmetry, it has similar weaknesses as the other tests based on the funnel plot, including the fact that several reasons other than publication bias can lead to asymmetry (1). Trim-and-fill also rests on the strong assumption that potentially unpublished studies have the most extreme results, so Lin et al. states that the method is commonly recommended only as a sensitivity analysis (40).

There are multiple variations to the trim-and-fill method. A fixed- or a random effects model can be utilized, both in the trim-and-fill phase and when estimating the new summary effect size (1). Different advice has been advocated on which combination to apply (1) and Duval recommends utilizing both the random-random, fixed-random, and fixed-fixed combinations, as the research on the best variation is inconclusive (58). Another variation is which estimator should be utilized to estimate the number of missing studies. The options are L_0, R_0 and Q_0 (58). L_0 and R_0 have been the recommended estimators, as Duval states that Q_0 is irrelevant, due to having worse properties than the two other estimators (58). However, Shi & Lin recommends utilizing all three methods in a sensitivity analysis approach (59). They also state that although the choice of estimator can heavily influence the results, few meta-analysts report the specific estimator they used (59).

Shi & Lin notes that it is important to specify the right direction regarding trim-and-fill. Some programs run Egger's regression to automatically determine the direction of missing studies, and other programs leave the direction as default. However, both options can be wrong. They recommend considering the stakeholders preferences when deciding direction. If the direction is wrong, the results are invalid (59). This could in practice be complicated, especially when results are closer to 0, and when studies in both directions potentially could be missing. Another caution meta-analysts should consider with trim-and-fill, is regarding outliers. According to Shi & Li, outliers could greatly influence the results of trim-and-fill. When outliers are present, they recommend doing sensitivity analyses which include and exclude the outliers to assess the impact (59).

In the matter of the profitableness of the method, Vevea et al. recommends using the method (1). Van Aert on the other hand, discourages the use of trim-and-fill because it can incorrectly add studies when none are missing, particularly under high heterogeneity (53).

PET-PEESE

PET-PEESE (60) is a variation of Egger's regression, consisting of two tests: PET and PEESE. PET (Precision-Effect Test) is mostly utilized to detect if publication bias is present or not. Then, PEESE (Precision-Effect Estimate with Standard Error) is utilized to obtain the adjusted estimated effect size. The test has several flaws, of which they share some with Egger's regression and other variants of Egger's test (1). This includes that it cannot incorporate random or between-studies heterogeneity, low power, poor performance in the presence of heterogeneity, and biased results (1). Vevea et al. recommend not utilizing the test, as there are other tests with more positives (1).

All funnel plot asymmetry tests

As noted earlier, other reasons than publication bias might cause funnel plot asymmetry (1). Consequently, results from funnel plot asymmetry tests indicating asymmetry, do not necessarily mean publication bias is present. Also, tests that are based on funnel plot asymmetry generally perform poorly on samples with high heterogeneity (61, 62), and Ioannidis & Trikalinos recommend no more than 50% I^2 (61).

Fail-safe N

Rosenthals Fail-safe N (63), is a publication bias method that calculates the number of missing studies with no effect that would be required to bring the meta-analytic findings to a non-significant level (2). There is a consensus against utilizing Rosenthal's Fail-safe N (1, 2, 41, 53). Thus, this version will not be discussed further. Alternatively, an updated version is available by Orwin (64). This method addresses some issues of the original Fail-safe, by enabling the researcher to specify the effect size of the missing studies (in Rosenthals version the effect size was 0), and specify a value that represents the lowest summary effect size that would be meaningful (2).

Selection models

Selection models are a class of publication bias methods that specifies a model describing the assumed mechanism behind study suppression (1). As such, they represent a fundamentally different type of method than the asymmetry-based methods described above. They are regarded as complex (1), requiring sophisticated assumptions and choices (53), and may therefore be utilized less. Furthermore, they generally have high sample size requirements (1, 65). Vevea et al. still states that selection models are recommended over other methods, because they tackle heterogeneity better. Additionally, they tend to perform well in simulations, and they allow the meta-analyst to test for publication bias under a range of different selection patterns (1). There are lots of different selection models that may be effective methods. However, because selection models are so rarely utilized by meta-analysts, and because there is a range of variations, we will only investigate a few.

Hedges (66), Vevea & Hedges (67), and Iyengar & Greenhouse (68) are all examples of three-parameter selection models. Firstly, these models estimate weights for the suppression mechanism, depending on the probability that different p-values get published.

Secondly, they estimate an effect size. Thirdly they may estimate heterogeneity in the study sample (69). A strength of these selection models is substantiated in that both the data model (which estimates the effect size and heterogeneity) and the selection model (which estimates the weights) are explicitly specified. According to McShane et al., this results in selection models being robust for different meta-analytic settings, among other advantages (69).

A weakness of these selection models is that with small data sets it is often impossible to estimate the weights for the p-values (1). However, the Vevea & Woods (70) method provides an adaptation to this problem, enabling users to specify their own weights and p-value cut points (1). This implies the meta-analyst can check if their results are robust to different publication bias patterns, without requiring lots of data to estimate the degree of publication bias.

To summarize the recommendations reviewed above, it is important to apply multiple methods to tackle publication bias (1, 41-43). Furthermore, meta-analysts ought to utilize a sensitivity analysis approach regarding publication bias methods (1, 44-46), implying that the meta-analysts run different publication bias methods, and tests whether the summary results are vulnerable to the different mechanisms that underlie the different methods (46). Meta-analysts also ought to utilize publication bias tests that perform reasonably well (44). Recommended methods to detect or adjust for publication bias include Egger’s regression (1, 50, 53), trim-and-fill (1, 56, 57), three-parameter selection models (1, 66, 69), and Vevea & Woods selection model (67). Trim-and-fill is a method with multiple possible variations, and it is recommended to apply most of the variations (58, 59). Significant funnel plot asymmetry tests do not necessarily mean that there is publication bias present, as other reasons might cause funnel plot asymmetry (1). Additionally, funnel plot asymmetry tests should not be utilized with a low number of individual studies, nor with a high amount of heterogeneity ($I^2 > 50\%$) (61). Hedges selection model, and other three-parameter selection models, require large data sets. However, Vevea & Woods selection model can still be utilized with smaller data sets (1). Lastly, Rosenthal's original method of Fail-Safe N should not be utilized (1, 2, 41, 53).

Given the advice from multiple sources, meta-analysts could detect and correct for publication bias. It is, however, uncertain whether researchers are aware of the advice, and if the advice is followed in meta-analyses, but a handful of findings seem to suggest that many are not, and do not.

Låg & Sæle (49) investigated the extent of publication bias in two selected educational interventions. They found that 80 % of the included meta-analyses used at least 1 publication bias method, with visual inspection of funnel plot being the most frequent. Only 12 of 51 meta-analyses (23.53%) utilized 3 or more methods, indicating that few followed the advice to rely on several methods in a sensitivity analysis approach. None of a total of 51 meta-analyses utilized selection models. According to reanalyzes of study-level data, meta-analytic estimates were slightly inflated. Kepes et al., demonstrated in their study that research in organizational sciences tends to pay little attention to publication bias (71). Furthermore, they recommend publication bias analyses in various scientific fields.

Ferguson & Brannick examined the use of publication bias methods in a recent sample of psychological meta-analyses. They found that nearly a third of all meta-analysts did not attend to publication bias at all. Furthermore, meta-analysts rarely utilized more than 1 method. They also reanalyzed the data utilizing Orwin’s fail-safe N, rank correlation or Egger’s regression and trim-and-fill. Their findings illustrated that approximately 20 % to 40 % of the meta-analyses were potentially affected by publication bias, depending on method for estimating publication bias (42). This suggests that psychology may be somewhat affected by inflated results.

Similarly, McClain et al. also surveyed psychological meta-analyses from 1980 to 2019. In their sample, 40 % of the meta-analyses did not utilize any methods to detect publication bias, and 24 % did not address publication bias at all. The most frequent methods were visual analysis of funnel plot (20%), Orwin's fail-safe N (19%), and Egger's regression (19%) (48).

An important area of research is the effectiveness of psychotherapy. Research in this area could potentially guide the practices of health care. Furthermore, even slightly more effective psychotherapy could imply that countless people suffering from mental health issues could benefit. There is a need to assess how large the problem of publication bias potentially is in this area of research, to ensure that mental health recommendations and practice is evidence based.

Niemeyer, Musch and Pietrowsky reanalyzed meta-analyses of the efficacy regarding psychotherapeutic interventions for depression. A one-sided test with rank correlation and Egger's regression found that both tests got significant results in about 29 % of the meta-analyses. For two-sided tests rank correlation test and Egger's regression yielded significant results in 16.13 % and 19.35 % of meta-analyses, respectively. For trim-and-fill, 12 meta-analyses indicated at least 1 imputed study. However, the authors noted that the power for a majority of reanalyzes were low, and that the estimates did not change enough to alter the meaning of the results in most cases (72).

To survey the potential implications of publication bias in psychotherapy research, we wanted to investigate how meta-analysts detect and control for publication bias in their meta-analyses. Furthermore, we also wanted to do reanalyzes of the meta-analyses' study-level data, to get a gauge of the potential influence of publication bias in a statistical manner. Therefore, our objective was divided into two components:

First, we systematically investigated researchers use of methods to detect and correct for publication bias and compared it with experts' recommendations, including the usage of methods, variants and estimators utilized, and how it was reported.

Second, we reanalyzed data sets and performed our own analyses to detect and correct for publication bias based on recommendations from the field. Thus, we were able to investigate if our reanalysis of the data sets yielded different outcomes than the initial meta-analyses. The findings could indicate whether the published estimates are vulnerable to publication bias.

To examine the potential implications of publication bias in perceived reliable methods that potentially can influence health care practice and policies, we decided to only study the meta-analyses that conducted experimental or quasi-experimental studies. We also wanted to survey and reanalyze a broad field in psychotherapeutic research, while still limiting the scope of the study to a defined set of disorders. For that reason, we decided to investigate meta-analyses regarding the effect of psychotherapeutic interventions for three large categories of mental disorders: Depression, anxiety and PTSD.

Method

Our study has two methodological components. 1: Describing the application of publication bias methods by meta-analysts in psychotherapy research. 2: To gauge the likelihood that publication bias has influenced published meta-analytic estimates in this area of research. Reporting will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (73).

Identification of meta-analyses

We searched and selected psychological journals/articles from 1 database, PsycINFO on Ovid. The search was conducted September 14th, 2021, with the controlled search term "psychotherapy/" exploded, and the results then limited to "meta-analyses", using the methodology limiter. The initial search yielded a total of 1095 articles.

Eligibility criteria

The objective was to review previously published meta-analyses. While a comprehensive review of a broader field might have been ideal, because of time constraints associated with the format of this study, we were forced to narrow the scope to a more manageable field. To find the best balance between the comprehensive and the achievable, we aimed to formulate eligibility criteria that would delimit a field, while simultaneously providing good coverage of that field.

We included studies whose primarily objective was to investigate the effects of psychotherapeutic interventions, conducting a standard meta-analysis. Standard meta-analyses refer to meta-analytic techniques that focuses on differences between two means, either with a fixed effects model or a random effects model. Studies were considered eligible if a psychotherapeutic intervention was compared experimentally or quasi-experimentally with psychotherapeutic interventions or a non-intervention control group. We included only meta-analyses using studies with primarily adult samples (18-65 years). Nevertheless, meta-analyses with some studies on samples of children or elderly were included if the primary focus was on the adult population. We only included interventions on populations with at least some symptoms of the following mental disorders: depression, anxiety and PTSD. Within these disorders we also included a variety of anxiety disorders such as generalized anxiety disorder (GAD), specific phobia, social anxiety, obsessive-compulsive anxiety disorder (OCD), and panic disorder with or without agoraphobia. Meta-analyses were excluded if their primary focus was on other disorders or diseases than our target disorders. Meta-analyses that contained some studies with other disorders were still included, if the meta-analyses did not focus on the non-target disorders. Finally, we excluded meta-analyses that were not written in a Scandinavian or the English language.

In addition to the eligibility criteria stated above, additional criteria had to be met to be eligible for reanalysis: 1. The meta-analyses had to have 10 or more effect sizes. 2. The meta-analyses had to use a controlled standardized mean difference score as their summary effect size. This includes Cohen's d and Hedges g. 3. The meta-analyses had to report a standardized mean difference score (SMD) on each individual study, or a reported mean and standard deviation for each group, so the SMD could be calculated. 4. The study had to report either a variance component around the effect size for each individual study; confidence intervals around the effect size for each study; or standard deviation for both the treatment group and the control group for each study.

Selection process

For the selection process we applied the screening tool Abstrackr (74) and based the implementation of the screening process on the best practice guidelines provided by Polanin and colleagues (75). Abstrackr is an automated abstract screening tool, which enabled us to connect and organize the results from the literature search in one place. We began by reading some of the abstracts from our search to get a general view of the literature before refining the eligibility criteria. We also conducted several pilot screening sessions using various eligibility criteria, to get a gauge of the potential number of included studies. After the eligibility criteria were finalized, we screened 30 abstracts collaboratively to ensure agreement on the implementation of the criteria. Subsequently, the rest of the abstracts were screened independently by at least two different authors. In case of disagreements, all three authors agreed whether the study met the eligibility criteria or not.

Data collection process

For the data collection process, we obtained the full texts of the meta-analyses that met the eligibility criteria after the screening process. Data was collected from the first 15 full texts by all three authors to ensure agreement on the coding rules. Following the first 15 full texts, data was independently collected from every article by one author each. This was done by examining the full text, tables, figures and the supplementary material. Any uncertainties throughout the coding process were discussed within the entire group and resolved. If the meta-analyses met the eligibility criteria for reanalysis, additional variables were coded.

If a full text had multiple meta-analyses, data from the meta-analysis with the greatest number of comparisons were coded. If two meta-analyses contained the same number of comparisons, data from the meta-analysis first reported were coded. However, if both a random effects meta-analysis and a fixed-effects analysis were reported with the same number of comparisons, data from the random-effects meta-analysis were coded.

Data items

The following variables were coded: target disorder, whether the meta-analysts suspected publication bias, number of methods for detecting and adjusting for publication bias, which methods was utilized to detect and adjust for publication bias, which variant of the methods was used, and whether the variant was reported.

For meta-analyses that met the eligibility criteria for reanalysis, additional variables were coded for each study included in the meta-analysis: authors, the mean of both groups (the treatment group and the comparison group), the standard deviation of both groups, number of participants in both groups, the total number of participants, the standardized mean difference (SMD), the confidence interval around the SMD, the variance of the SMD, and the standard error of the SMD. If the standard error or the variance of the mean was not reported, we calculated it. To calculate the standard error, we subtracted the means from each study from their upper limit of the confidence interval, and then divided it by 1.96. If the confidence interval was unavailable, we calculated it based on the method by Lipsey & Wilson, which uses the mean of each study, along with the number of participants in the studies' treatment group and control group (76).

Effect measures

For the descriptive results, the frequency and percentage of our outcome variables were calculated. The variables of interest were the number of methods for detecting or adjusting for publication bias, percentage of studies that had utilized different numbers of methods for detecting and adjusting for publication bias, and frequency and percentage of studies that used each publication bias method.

Additionally, our reanalysis yielded the following variables of interest: a descriptive summary about the frequency of methods that detected or corrected for publication bias, and the mean difference between the original summary estimate and the reanalyzed summary estimates from trim-and-fill and Vevea & Woods' selection model.

Synthesis methods

For our reanalysis, all statistical analyses were conducted in R (77). Some studies had large differences between their heterogeneity estimates and ours. We therefore decided to exclude studies with more than 10 percent difference in I^2, to maintain similarity to the original studies. Furthermore, following Ioannidis & Trikalinos advice (61), we only performed funnel plot asymmetry tests (trim-and-fill, Egger's regression and rank correlation test) on meta-analyses where we obtained under 50% I^2 from our reanalysis.

Depending on which method the included meta-analysis conducted, we ran a random-effects model or fixed-effects model. When random effects meta-analysis was applied, we initially tested statistical analyses with both the maximum likelihood (ML) estimator and the restricted maximum likelihood (REML) estimator. We eventually choose the ML estimator because it initially obtained closer I^2 estimates to the meta-analyses' own estimates, than with the REML estimator.

When analyzing study level data, we applied Egger´s regression intercept (50), rank correlation test (54), trim-and-fill (56, 57) and a weight-function model by Vevea & Woods (70).

The R package *metafor* was utilized to calculate asymmetry tests in RStudio (78). Egger´s regression intercept was performed with both RMA and LM models. Rank correlation was conducted as a statistical analogue to the funnel graph (54).

For the trim-and-fill analyses we utilized both a random effects model and a fixed effects model to estimate the number of imputed studies. On analyses that were originally conducted using a fixed effects model, we only utilized the fixed effects model to estimate the number of imputed studies. For the reanalyzes of the updated data (including the imputed studies), we utilized random effects model for the analyses that were originally conducted with a random effects model. Likewise, we utilized a fixed effects model for the analyses that were originally conducted with a fixed effects model. For all combinations of models, we additionally estimated the number of imputed studies applying both the L_0 and the R_0 estimator. Furthermore, we had to ascertain whether the trim-and-fill analysis should be calculated on either the right or left side. The value of effect size (positive or negative) determined the command (right side or left side) executed in the console. In certain cases where the effect size estimate was close to 0, we ran the trim-and-fill analyses on both right and left side.

The R package *weightr* (79) was applied for the Vevea & Woods version of the weight-function model (70). We modelled a weighted p-value (p>.025) interval at .2, .5 and

.8 to assess severe, moderate and mild publication bias for non-significant studies, respectively (70). For analyses where the suppressed studies theoretically were on the right side, the inputs in the model had to be changed. The p-value interval was then changed to $p > .975$, and the weights were reversed. For the Vevea & Woods model, the same method was applied for deciding which side to run the analysis on.

For both trim-and-fill and the Vevea & Woods model, the side which indicated the highest number of imputed studies were utilized as data in our analyses, although both sides are reported in our tables.

Results

Study selection

During the abstract screening process, 936 articles were excluded due to not meeting the eligibility criteria. This resulted in 86 meta-analyses meeting the eligibility criteria for our study. Additionally, 37 of 86 meta-analyses met the eligibility criteria for reanalysis. The process of study selection is illustrated in Figure 1.

Study characteristics and results of individual studies

Each meta-analysis included in this study, along with some key characteristics, are illustrated in Table S1 in the supplementary material. Table S2 is for meta-analyses included in our reanalyzes. A full list of the included meta-analyses' references are displayed in S3.

The results of the reanalyzes on each meta-analysis are displayed in the supplemental material. Table S4 displays the results of Egger's test and rank correlation. Table S5 displays the results of trim-and-fill with meta-analyses conducted with random effects, Table S6

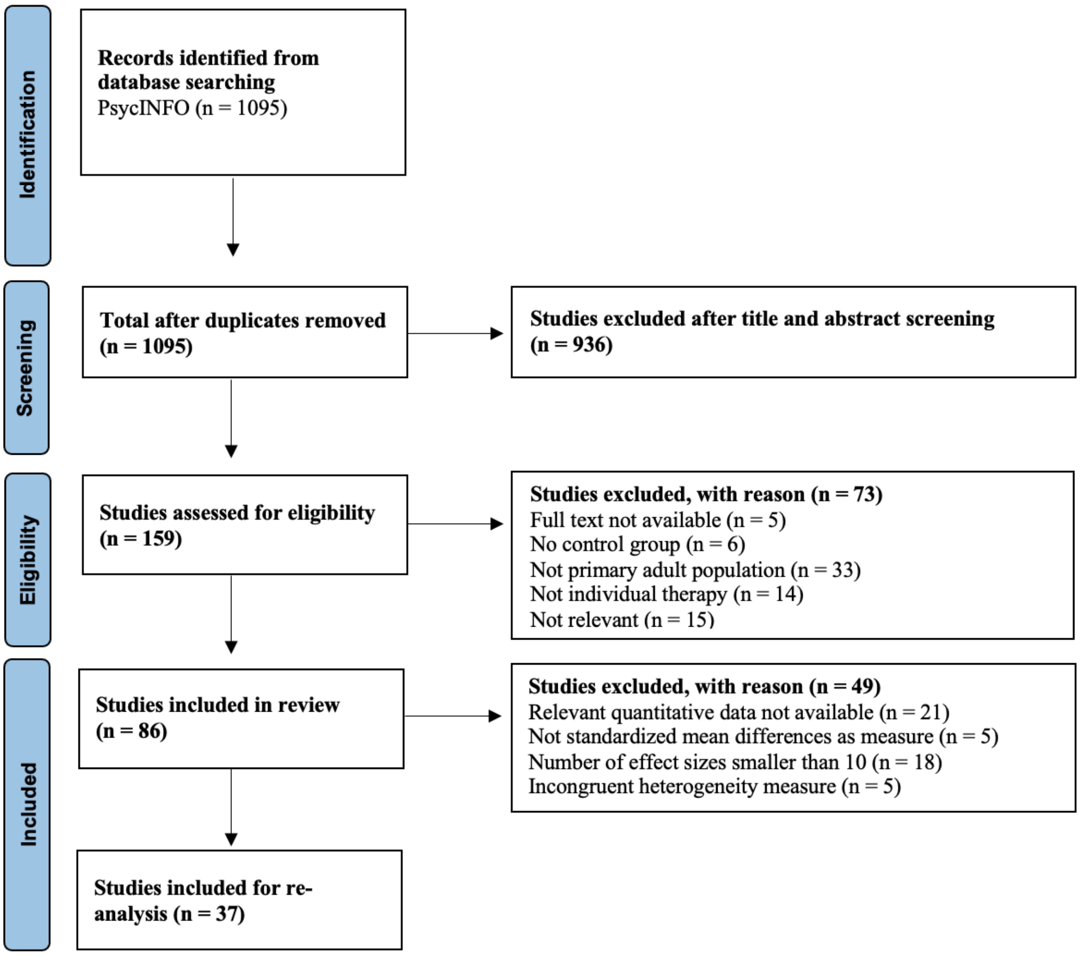

Figure 1. PRISMA flowchart illustrating the study selection process.

displays the results of trim-and-fill with meta-analyses conducted with fixed effects, and finally, Table S7 displays the results of Vevea & Woods' selection model.

Results of syntheses

Descriptive results, concerning the use of publication bias methods are displayed in Table 1. Of our included meta-analyses 20 of 86 (23.3%) applied 0 publication bias methods. 12 of 86 (14.0%) applied 1 publication bias method. 22 of 86 (25.6%) applied 2 publication bias methods. 29 of 86 (33.7%) applied 3 publication bias methods. 3 of 86 (3.49%) applied 4 publication bias methods. 66 of 86 (76,7%) of the included meta-analyses

Table 1
Meta-analysists use of methods for detecting or correcting for publication bias

	Number used	Total number	Percent
Used trim-and-fill (TaF)	45	86	52.3%
Used at least two versions of TaF	0	45	0.000%
Reported estimator used (L0, R0 or Q0)	0	45	0.000%
Reported method to estimate imputed studies in TaF	3	45	6.67%
Used more than one variation of TaF (either method or estimator)	0	45	0.000%
Used Egger's regression	33	86	38.4%
Used both models of Egger's test (RMA and LM)	0	33	0.000%
Reported what model is used in Egger's test	4*	33	12.1%
Used Fail safe N	6	86	6.98%
Used Orwin's Fail safe N	5	86	5.81%
Used rank correlation	6	86	6.98%
Used funnel plot to inspect for PB	59	86	68.6%
Reader can see funnel plot	24	59	40.7%
Used a selection model	0	86	0.000%
Used at least one statistical PB method	66	86	76.7%
Used 0 statistical PB methods	20	86	23.3%
Used 1 statistical PB method	12	86	14.0%
Used 2 statistical PB methods	22	86	25.6%
Used 3 statistical PB methods	29	86	33.7%
Used 4 statistical PB methods	3	86	3.49%
Searched for unpublished studies	22	86	25.6%
Reports that publication bias might have influenced the summary estimate	23	86	26.7%

Note: *= model not explicitly stated, only inferred based on whether z-values or b-values were used.

applied at least 1 publication bias method. 32 of 86 (37.2%) of the included meta-analyses applied at least 3 publication bias methods. The average number of statistical publication bias methods was: 1.80.

Table 1 also visualizes the most employed methods for detecting or correcting for publication bias, which was visual inspection of a funnel plot, with 59 of 86 studies (68.6%) which employed the method. The second most utilized test was Duval & Tweedie's trim-and-fill, with 45 of 86 studies (52.3%) that employed the method. Egger's test was third with 33 of 86 (38.4%) studies that employed the method. Rank correlation, Fail Safe N and Orwin's Fail Safe N were employed 6 (6.98%), 6 (6.98%), and 5 (5.81%) times respectively. No other methods for detecting or correcting for publication bias was employed.

Additionally, no studies employed more than 1 variant of a single test (relevant for trim-and-fill and Egger's regression). 20 of 86 studies (23.3%), reported that publication bias

Table 2
Distribution of publication bias methods by decade.

Decade	*N*	Funnel Plot		Orwin´s Failsafe N		Rosenthal´s Failsafe N		Rank Correlation		Egger´s Regression		Trim-and-fill	
1990-1999	1	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
2000-2009	14	6	43%	2	14%	2	14%	0	0%	1	7%	5	36%
2010-2019	57	42	74%	2	4%	4	7%	3	5%	24	42%	34	60%
2020-	14	11	79%	0	0%	1	7%	2	14%	8	57%	7	50%
Total	86	59	69%	4	5%	7	8%	5	6%	33	38%	46	53%

might have been likely to influence their summary estimate. Lastly, Table 2 illustrates the utilization of different publication bias methods over the past years.

Table 1 displays that 0 of the 45 studies that employed the trim-and-fill method, reported which type of estimator was utilized. Only 3 of 45 studies (6.67%) reported whether a random effects model or a fixed effects model were applied to estimate the number of imputed studies. For Egger's test, 0 of the studies reported explicitly what model was applied to estimate the output (LM or RMA). However, 4 of 33 studies (12.1%) reported it indirectly by reporting either a *z*-value or a *b*-value in their results.

Of the included meta-analyses, 37 of 86 (43.0%) met the criteria for reanalysis (see Figure 1). 20 of the meta-analyses we reanalyzed, resulted in > 50% I^2. Consequently, funnel plot asymmetry tests were not applied on the 20 meta-analyses it concerned, leaving 17 meta-analyses. 3 of the included meta-analyses were completed utilizing a fixed effects model by the original authors and was therefore also reanalyzed with fixed effects. The number of meta-analyses where the publication bias tests suggested publication bias is illustrated in Table 3.

4 of 17 (23.5%) meta-analyses reanalyzed with the RMA model of Egger's regression obtained significant results at a p-value of .05. For the LM model of Egger's regression, 3 of 17 (17.7%) got significant results at the same p-value. All 3 significant results from the LM model were also significant on the RMA model. However, the RMA

Table 3

Number of meta-analyses potentially affected by publication bias based on results from reanalysis

Test variation	Tests suggests PB	Total number of meta-analyses analysed	Percent	n of invalid results
Egger's (RMA)	4	17	23.5%	20
Egger's (LM)	3	17	17.7%	20
Rank correlation				
Trim-and-fill random-random (L_0)	7	14	50.0%	1
Trim-and-fill random-random (R_0)	9	14	64.3%	1
Trim-and-fill fixed-random (L_0)	6	15	40.0%	0
Trim-and-fill fixed-random (R_0)	10	15	66.7%	0
Trim-and-fill fixed-fixed (L_0)	1	2	50.00%	1
Trim-and-fill fixed-fixed (R_0)	2	2	100%	1

model had one more significant result than the LM model. That implies that there were 4 meta-analyses with at least 1 significant result on Egger's regression.

For rank correlation, 9 of 17 (52.9%) analyses resulted in a warning message (see Table S4). 3 of 17 (17.7%) meta-analyses obtained a significant result at .05 level. Of these, 2 of 3 (66.7%) received a warning message. Significant results indicate a risk for publication bias or funnel plot asymmetry. The results are displayed in Table S4.

The results from our trim-and-fill reanalyzes on the studies with $I^2 < 50$ % are also displayed in Table S5. Additionally, the number of trim-and-fill reanalyzes which indicated at least one missing study are displayed in Table 3. Random-random with L_0 estimator indicated that 7 of 14 (50.0%) had missing studies. Random-random with R_0 estimator indicated that 9 of 14 (64.3%) had missing studies. Fixed-random with L_0 estimator indicated that 6 of 15 (40.0%) had missing studies. Fixed-random with R_0 estimator indicated that 10 of 15 (66.7%) had missing studies. Regarding the 3 studies conducting a fixed effects model in our reanalysis, the fixed-fixed effects version of trim-and-fill with the L_0 estimator indicated that

1 of 2 meta-analyses had missing studies. The fixed-fixed model with the R_0 estimator indicated that 2 of 2 meta-analyses had missing studies.

Combining both the random effects and fixed effects meta-analyses, 12 of the 17 (70.6%) meta-analyses obtained missing studies in *at least one* of the four variants of trim-and-fill employed. Furthermore, 5 of 16 (31.3 %) meta-analyses obtained missing studies in *all* trim-and-fill variants. One meta-analysis obtained missing studies on both estimators in the fixed-random model. However, the trim-and-fill analysis did not converge on the random-random model and was therefore not included in that number.

3 of 17 reanalyzes (17.6%) yielded a greater difference than 0.10 in SMD between the trim-and-fill adjusted estimates and the unadjusted estimates in *at least one* of the trim-and-fill variants. 0 of 16 reanalyzes yielded a greater difference than 0.10 SMD, in *all* the trim-and-fill variants.

The average number of missing studies and change in effect size for the different variants of trim-and-fill is displayed in Table 4. The mean number of missing studies for all variants of trim-and-fill was 2.55. The mean change of the standardized mean difference after applying trim-and-fill was 0.041 for all variants. Lastly, the difference between the adjusted and unadjusted summary estimate was greater than 0.10 in *at least one* of the variants, in 3 of 17 (17.7%) meta-analyses. The variation between the random-random and fixed-random results displayed in Table 4, are biased by an outlier with a high number of missing studies, which could not be computed with the random-random version, thus causing the fixed-fixed results to estimate higher numbers of publication bias than random-random.

The results of Vevea & Woods selection model are displayed in Table S7. 6 of 37 (16.2%) of the included meta-analyses obtained a warning message in RStudio. Additionally, 2 meta-analyses (5.41%) obtained an error message. One study obtained both

Table 4
Descriptive statistics from trim-and-fill and Vevea & Woods' selection model

Publication bias method	N total missing studies	k number of meta-analyses analysed	Mean number of missing studies	SMD change total	Mean SMD change
R-R L_0	30	14	2.21	0.42	0.03
R-R R_0	21	14	1.50	0.31	0.022
F-R L_0	39	15	2.60	0,67	0.044
F-R R_0	62	15	4.13	0.90	0.060
F-F L_0	2	2	1	0.06	0.004
F-F R_0	4	2	2	0.16	0.011
Total: All trim-and-fill analyses	**158**	**62**	**2.55**	**2.52**	**0.041**
Vevea & Woods Low pb.	-	30	-	1.12	0.037
Vevea & Woods, mod. Pb.	-	30	-	3.16	0.11
Vevea & Woods high pb.	-	30	-	7.22	0.24
Total: Vevea & Woods	**-**	**90**	**-**	**11.50**	**0.13**
Total	**-**	**152**	**-**	**14,02**	**0.092**

Note. SMD = standardized mean difference. R-R = the random-random variant of trim-and-fill. F-R = the fixed-random variant of trim-and-fill. F-F = the fixed-fixed variant of trim-and-fill (only applied on fixed meta-analyses). Low pb. = weights that indicate low probability of publication bias. Mod. Pb. = weights that indicate moderate probability of publication bias. High pb. = weights that indicate high probability of publication bias.

the warning message and the error message. We did not include the 7 adjusted estimates where there were warning messages or error messages. Table S7 visualize which meta-analyses received the messages, and the content of the messages.

For the weight equivalent to *low* probability of publication bias, 1 of 30 (3.33%) meta-analyses yielded a greater difference than 0.10 SMD between the weighted estimate and the unadjusted estimate. For the weight equivalent to *moderate* probability of publication bias, 18 of 30 (60.0%) meta-analyses yielded a greater difference than 0.10 SMD between the weighted estimate and the unadjusted estimate. Lastly, for the weight equivalent to *high* probability of publication bias, 29 of the 30 (96.7%) meta-analyses yielded a greater difference than 0.10 SMD between the weighted estimate and the unadjusted estimate. The mean change of the standardized mean difference (SMD) after applying the Vevea & Woods' selection model, is displayed the lower half of Table 4 for the three different weights. The mean amount of change in SMD for all weights was 0.13.

Of the meta-analyses included for reanalysis, 27 of 37 (73.0%) did not explicitly state that publication bias had an impact on their results. Of these, 11 of 27 (40.7%) were removed due to > 50% heterogeneity. When meta-analyses with > 50% heterogeneity were removed, 4 of the 16 (25.0%) meta-analyses which did not suspect publication bias, obtained a significant result on either one of the Eggers' variants or rank correlation. Of the 4 significant results, 3 of 4 (75.0%) obtained a significant result at the RMA model, 2 of 4 (50.0%) on the LM model, and 2 of 4 (50.0%) on the rank correlation test. For trim-and-fill, 3 of 4 (75.0%) meta-analyses with significant results had missing studies on *all* variants. The aggregated difference in SMD in these meta-analyses were: 0.44. 0.088 and 0.03 for all relevant variants of trim-and-fill.

Regarding the 4 meta-analyses that did not suspect publication bias *and* obtained significant results on either the Egger´s variants or rank correlation, 1 of 3 (33.3%) yielded more than 0.10 difference in SMD with moderate probability of publication bias in Vevea and Woods model (1 of the meta-analyses were removed due to a warning message).

Discussion

Our results suggest an overall need for improvement in applying publication bias methods in research of psychotherapeutic interventions. Demonstrating that 76.7% applied at least 1 publication bias method. Furthermore, only 37.2% utilized at least three publication bias methods. Surprisingly, 23.3% of the included meta-analyses utilized 0 publication bias methods. In comparison to our results, Ferguson and Brannick's results (2012) advocates that 70% of psychological research assessed publication bias statistically (42). Låg & Sæle's (2021) study from educational research demonstrated that 80% of the meta-analyses in their sample applied at least 1 method to assess publication bias (49). Their estimates are approximately consistent with our findings.

The literature and previous research converge on the consensus that utilizing several and different statistical publication bias methods is the preferred approach (1, 41-43). Contrary to this recommendation, our results present a divergent use of this recommendation. Thus, indicating that publication bias may threaten the respective meta-analyses utilizing 0 or 1 method. Consequently, the effect size of the interventions in these meta-analyzes could be inflated.

Of all 86 meta-analyses surveyed, none utilized more than 1 variant of a publication bias method. Following the advice of utilizing publication bias methods as sensitivity analyses (1, 44-46), it seems appropriate to analyze different versions of a test. Especially when a small number of publication bias methods are utilized. This mostly pertains to Egger's and trim-and-fill, where different models and estimators can be selected. There is also explicitly stated advice from Duval to report all model variations of trim-and-fill (58).

The most frequent method for detecting or adjusting for publication bias was visual inspection of the funnel plot, which was applied in 68.6% of the meta-analyses. This is consistent with the results from Låg & Sæle (49) and McClain et al. (48), although the frequency differed. The second most utilized method was trim-and-fill (52.3%), and the third most utilized method was Egger's regression (38.4%). Notably, all three methods are based on funnel plot asymmetry. Firstly, this implies that high heterogeneity and low power can invalidate the results of the test. Secondly, meta-analysts could benefit from applying another type of method, as other types of methods have other assumptions, strengths and weaknesses. Selection models were never utilized in our sample of meta-analyses, which could provide a useful addition to the funnel plot asymmetry tests. Lastly, 6.98 % of our sample utilized the original fail-safe N, which is strongly advised against by multiple authors (1, 2, 41, 53).

To summarize, although many meta-analysts in our sample were aware of publication bias, a relatively high number also did nothing to tackle potential publication bias. The use of multiple methods is varied among meta-analysts, with many only utilizing 1 or 2 methods. Furthermore, meta-analysts do not utilize different variants of tests, and generally do not report which variant is utilized. Consequently, it appears as if meta-analysts still have more to improve on regarding the detection and adjustment for publication bias. Positively, our numbers indicate that a considerable number of meta-analysts utilize more than one publication bias method.

Signs of publication bias from our reanalysis

Regarding Egger´s interception test, 23.53% had at least 1 significant result either on the RMA- or LM model. Furthermore, 66.67% of the meta-analyses had at least 1 missing study, in at least 1 of the 4 versions of trim-and-fill. We accounted for the issues regarding funnel plot asymmetry tests by not including meta-analyses with less than 10 effect sizes, and not analyzing meta-analyses with $I^2 > 50\%$. Therefore, the issues regarding false positives in funnel plot asymmetry tests should be at least somewhat counteracted. These results indicate that a substantial number of meta-analyses may be affected by at least some level of publication bias. Furthermore, meta-analytic effect size estimates of psychotherapeutic interventions on depression, anxiety and PTSD may be inflated.

A majority of the meta-analyses reanalyzed with trim-and-fill yielded at least 1 missing study. However, only 3 of 17 resulted in a greater difference than 0.10 in SMD between the adjusted and the unadjusted effect sizes. Furthermore, the mean change of the SMD after applying trim-and-fill for all variants, was only 0.041, and the mean number of imputed studies was 2.55. This indicates that many reanalyzes may be affected by some level of publication bias. However, the general meaning of the findings may still be robust.

The results from the Vevea & Woods selection model indicates that the meta-analyses that were reanalyzed generally are mostly unaffected by low levels of publication bias. Results with a moderate level of publication bias illustrates that 60 % of the meta-analyses yields a greater than 0.10 difference in SMD between the weighted estimate and the unadjusted estimate. The mean change in SMD for moderate publication bias was 0.11. Regarding high levels of publication bias, 96.66% yields a greater than 0.10 difference in SMD, with a mean difference of 0.24. The results from the Vevea & Woods model thus indicate that meta-analytic results in our sample of reanalyzes are generally robust to low levels of publication bias, but vulnerable to moderate to high levels of publication bias. Furthermore, high levels of publication bias could potentially alter the meaning of the studies' results.

There are indications of publication bias in some of the reanalyzed meta-analyses where the meta-analysts did not suspect that publication bias affected their results. 4 of 16 meta-analyses has significant results in either Egger's regression or rank correlation. 3 have a significant result in Egger's regression or rank correlation, as well as imputed studies in all trim-and-fill variants. Furthermore, 2 of the significant meta-analyses yielded more than a 0.10 difference in SMD between the adjusted and the unadjusted effect size. This indicates that even meta-analyses where it is stated that publication bias may be a problem, could have considerable publication bias. This poses an even higher risk of overestimating the effects of psychotherapeutic interventions, because the estimates could be viewed as unbiased.

To summarize, tests utilized in our reanalysis indicate that a moderate portion of the meta-analyses in our sample may be affected by publication bias. Nevertheless, adjusted estimates from trim-and-fill suggest that only a few meta-analyses may be considerably affected. Thus, the general meaning of the results may remain mostly the same. However, the Vevea & Woods model suggests that moderate to high levels of publication bias could substantially affect the findings of the studies. Lastly, even if meta-analysts report that publication bias most likely has not affected their results, their meta-analytic estimate could still be affected by a considerable degree of publication bias. The evidence put forth suggests that the estimated effectiveness of psychotherapeutic interventions on depression, anxiety and PTSD, from meta-analyses using experimental studies, might be slightly to moderately inflated.

Limitations

To our knowledge, there are no methods currently available to directly measure publication bias. This implies that the evidence of potential publication bias we obtained from reanalyzing meta-analyses, is contingent on the publication bias tests themselves being accurate. Although these tests often have been under simulation studies where their accuracy has been tested (80-82), there is no way of knowing exactly how effective publication bias methods are, to our knowledge. This may be a reason why a sensitivity analysis approach is recommended. Furthermore, while methods for preventing publication bias are not a focus in this study, prevention methods are a crucial part of tackling publication bias.

There also exist multiple other publication bias methods, which we have not covered. The literature recommends utilizing several and different methods. The addition of other methods could potentially have improved the accuracy of our reanalyzes. Contrary, it is not advocated a golden number of methods to apply, and we included the methods which the literature and prior research emphasize. For instance, we only utilized 1 selection model due to our study's time constraints, although ideally more methods could have been applied.

Hedges' selection model (66) is an example of this. On the other hand, a minority of the meta-analyses we included had enough studies that met the requirements to apply such methods. Additionally, the Vevea & Woods selection model could also have been applied with more variations, in line with the sensitivity analysis approach. Finally, it is conceivable that we included enough methods to assume that our results from reanalysis are somewhat reliable.

Furthermore, a lot of the included meta-analyses had high heterogeneity. We decided to follow Ioannidis (56) advice of not using funnel plot asymmetry tests on samples with remarkably high heterogeneity ($I^2 > 50\%$), and the number of funnel plot asymmetry tests that were conducted was thus greatly reduced. This is a problem within our study, and problematic for entire fields of research with remarkably high heterogeneity. Given the advice on triangulation and sensitivity analysis approach, refraining from utilizing a large and recommended set of publication bias tests can be problematic. Especially if other tests are not utilized instead. If fewer publication bias tests are utilized overall, as in our study, it weakens the basis on which to judge whether there is publication bias or not. In our study, only 17 of 37 included meta-analyses could be analyzed by funnel plot asymmetry tests, leaving 20 meta-analyses where only the Vevea & Woods selection model was applied. On the other hand, the range of publication bias methods we employed on the meta-analyses with $I^2 < 50\%$ may be a strength of our study. We applied different variations of Egger's regression and trim-and-fill, as well as rank correlation test and Vevea & Woods' selection model.

A potential limitation of our survey process was the fact that our outcome variables were not clearly defined from the start, since publication bias is a complex topic. Some outcome variables we coded for in the beginning were discarded, due to challenges with coding. Also, the importance of different variables changed in some degree during the coding process. Potentially, this could have contributed to publication bias because the results of our variables might have affected our decisions in the coding process. However, since generalizing is not a goal for this study, the results are not affected by sample error. Therefore, it should not be an excessive concern, as the results will be accurate.

As we did not use any formal probability tests, our study should be regarded as a descriptive survey of a limited area of research. Therefore, it cannot generalize to other populations. It can, however, tell us something about the potential extent of publication bias in meta-analyses of the effectiveness of psychotherapeutic interventions for depression, anxiety and PTSD, and how the meta-analysts deal with this problem.

Implications

There appears to be some publication bias affecting the results of meta-analytic estimates of psychotherapeutic interventions for depression, anxiety and PTSD. The meta-analysts in our sample generally applied measures to counteract this issue. However, some meta-analysts, even from recent time, applied 0 tests to detect or adjust for publication bias, and the majority applied only 1 or 2 tests. Furthermore, even meta-analyses where publication bias were not reported as being likely to have influenced the results, could still be affected by considerable publication bias. Overall, this could have led to an overestimation regarding the effectiveness of psychotherapeutic interventions. A risk of this, is that psychotherapeutic interventions may be chosen based on the degree of overestimation regarding their effectiveness, which could lead to suboptimal help for people struggling with mental health issues. However, the meta-analytic estimates only appear to be marginally to moderately affected by publication bias, which means that the general results of the psychotherapeutic interventions in our sample, may still be valid and reliable.

A major purpose of this study is to motivate, raise awareness and guide researchers to act against publication bias, particularly within psychology. We hope that future meta-analysts aim to follow the recommendations given, to counteract the impacts of publication bias. To discover the most optimal way of helping individuals with mental health issues, one must strive to obtain unbiased evidence of psychotherapeutic interventions.

We encourage further research to continue examining publication bias in psychotherapeutic interventions with larger samples and with other mental disorders. Additionally, a comprehensive practical guide, summarizing research recommendations regarding the application of publication bias methods could be useful for meta-analysts. This could support and guide meta-analysts to make better decisions to address the issue of publication bias.

Conclusions

The first aim of our study was to describe the application of publication bias methods by meta-analysts in psychotherapy research. Meta-analysts in our limited area of research generally provided measures to tackle publication bias. Simultaneously our findings demonstrate that it is still room for improvement regarding the use of publication bias methods. This especially concerns the number of utilized publication bias methods and the number of variants utilized. The second aim of our study was to gauge the likelihood that publication bias has influenced published meta-analytic estimates in this area of research. Our findings indicate some degree of publication bias in our limited area. In conclusion, the combination of publication bias and suboptimal utilization of publication bias methods could potentially have led to overestimated effects of psychotherapeutic interventions.

Acknowledgements

This study was supervised by Torstein Låg.

Funding

This dissertation is associated with UiT (The Arctic University of Norway, Tromsø).

Availability of data and materials

The survey and all raw data are available on request.

Authors´ contributions

All authors contributed to all aspects of the research process. All authors read and approved the final version.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be constructed as a potential conflict of interest.

References

1. Vevea JL, Coburn K, Sutton A. Publication bias. In: Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 3. ed: Russel Sage Foundation; 2019.
2. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley and Sons, Ltd; 2009.
3. Hussain R, Hassali MA, Patel M, Babar Z-U-D. Publication Bias. Encyclopedia of Pharmacy Practice and Clinical Pharmacy. 2019:53.
4. Boutron I, Page MJ, Higgins JP, Altman DG, Lundh A, Hróbjartsson A, et al. Considering bias and conflicts of interest among the included studies. Cochrane Handbook for Systematic Reviews of Interventions. 2019:177-204.
5. Rothstein HR, Sutton AJ, Borenstein M. Publication bias in meta-analysis: Prevention, assessment and adjustments: John Wiley & Sons; 2006.
6. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. Health Technol Assess. 2000;4(10):1-115.
7. Page MJ, Higgins JP, Sterne JA. Assessing risk of bias due to missing results in a synthesis. Cochrane handbook for systematic reviews of interventions. 2019:349-74.
8. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 1988;151(3):419-45.
9. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. Journal of clinical epidemiology. 2000;53(2):207-16.
10. Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. The Lancet. 1991;337(8746):867-72.
11. Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA. 1990;263(10):1385-9.
12. Koricheva J. Non-significant results in ecology: a burden or a blessing in disguise? Oikos. 2003;102(2):397-401.
13. Field A. Discovering statistics using IBM SPSS statistics. 5th ed. London: SAGE Publications Ltd.; 2018.
14. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US States Data. PloS One. 2010;5(4):e10271.
15. Felson DT. Bias in meta-analytic research. Journal of clinical epidemiology. 1992;45(8):885-92.
16. LeVois ME, Layard MW. Publication bias in the environmental tobacco smoke/coronary heart disease epidemiologic literature. Regulatory Toxicology and Pharmacology. 1995;21(1):184-91.
17. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA. 1992;267(3):374-8.
18. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003;326(7400):1167-70.
19. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The Extent and Consequences of P-Hacking in Science. PloS Biology. 2015;13(3):e1002106.
20. Simonsohn U, Nelson LD, Simmons JP. P-curve: A key to the file-drawer. Journal of Experimental Psychology: General. 2014;143(2):534-47.
21. Chuard PJC, Vrtílek M, Head ML, Jennions MD. Evidence that nonsignificant results are sometimes preferred: Reverse P-hacking or selective reporting? PloS Biology. 2019;17(1):e3000127.
22. Ioannidis JP. Why most published research findings are false. PloS Medicine. 2005;2(8):e124.
23. Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, et al. Publication Bias in Editorial Decision Making. JAMA. 2002;287(21):2825-8.
24. Lynch J, Cunningham M, Warme W, Schaad D, Wolf F, Leopold S. Commercially Funded and United States-Based Research Is More Likely to Be Published; Good-Quality Studies with Negative Outcomes Are Not. The Journal of bone and joint surgery American volume. 2007;89:1010-8.

25. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8):iii, ix-xi, 1-193.
26. Senn S. Misunderstanding publication bias: editors are not blameless after all. F1000Res. 2012;1:59-.
27. Pigott TD, Polanin JR. Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review. Review of Educational Research. 2020;90(1):24-46.
28. Hollon D, Miller IJ, Robinson E. Criteria for evaluating treatment guidelines. American Psychologist. 2002;57(12):1052-9.
29. Egger M, Smith GD, O'Rourke K. Introduction: rationale, potentials, and promise of systematic reviews. In: Egger M, Smith GD, O'Rourke K, editors. Systematic reviews in health care: meta-analysis in context. 2nd. ed. London: BMJ Publishing Group; 2001. p. 1-19.
30. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. BMJ. 2012;344:d7762.
31. Polanin JR, Tanner-Smith EE, Hennessy EA. Estimating the Difference Between Published and Unpublished Effect Sizes: A Meta-Review. Review of Educational Research. 2016;86(1):207-36.
32. Fanelli D. “Positive” results increase down the hierarchy of the sciences. PloS One. 2010;5(4):e10068.
33. Schäfer T, Schwarz M. The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. Frontiers in Psychology. 2019;10:1-13.
34. Song F, Hooper L, Loke Y. Publication bias: what is it? How do we measure it? How do we avoid it? Open Access Journal of Clinical Trials. 2013;2013(5):71-81.
35. Joober R, Schmitz N, Annable L, Boksa P. Publication bias: what are the challenges and can they be overcome? Journal of psychiatry & neuroscience: JPN. 2012;37(3):149.
36. Stern J, Simes J. Publication Bias: Evidence of Delayed Publication in a Cohort Study of Clinical Research Projects. BMJ (Clinical research ed). 1997;315:640-5.
37. Paulson K, Saeed M, Mills J, Cuvelier GDE, Kumar R, Raymond C, et al. Publication bias is present in blood and marrow transplantation: an analysis of abstracts at an international meeting. Blood. 2011;118(25):6698-701.
38. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. The British Journal of Psychiatry. 2010;196(3):173-8.
39. Driessen E, Hollon SD, Bockting CLH, Cuijpers P, Turner EH. Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials. PloS One. 2015;10(9):e0137864-e.
40. Lin L, Chu H, Murad MH, Hong C, Qu Z, Cole SR, et al. Empirical Comparison of Publication Bias Tests in Meta-Analysis. J Gen Intern Med. 2018;33(8):1260-7.
41. Marks-Anglin A, Chen Y. A historical review of publication bias. Res Synth Methods. 2020;11(6):725-42.
42. Ferguson CJ, Brannick MT. Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. Psychological methods. 2012;17(1):120.
43. Banks GC, Kepes S, Banks KP. Publication Bias: The Antagonist of Meta-Analytic Reviews and Effective Policymaking. Educational Evaluation and Policy Analysis. 2012;34(3):259-77.
44. Carter EC, Schönbrodt FD, Gervais WM, Hilgard J. Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. Advances in Methods and Practices in Psychological Science. 2019;2(2):115-44.
45. van Assen MALM, van Aert RCM, Wicherts JM. Meta-analysis using effect size distributions of only statistically significant studies. Psychological Methods. 2015;20(3):293-309.
46. Greenhouse JB, Iyengar S. Sensitivity analysis and diagnostics. The handbook of research synthesis and meta-analysis, 2nd ed. New York, NY, US: Russell Sage Foundation; 2009. p. 417-33.
47. Light RJ, Pillemer DB. Summing up: the science of reviewing research. 1984.

48. McClain MB, Callan GL, Harris B, Floyd RG, Haverkamp CR, Golson ME, et al. Methods for addressing publication bias in school psychology journals: a descriptive review of meta-analyses from 1980 to 2019. Journal of school psychology. 2021;84:74-94.
49. Låg T, Sæle RG. Inflated summary effect sizes in meta-analyses of flipped classroom and collaborative learning interventions? Paper presented at American Educational Research Association (AERA) Annual Meeting; Virtual meeting; 2021 April 8-12. p. 1-22.
50. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.
51. Sterne JAC, Sutton AJ, Ioannidis JPA, 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.
52. Pustejovsky JE, Rodgers MA. Testing for funnel plot asymmetry of standardized mean differences. Research Synthesis Methods. 2019;10(1):57-71.
53. Van Aert RC, Wicherts JM, Van Assen MA. Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. PloS One. 2019;14(4):e0215052.
54. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994:1088-101.
55. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. Journal of clinical epidemiology. 2000;53(11):1119-29.
56. 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.
57. Duval S, Tweedie R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. Journal of the american statistical association. 2000;95(449):89-98.
58. Duval S. The trim and fill method. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication bias in meta-analysis: Prevention, assessment and adjustments. Chichester: John Wiley & Sons; 2006. p. 127-44.
59. Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. Medicine (Baltimore). 2019;98(23):e15987. 10.1097/MD.0000000000015987
60. Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods. 2014;5(1):60-78.
61. Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. Cmaj. 2007;176(8):1091-6.
62. Rothstein HR, Bushman BJ. Publication bias in psychological science: Comment on Ferguson and Brannick (2012). 2012.
63. Rosenthal R. The file drawer problem and tolerance for null results. Psychological bulletin. 1979;86(3):638.
64. Orwin RG. A fail-safe N for effect size in meta-analysis. Journal of educational statistics. 1983;8(2):157-9.
65. Field AP, Gillett R. How to do a meta-analysis. British Journal of Mathematical and Statistical Psychology. 2010;63(3):665-94.
66. Hedges LV. Modeling publication selection effects in meta-analysis. Statistical Science. 1992;7(2):246-55.
67. Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. Psychometrika. 1995;60(3):419-35.
68. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. Statistical Science. 1988:109-17.
69. McShane BB, Böckenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspectives on Psychological Science. 2016;11(5):730-49.

70. Vevea JL, Woods CM. Publication Bias in Research Synthesis: Sensitivity Analysis Using A Priori Weight Functions. Psychol Methods. 2005;10(4):428-43.
71. Kepes S, Banks GC, McDaniel M, Whetzel DL. Publication bias in the organizational sciences. Organizational Research Methods. 2012;15(4):624-62.
72. Niemeyer H, Musch J, Pietrowsky R. Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for depression. Journal of Consulting and Clinical Psychology. 2013;81(1):58-74.
73. 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. Systematic Reviews. 2021;10(1):89.
74. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA, editors. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. Proceedings of the 2nd ACM SIGHIT international health informatics symposium; 2012.
75. Polanin JR, Pigott TD, Espelage DL, Grotpeter JK. Best practice guidelines for Abstract screening large-evidence systematic reviews and meta-analyses. Research Synthesis Methods. 2019;10(3):330-42.
76. Lipsey MW, Wilson DB. Practical meta-analysis: SAGE publications, Inc; 2001.
77. R Development Core Team. R: A language and evironment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. p. Retrieved from https://www.r-project.org/.
78. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of statistical software. 2010;36(3):1-48.
79. Coburn KM, Vevea JL, Coburn MKM. Package ‘weightr’. Estimating Weight-Function Models for Publication Bias. 2019.
80. Moreno SG, Sutton AJ, Ades A, Stanley TD, Abrams KR, Peters JL, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC medical research methodology. 2009;9(1):1-17.
81. Bürkner PC, Doebler P. Testing for publication bias in diagnostic meta-analysis: a simulation study. Statistics in medicine. 2014;33(18):3061-77.
82. Schwarzer G, Antes G, Schumacher M. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. Statistics in medicine. 2002;21(17):2465-77.