ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

ΠΜΣ «Μεθοδολογία Βιοϊατρικής Έρευνας, Βιοστατιστική και Κλινική Βιοπληροφορική»

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Μέθοδοι εύρεσης του σφάλματος δημοσίευσης στην μετα-ανάλυση»

«Methods to detect publication bias in meta-analysis»

ΓΕΩΡΓΙΟΣ ΚΑΡΑΜΑΛΗΣ

ΙΑΤΡΟΣ

ΜΑΙΕΥΤΗΡΑΣ-ΓΥΝΑΙΚΟΛΟΓΟΣ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ:

- **ΔΑΡΔΙΩΤΗΣ ΕΥΘΥΜΙΟΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ ΝΕΥΡΟΛΟΓΙΑΣ (ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ)**
- **ΣΤΕΦΑΝΙΔΗΣ ΙΩΑΝΝΗΣ, ΚΑΘΗΓΗΤΗΣ ΝΕΦΡΟΛΟΓΙΑΣ**
- **ΔΟΞΑΝΗ ΧΡΥΣΟΥΛΑ, ΙΑΤΡΟΣ, ΑΚΑΔΗΜΑΪΚΗ ΥΠΟΤΡΟΦΟΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΒΙΟΜΕΤΡΙΑ**

ΜΑΪΟΣ 2023

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

ΠΕΡΙΕΧΟΜΕΝΑ

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

ΠΕΡΙΛΗΨΗ

Η μετα-ανάλυση έχει καθιερωθεί ως ένα δυνατό στατιστικό εργαλείο το οποίο συνθέτει δύο ή περισσότερες μελέτες ενός κοινού φαινομένου, με το να συγκεντρώνει τα αποτελέσματα και να εκτιμά ένα μέσο μέγεθος επίδρασης. Για να τύχει γενικής εφαρμογής αυτό το μέγεθος επίδρασης, ο πληθυσμός των μελετών θα πρέπει να είναι σωστά ορισμένος και, πιθανόν πιο σημαντικό, το σύνολο των μελετών που περιλαμβάνονται στην μετα-ανάλυση, μετά από μια συστηματική ανασκόπηση της βιβλιογραφίας, θα πρέπει να είναι περιεκτικό ή τουλάχιστον αντιπροσωπευτικό αυτού του πληθυσμού.

Εάν η πιθανότητα για να δημοσιευθεί μια μελέτη επηρεάζεται από την στατιστική σημαντικότητα ή την κατεύθυνση των αποτελεσμάτων της, τότε ανακύπτει το σφάλμα δημοσίευσης. Αυτό μπορεί να επηρεάσει την εγκυρότητα και την επαναληψιμότητα των αποτελεσμάτων.

Σε αυτήν την εργασία αναζητήσαμε μεθόδους οι οποίες αναπτύχθηκαν για να ανιχνεύσουν την παρουσία του σφάλματος δημοσίευσης και, ενδεχομένως, να διορθώσουν το μέγεθος επίδρασης, συνοδευόμενες από τα πλεονεκτήματα και μειονεκτήματά τους.

Πραγματοποιήθηκε μια συστηματική αναζήτηση σε βάσεις δεδομένων όπως PubMed, Cochrane Library, ακόμα και στο Google Scholar.

Οι μέθοδοι αυτές κατηγοριοποιήθηκαν σε τρεις ομάδες, α) μέθοδοι Fail-safe N (αποτυχίας και ασφάλειας), β) βασισμένες σε διαγράμματα, στατιστικές δοκιμασίες για την ασυμμετρία και ανάλυση μετα-παλινδρόμησης και γ) μέθοδοι επιλογής.

Παρά την σημαντική πρόοδο σε αυτό το πεδίο, καμία μέθοδος δεν υπερτερεί των άλλων σε όλα τα είδη μελετών και υπάρχει ανάγκη να αναπτυχθούν νέες μέθοδοι, αλλά επίσης θα πρέπει να δοθεί βάρος στις πολιτικές έρευνας και δημοσίευσης.

Λέξεις-κλειδιά: μετα-ανάλυση, σφάλμα δημοσίευσης, μελέτη προσομοίωσης, ανασκόπηση.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

ABSTRACT

Meta-analysis has been established as a powerful statistical tool which synthesizes two or more studies of a common phenomenon by pooling the results and estimating an average effect size. In order to generalize this effect, the population of studies should be correctly defined and, maybe most important, the set of studies included in the meta-analysis, after a systematic review of the literature, should be comprehensive or at least representative of this population.

When the probability of a study getting published is affected by the statistical significance or the directionality of its results, then publication bias emerges. This can affect the validity and reproducibility of the results.

In this essay, we looked for methods developed in order to detect the presence of and even adjust for publication bias, along with advantages and disadvantages. A systematic search was performed in databases such as PubMed, Cochrane Library, even Google Scholar.

These methods were categorized in three groups, a)Fail-safe N methods, b) graph-based methods and statistical tests for the asymmetry and meta-regression analysis, c) selection methods.

Despite the great advance in this field, no one method is optimal across all settings and it is needed to develop new methods, but also there should be a focus on the researching and publishing policies.

Keywords: meta-analysis, publication bias, simulation studies, review

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Methods to detect publication bias in meta-analysis

Introduction

Meta-analysis is the method which combines usually conflicting evidence from different studies performed on a particular topic, mainly clinical trials evaluating the effectiveness of therapies or tests. Meta-analysis is defined as “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” [1], it pools their results to report an overall effect size, an overall Confidence Interval and an overall p-value. The derived pooled evidence can provide much more precise and reliable answer than any single study can. We also assess the robustness of the results, search for heterogeneity patterns and test hypotheses on how effects differ through sensitivity analyses [2].

Meta-analysis is based on a rigorous systematic review, considering that the data collected are comprehensive, or representative of the field under examination. However, relevant study results could be missing due to selective publication and non or insufficient dissemination. Even a most comprehensive search is likely to miss study data which is not published at all (supplemental unpublished data related to published trials, data obtained from regulatory authorities or post marketing analyses hidden from the public). Additionally, study data not published in conventional journals, the so-called grey literature [3], is not indexed in electronic data bases and likely not to be identified (print or electronic information not controlled by commercial or academic publishers, including non-indexed conference abstracts frequently published in journal collections, dissertations, press releases, government reports, policy documents, book chapters or data obtained from trial registers) [4].

This absence could introduce bias in the results of the meta-analysis, which can lead to under-estimation or over-estimation of the true intervention effect, varying in magnitude and direction. It may distort scientists’ perception of the existing evidence, time and research funds are wasted (since researchers have chosen not to publish or partially publish their results, because of “non-significance”), clinicians’ decisions and recommendations about therapies and interventions will be based on non-well documented evidence, influence government policies and, the most important, expose patients to unnecessary and preventable risks [4].

This phenomenon is known as ‘publication bias’ which exists when the probability of a study getting published is affected by its results. The probability depends on the direction and the statistical significance of their findings, with being high when there is statistical significance (mainly based on a p-value <0.05). On the contrary, studies with negative or neutral results are less favorable to be published, so they are missing in the meta-analysis data set [5][6] .

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Publication Bias is considered only one type of Reporting Bias, as there are several other types that affect the publication process and finally distort the obtained evidence from meta-analysis [7].

Type of reporting bias	Definition
Publication bias	The *publication* or *non-publication* of research findings, depending on the nature and direction of the results.
Time-lag bias	The *rapid* or *delayed* publication of research findings, depending on the nature and direction of the results.
Language bias	The publication of research findings *in a particular language*, depending on the nature and direction of the results.
Citation bias	The *citation* or *non-citation* of research findings, depending on the nature and direction of the results.
Multiple (duplicate) publication bias	The *multiple* or *singular* publication of research findings, depending on the nature and direction of the results.
Location bias	The publication of research findings in journals with different *ease of access* or *levels of indexing* in standard databases, depending on the nature and direction of results.
Selective (non-) reporting bias	The *selective reporting* of some outcomes or analyses, but not others, depending on the nature and direction of the results.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

These types of non-reporting biases make it hard for the researchers to find existing evidence. Furthermore, one should take account Questionable Research Practices (QPRs) which researchers may have been applied when analyzing and reporting their findings ("researcher's degree of freedom", p-hacking, i.e. tweaking analyses till $p<0.05$ is reached, HARKing, i.e. hypothesizing after results are known by running various tests on a dataset and then invent hypotheses for the significant ones) [2].

It is clear that publication bias, the other types of reporting bias, along with QPRs, constitute a deleterious threat for the validity of the meta-analytic results.

These reporting biases have been recognized for centuries (for example Robert Boyle 1671, Ferriar and other scientists by the 18th century, The Boston Medical and Surgical Journal August 1909, Bradford Hill 1959) [5][8]. It was the accumulation of studies over time in different scientific domains and the development of statistical methods that allowed for the formal study of publication bias (PB), its existence and effects.

The first studies on PB comes from Sterling in 1959 [9], from the field of Psychology, and Smart in 1964 [10],from the field of Education, who reported over-representation of published studies rejecting the null hypothesis and a lack of replicated studies (Sterling reported the same on an updated review 30 years later [11]) . Greenwald in 1975[12] presented a review against the prevailing attitudes on the null hypothesis among behavioral scientists, that is only findings which reject the null hypothesis could advance science. The misuse of P values in the early-mid 20th century produced false-positive and non-replicable results, a phenomenon which persists in the 21st century [13].

Since then, there was an increasing awareness and extensive work with strong evidence that PB exists in social and biomedical sciences. But also there was a need for statistical methods to detect and assess the publication bias and even correct the effect size, should PB is present.

The first approach for dealing with PB was introduced by Rosenthal in 1979 [14], the famous file-drawer number or Fail-safe N method.

Afterwards, an extensive body of research dealing with methods for detecting, quantifying and adjusting for PB, even in the broader definition of dissemination bias, was produced, as it is shown in the following Figure 1 [15].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Figure 1: Timeline of methodological development for publication bias

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Methods

Aiming to reporting the methods for detection, assessment of publication bias and adjustment of effect size, their comparative and/or synthetic use, a thorough search of electronic bibliography in PUBMED, Cochrane databases and in Google Scholar was performed in order to find the relevant articles.

Key words used were: publication bias, meta-analysis, simulation studies, review, using simple, Boolean and Advanced Search methods, trying to discover articles which were reviewing methods for detection of and correction for Publication Bias, exploring their relative effectiveness, their advantages and disadvantages.

Their references were also explored and related articles were retrieved and included.

Articles dealing with recommendations on how to avoid publication bias, or the more extended definition of reporting bias, were also retrieved.

Although search was not limited in terms of year of publication, it was focused mainly on the recent advances in the field during the last 5 years.

Taking also account the elaboration of computational and programming methods, mainly in the last 10 years, there was an effort to retrieve articles which encompassed sections or reported programming code for the corresponding method(s).

Search was limited to articles in English language.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Results

There is a significant number of methods for the detection and/or correction for publication bias reported in bibliography, as pictured in Figure 1. These methods are generally divided into three main categories:

1. Fail-safe N methods
2. Methods based on graphs (Funnel plot, Contour Enhanced, Meta-plot, statistical tests for funnel plot asymmetry)
3. Selection models

- **1st category** includes the first attempt to deal with the problem of PB. Rosenthal introduced his "file-drawer" concept [14], in which only statistically significant results are published and non-significant were remaining in the "drawer".

Under the extreme assumption that the true effect is null and that journals are filled with the 5% of studies with Type-I errors, he developed a formula for estimation of the number N of these non-significant articles needed to average the null effect, to reduce the overall effect to null. The fail-safe N number of these studies is calculated by the formula:

$$N > k\left[\frac{Z_S}{Z_\alpha}\right]^2 - k.$$

Where k is the number of studies included in the analysis, Z_α is the α level upper tail critical value of the normal distribution (for α=0.05, one-tailed $Z_\alpha = 1.645$). Z_s is the sum of z-scores corresponding to the observed p-values divided by the square root of k.

$$Z_S = \frac{\sum_{i=1}^{k} z_i}{\sqrt{k}},$$

Rosenthal argued that if N is large relative to k, the results of the meta-analysis may be considered robust to publication bias. He also proposed a 'rule of thumb' that raises concerns if $N < 5k + 10$.

Orwin[16] proposed a fail-safe N based on true effect size different from the null and Rosenberg[17] made a modification which accounts for weighting of the observed or unpublished studies by study-size.

Other modifications by Glacier and Olkin [18] tried to estimate the numbers of unpublished articles that may exist, based on selection modeling approach, using the p-values observed in the studies and assuming that the null hypothesis is correct.

A Bayesian hierarchical selection model was also proposed by Eberly and Casela [19], for the distribution of total number of studies, both published and unpublished, dependent on the probability of publication, assuming all studies significant at level α are published, while non-significant studies are published with a selection probability ρ.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

The first fail-safe N methods lack a statistical model or distributional assumptions for the unpublished data and there is no clear-cut and justifiable statistical criterion for what consists of a large fail-safe N.

The latter 2 methods based on selection models are complex and dependent on their assumptions, mainly the null hypothesis and the prior distribution of the probability of publication.

All these fail-safe N methods lead to widely varying and conflicting N numbers of additional studies, so their use is limited and tends to be abandoned, as they are not recommended [5][20].

- **The 2nd category** comprises of graph-based tools and statistical tests assessing their asymmetry and meta-regression to adjust for publication bias.

Funnel plot is an intuitive and easy to implement tool, frequently used in meta-analyses for the investigation of publication bias. It was first used by Light and Pillemer in 1984 [21], in educational and psychology research.

They are scatter plots of effect sizes estimated from individual studies on the x-axis, against a measure of precision on the y-axis (standard error, inverse standard error, sample size). Since the precision in the estimation of treatment effect increases as the sample size of a study increases, results from small studies will spread at the bottom of the graph and effect sizes scatter more heavily to the left and right of the pooled effect. The spread is narrowing for larger, or more powerful, studies, towards the top of the plot, not far away from the pooled effect size.

A triangular region is also plotted, within which 95% of studies would be expected to lie in the absence of both biases and heterogeneity.

If bias is absent (and studies estimate the same underlying effect) the plot resembles a symmetrical inverted funnel.

If bias exists (for example if smaller studies with statistically non-significant effects are unpublished), the funnel plot appears asymmetrical, with a gap in the bottom right side of the graph [7][22][23].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

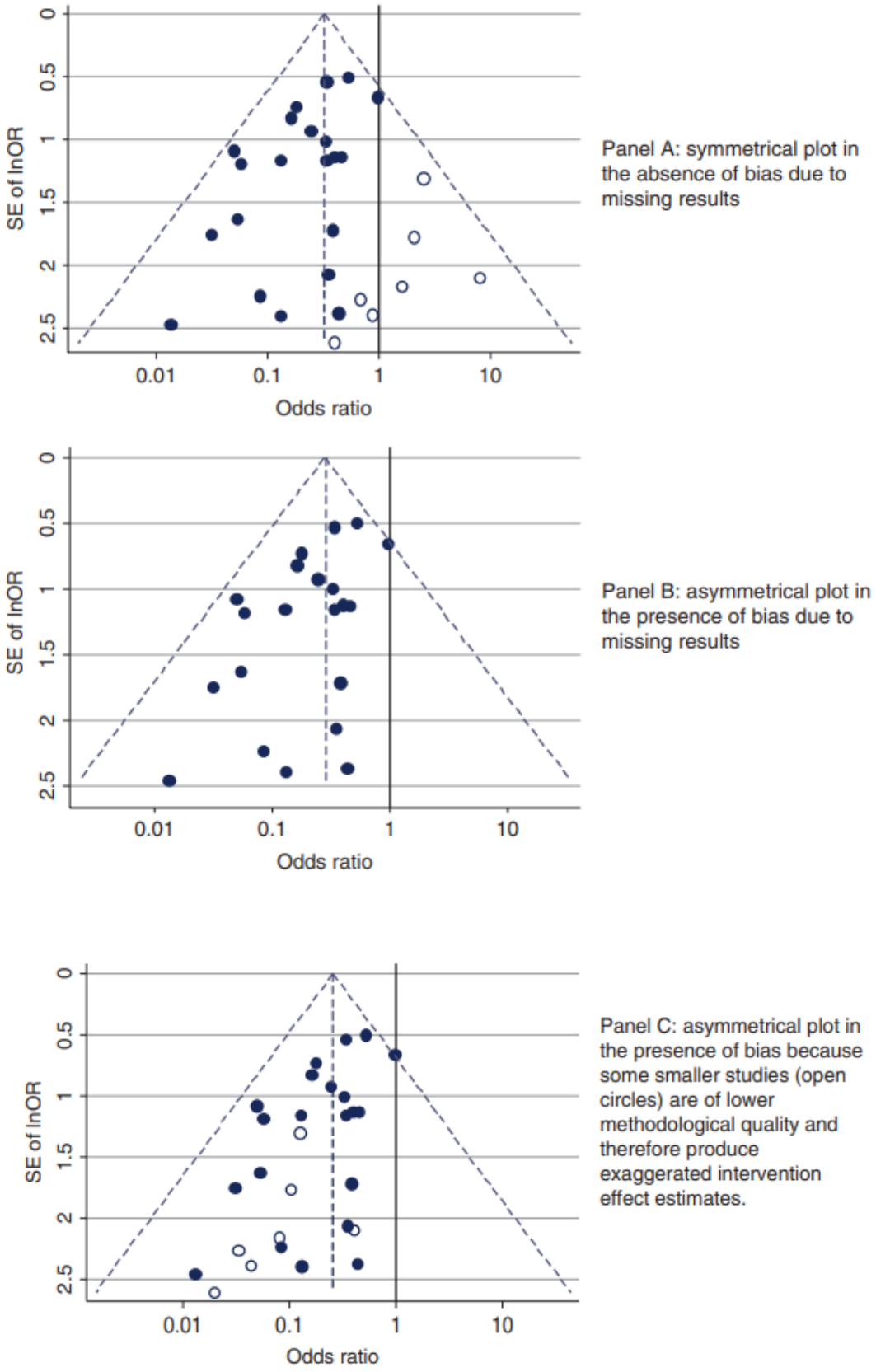

Figure 2. Funnel plot**,** adapted from Cochrane Handbook [7].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Funnel plots were proposed as a means to examine for publication bias. But it is known that small studies of low quality tend to exaggerate their treatment effect, differing from those estimated in larger studies, a phenomenon called “small-study effect” [2][5]. Funnel plot cannot distinguish between publication bias and other sources of asymmetry and it is now considered as a method to inspect the small-study effects, keeping in mind that asymmetry of the funnel plot may be attributed to other reasons, even to pure chance, as shown in the following table [7].

Table 13.3.b Possible sources of asymmetry in funnel plots. Adapted from Egger et al (1997)

1) Non-reporting biases
- Entire study reports, or particular results, of smaller studies are unavailable because of the nature of the findings (e.g. statistical significance, direction of effect).

2) Poor methodological quality leading to spuriously inflated effects in smaller studies
- Trials with less methodological rigour tend to show larger intervention effects (Page et al 2016a). Therefore, trials that would have been ‘negative’, if conducted and analysed properly, may become ‘positive’. Asymmetry can arise when some smaller studies are of lower methodological quality and therefore produce larger intervention effect estimates (Figure 13.3.b, Panel C).

3) True heterogeneity
- Substantial benefit may be seen only in patients at high risk for the outcome that is affected by the intervention, and usually these high-risk patients are more likely to be included in small, early studies (Davey Smith and Egger 1994).
- Some interventions may have been implemented less thoroughly in larger trials and may, therefore, have resulted in smaller estimates of the intervention effect (Stuck et al 1998).

4) Artefactual
- Some effect estimates (e.g. odds ratios and standardized mean differences) are naturally correlated with their standard errors, and this can produce spurious asymmetry in a funnel plot (Sterne et al 2011, Zwetsloot et al 2017).

5) Chance

The interpretation of a funnel plot is visual and subjective, may be erroneous and interobserver variability is also expected. There is also a question regarding the number of studies in a meta-analysis required before using funnel plots. Informally, it is unlikely that funnel plots are useful in meta-analyses containing a small number of studies (e.g. < 10) [24][25][26].

The need for more objective methods for examining and interpreting asymmetry led to the development of modifications of the funnel plot and statistical tests for correlation between observed effect sizes and precisions.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

The contour-enhanced funnel plot was proposed by Peters et al. in 2008 [27] to counteract the limitations of the funnel plot. It differs in two ways, first it is centered at zero whereas the funnel plot is centered at the meta-analytic effect size estimate. Second, contour lines are added corresponding to the p-values of studies (dark gray two-tail p-values 0.05-0.1, gray 0.01-0.05 and outside the funnel 0 and 0.01). These lines help distinguishing asymmetry caused by publication bias or from other causes (because they show whether statistically significant studies are missing in the meta-analysis). If studies appear to be missing in areas where results would be statistically non-significant, then this is an indication that the asymmetry is due to reporting biases. Conversely, if the supposed missing studies are in areas where results would be statistically significant and favorable to the experimental intervention, this would suggest the cause of the asymmetry is more likely to be due to factors other than reporting biases.

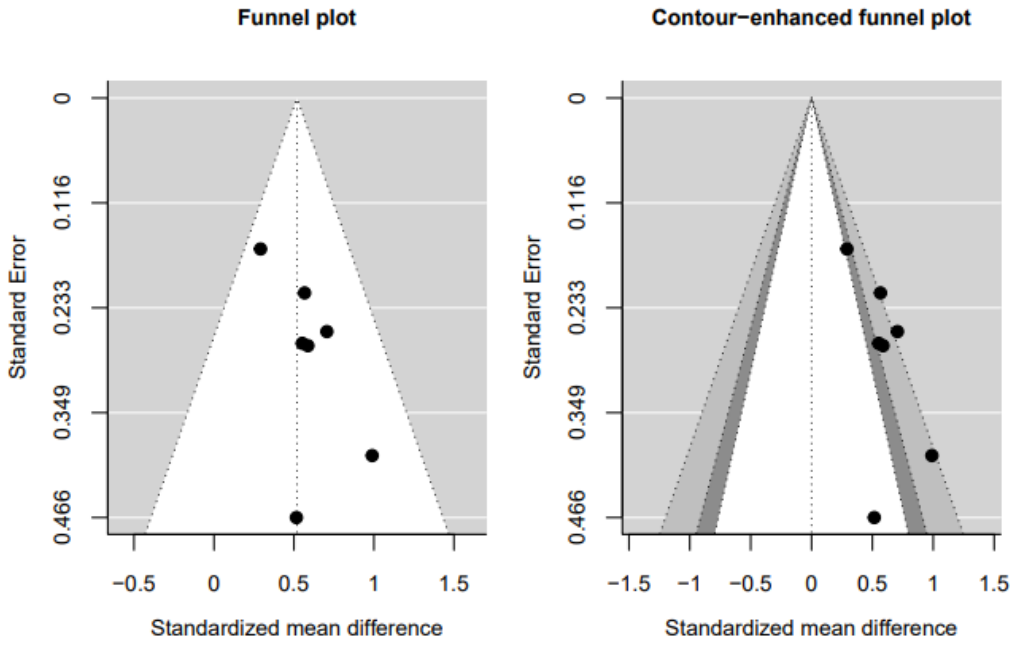

Figure 3. Contour-enhanced funnel-plot (adapted from Avoiding Questionable Research Practices in Applied Psychology [20]).

A new graphical display was examined by Furuya-Kanamori et al in 2018 [28], the Doi plot, to visualize asymmetry and a new measure (LFK index) to detect and quantify study asymmetry of study effects in Doi plots. They demonstrated a better visual representation of asymmetry for the Doi plot when compared to funnel plot and LFK index outperformed Egger's p-value for detection of asymmetry (the Doi plot and the LFK index have been implemented into MetaXL version 5.3, an add-in for Microsoft Excel that can be freely downloaded from www.epigear.com).

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Another graphical method that was recently proposed to assess publication bias in a meta-analysis is the meta-plot (Van Assen et al. 2020) [20][29]. It shows the precision of a study (i.e., reciprocal of its standard error) on the x-axis and the effect size on the y-axis. The circles in the meta-plot are the average effect size estimates of a cumulative random-effects meta-analysis (in a cumulative meta-analysis multiple meta-analyses are conducted, where the first meta-analysis is based on a single study and in each subsequent meta-analysis a study is added). The order of the studies being added to the cumulative meta-analysis in the meta-plot is based on studies' precision. The rightmost dot is the meta-analysis based on only the study that is most precise and the leftmost dot is the meta-analysis based on all studies. Each dot is accompanied by its 95% CI.

The meta-plot in this figure shows a decreasing trend in the cumulative meta-analysis from left to right. This is indicative for small-study effects, because the average effect size estimate of the meta-analysis based on all studies is larger than meta-analyses based on more precise studies. An advantage of the meta-plot over the funnel plot is that small-study effects are more visible as the effect size in the plot refers to the results of meta-analyses rather than individual studies.

The meta-plot also contains other relevant information for meta-analysts. First, it states the percentage of statistically significant results in the meta-analysis [71.4% in the meta-analysis of Cowlishaw et al. (2012)].

Second, it shows information about the statistical power of the studies in the meta-analysis at the top of the plot. The leftmost percentage indicates the percentage of studies whose statistical power was insufficient (less than 80%) to detect a large population effect. The remaining three percentages at the top of the plot describe the percentages of studies with sufficient statistical power to detect a large (L), medium (M), and small (S) effect, respectively.

Finally, the asterisks in the meta-plot refer to the expected estimates in the cumulative meta-analysis if the population effect size is zero, combined with extreme publication bias (i.e., only statistically significant studies get published). Asterisks that are larger than the dots imply that the results of the meta-analysis can also be explained by extreme publication bias in combination with no effect. This is the case for the meta-plot in Figure 4, so authors are recommended to be cautious when interpreting the results of this meta-analysis.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

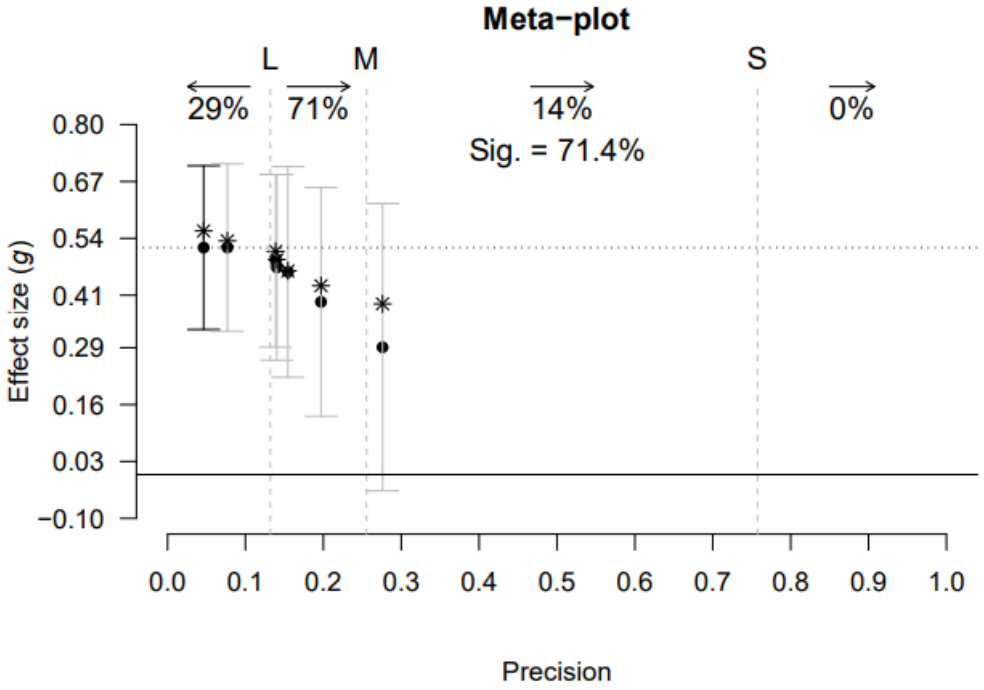

Figure 2: Meta-plot of the meta-analysis by Cowlishaw et al. (2012).

Figure 4. Meta-plot (adapted from Avoiding Questionable Research Practices in Applied Psychology [20]).

Due to the previously mentionned drawbacks of funnel plots, mainly its visual and subjective interpretation, the need for a more accurate and objective appraisal of its asymmetry led to the development of statistical tests in order to quantify it.

These studies examine if the association between estimated effect sizes and a measure of study size, mainly its precision (the standard error of the effect or the inverse of this standard error), is greater than expected to occure by chance.

Begg and Mazumbar in 1994 [30] first described an adjusted rank method to examine the association between the effect estimates and their sampling variances (this method is not recommended due to its low power-less than Egger's) [5][31].

Since then, a plethora of tests is proposed in order to examine the presence of asymmetry, including methods to estimate effect size in the presence of publication bias. Like the visual inspection of funnel plot, these tests identify small-study effects and not tell us if publication bias exists.

Then, it was Egger et al [32] that introduced in 1997 a linear regression approach in which the standard normal deviate z_i (defined as $z_i = \theta_i / s_i$) is regressed against its precision $prec_i$ ($prec_i$ defined as = $1 / s_i$), denoting the intervention effect estimate, e.g. standardized mean difference or log odds ratio, from study i as θ_i, and its corresponding variance and standard error as v_i and s_i respectively [9].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

$$E[z_i] = \beta_0 + \beta_1 prec_i.$$

A test of null hypothesis β_0 (the intercept) = 0 can be derived from the regression output using statistical packages, as a t-test.

A significant level of 0.1 is recommended for hypothesis testing [32].

In the absence of funnel plot asymmetry, the points in a plot of z_i against $prec_i$ will scatter about a line which runs through the origin at standard normal deviate zero, since the intercept $\beta_0 = 0$, with the slope β_1 indicating the size and direction of effect.

If there is funnel asymmetry, the regression line will not run through the origin and the intercept provides a measure of asymmetry, the larger its deviation from zero, the more pronounced the asymmetry.

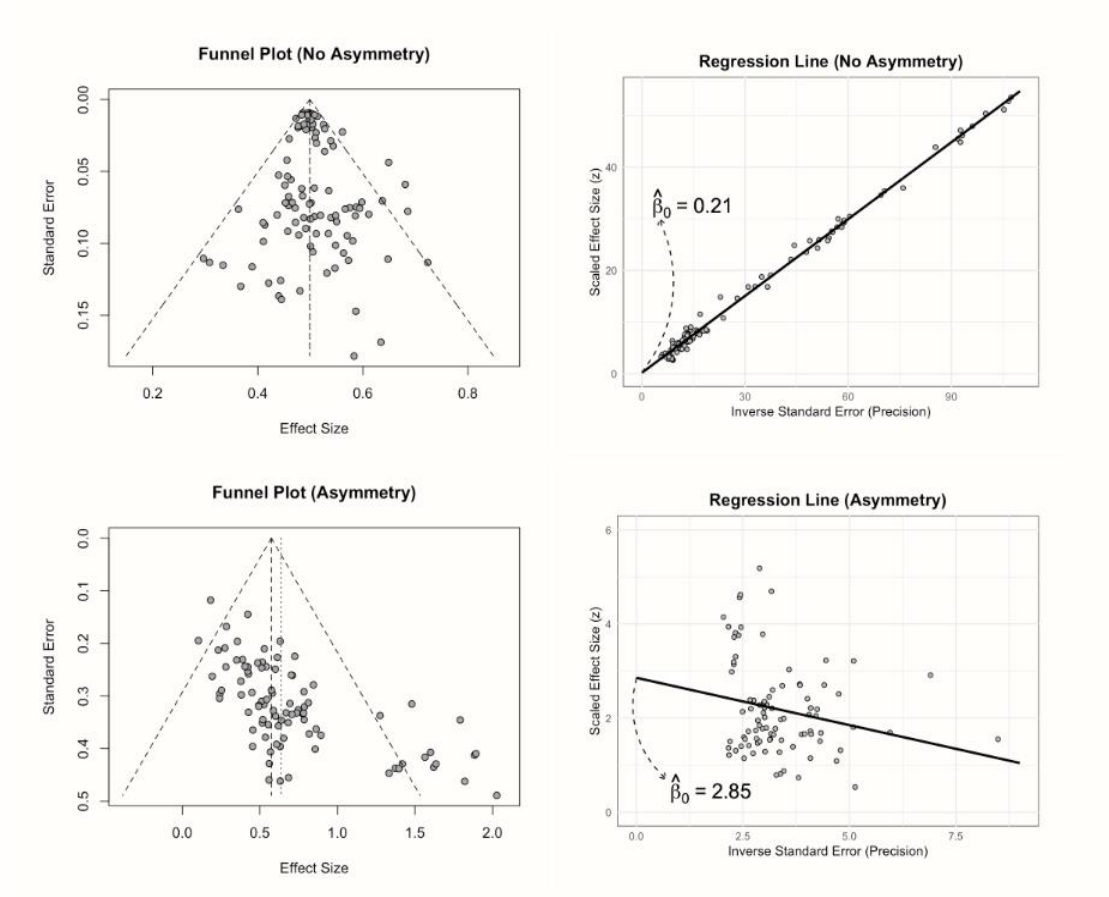

Figure 5. Regression line (adapted from Doing Meta-Analysis with R: A Hands-On Guide [2])

Egger's test has been extensively studied for binary outcomes, but not for continuous ones. Continuous outcomes are commonly measured on an absolute (mean) difference scale, and it is not uncommon for the magnitude of effect to be related to response in the control arm (i.e. baseline risk). When this is the case, funnel plots can appear highly asymmetric, even when publication bias is not present since correlations between

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

outcome and both effect size and its standard error exist, so Egger's test is potentially misleading with inflation of false positive results (inflated Type I error) [33].

Two methods are proposed to address this problem, the first by Pustejovsky and Rodgers in 2019 [34], based on either a simple modification to the conventional standard error formula or a variance-stabilizing transformation, minimizing Type I errors.

The second method was proposed by Doleman, Sutton et al in 2020 [35], a test which regresses the residuals from a meta-regression model, including baseline risk as a study-level covariate, against inverse sample size, showing better statistical properties.

Egger's work was followed by a variety of modified statistical tests, due to its limitations (low power, inflated type I errors, problematic interpretation when heterogeneity between studies exists or there is small number of studies) [15] . Some of these methods are presented in the following list (adapted from https://handbook-5-1.cochrane.org/):

Proposed tests for funnel plot asymmetry

Reference	**Basis of test**
All outcomes	
Begg and Mazumdar (1994)[3]	Rank correlation between standardised intervention effect and its standard error
Egger et al (1997)[4]	Linear regression of intervention effect estimate against its standard error, weighted by the inverse of the variance of the intervention effect estimate
Tang and Liu (2000)[5]	Linear regression of intervention effect estimate on $1/\sqrt{N_{tot}}$, with weights N_{tot}
Dichotomous outcomes only	
Macaskill et al (2001)[6]*	Linear regression of intervention effect estimate on N_{tot}, with weights $S \times F/N_{tot}$
Deeks et al (2005)[7]*	Linear regression of log odds ratio on $1/\sqrt{ESS}$ with weights ESS, where effective sample size $ESS = 4N_E \times N_C / N_{tot}$
Harbord et al (2006)[8]*	Modified version of the test proposed by Egger et al, based on the 'score' (O–E) and 'score variance' (V) of the log odds ratio
Peters et al (2006)[9]*	Linear regression of intervention effect estimate on $1/N_{tot}$, with weights $S \times F/N_{tot}$
Schwarzer et al (2006)[10]*	Rank correlation test, using mean and variance of the non-central hypergeometric distribution
Rücker et al (2008)[11]	Test based on arcsine transformation of observed risks, with explicit modelling of between-study heterogeneity

N_{tot} is the total sample size, N_E and N_C are the sizes of the experimental and control intervention groups, S is the total number of events across both groups and $F=N_{tot} - S$.

These modifications improved type I errors but other issues persist. Their power is an issue, mainly when the sample size is small, they also may lead to incosistent conclusions as no one test is optimal in all meta-analytic settings [15][36][37].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Then, in 2011 Sterne and a vast group of experts (including writers of the proposed methods) published an article proposing recommendations regarding the implementation of these tests [31].

Box 2: Recommendations on testing for funnel plot asymmetry

All types of outcome

- As a rule of thumb, tests for funnel plot asymmetry should not be used when there are fewer than 10 studies in the meta-analysis because test power is usually too low to distinguish chance from real asymmetry. (The lower the power of a test, the higher the proportion of "statistically significant" results in which there is in reality no association between study size and intervention effects). In some situations—for example, when there is substantial heterogeneity—the minimum number of studies may be substantially more than 10
- Test results should be interpreted in the context of visual inspection of funnel plots— for example, are there studies with markedly different intervention effect estimates or studies that are highly influential in the asymmetry test? Even if an asymmetry test is statistically significant, publication bias can probably be excluded if small studies tend to lead to lower estimates of benefit than larger studies or if there are no studies with significant results
- When there is evidence of funnel plot asymmetry, publication bias is only one possible explanation (see box 1)
- As far as possible, testing strategy should be specified in advance: choice of test may depend on the degree of heterogeneity observed. Applying and reporting many tests is discouraged: if more than one test is used, all test results should be reported
- Tests for funnel plot asymmetry should not be used if the standard errors of the intervention effect estimates are all similar (the studies are of similar sizes)

Continuous outcomes with intervention effects measured as mean differences

- The test proposed by Egger et al may be used to test for funnel plot asymmetry.[1] There is no reason to prefer more recently proposed tests, although their relative advantages and disadvantages have not been formally examined. General considerations suggest that the power will be greater than for dichotomous outcomes but that use of the test with substantially fewer than 10 studies would be unwise

Dichotomous outcomes with intervention effects measured as odds ratios

- The tests proposed by Harbord et al[26] and Peters et al[27] avoid the mathematical association between the log odds ratio and its standard error when there is a substantial intervention effect while retaining power compared with alternative tests. However, false positive results may still occur if there is substantial between study heterogeneity
- If there is substantial between study heterogeneity (the estimated heterogeneity variance of log odds ratios, τ^2, is >0.1) only the arcsine test including random effects, proposed by Rücker et al, has been shown to work reasonably well.[28] However, it is slightly conservative in the absence of heterogeneity and its interpretation is less familiar than for other tests because it is based on an arcsine transformation.
- When τ^2 is <0.1, one of the tests proposed by Harbord et al,[26] Peters et al,[27] or Rücker et al[28] can be used. Test performance generally deteriorates as τ^2 increases.

Lin in 2020 [37] proposed a hybrid test that incorporates the strenghts of all these tests to maximize power across different settings. The statistic test is based on a set of tests T to detect PB and P_x the p-values of these tests. Then, through the resampling method, the test statistic is calculated as the minimum of p-values of these tests.

$$T_{\text{hybrid}} = \min_{X \in \mathcal{T}} P_X$$

It does not require to choose a single publication bias test from a large pool of candidates and draw a conclusion based entirely on this single test; it permits them to combine various candidates into synthesized evidence for evaluating publication bias.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Although powerful, it may have limitations, mainly heterogeneity, contamination by tests that have poor performance and it does not adjust for the bias. The article includes the R code for its implementation.

All these tests examine the presence of small-study effects in a meta-analysis (as a proxy that may point to publication bias). The reserchers are also interested in the magnitude of the bias (is a slight or a massive enough to change the interpretation of their data). In order to explore the impact of these biases on the results of a meta-analysis, correction methods were intorduced to estimate effect size in the presence of publication bias.

The most often used correction method is the non parametric trim-and-fill (Duval and Tweedie in 2000) [38]. It is an iterative procedure that trims the most extreme effect sizes from the right-hand side of the funnel plot and the pooled effect is recalculated without them.

The recalculated pooled effect is now assumed to be the center. For each trimmed study, one additional effect size is added, mirroring its results on the other side of the funnel (if now the center is 0.5 and the trimmed has an effect of 0.8, the mirrored study will be given an effect of 0.20). Based on all data, trimmed and imputed effect sizes, the average effect is recalculated, using a random-effects model. The result is an estimate of the corrected pooled effect size.

It can lead to inappropriate adjustment in the presence of heterogeneity and result in inflated summary estimate because it imputes studies with the most extreme values. It is built on the strong assumption that the funnel plot should be symmetric and the imputed studies are ‘fictional’. For these reasons, the results should be interpreted cautiously and the method be used for sensitivity analysis to assess the impact of missing studies, rather than as a means of adjusting results per se [15].

Lin and Chu [39] developed a skewness measure for quantifying asymmetry in a funnel plot. They used study-specific standardized deviates from the mean (skewness=0 means absence of asymmetry). Even though it has high power, as with the graph-based methods, publication bias is only one source of asymmetry and absence of skewness does not necessarily imply symmetry.

Zhu et al [40] proposed in 2018 a parametric approach for estimation of the adjusted treatment effect and the severity of publication bias by formulating study omission as truncation of a normal distribution, as studies with effect sizes below a certain threshold or p-values above are truncated. They derived estimators for the overall mean and the truncation proportion using maximum likelihood estimation and method of moments for fixed and random-effects models, respectively. The simulation studies performed consistently well, especially compared to trim-and-fill method. This method is susceptible to outliers, complicated in implementing for the non-statistician, but it can provide distinction between heterogeneity and publication bias, when formulated as a random-effects model.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Meta-regression methods have also been proposed in order to obtain an adjusted summary effect size, while accounting for true test heterogeneity through the inclusion of study-level covariates. Methods proposed by Weinhandle-Duval [41] and Moreno[42] displayed promising results compared to trim-and-fill, but they suffered under high levels of heterogeneity.

Rucker et al [43] provided the limit meta-analysis that better distinguishes between small-study effects and heterogeneity, by diminishing within-study variability but retaining between-study variability.

From the field of economics, Stanley and Doucouliagos (Deakin University, Victoria, Australia) have produced significant work in the field of meta-analysis and the detection and correction of publication bias (since this problem seems still significant for social sciences). They have published a series of papers, the most popular of which is on FAT-PET-PEESE approach, a family of meta-regression tests for funnel plot asymmetry similar to Egger's test (FAT), the precision-effect test (PET) for effects adjusted for publication bias when the true effect is zero, the precision effect estimate with standard error (PEESE) if the true effect estimate is not zero. Finally, they combined these tests in PET-PEESE test which is aimed at small-study effects (as a potential indicator of publication bias) [2][44].

The effect size estimates of PET and PEESE are the values where the slope of the regression line is 0 (i.e., the estimate of the intercept).

Limitations of the method are that it actually corrects the effect size for small-study effects rather than publication bias. Hence, the method becomes biased if there is large heterogeneity in a meta-analysis. Moreover, applying the method is also discouraged if there are less than 10 studies in the meta-analysis or the precision of the studies is similar, because this makes it difficult to fit the regression line and results in an imprecise estimate [2][45].

Recently, with the collaboration of Ioannidis et Carter [46], they introduced and evaluated three tests for publication selection bias based on excess statistical significance (ESS). The test of excess statistical significance (TESS), the proportion of statistical significance test (PSST) and their combination (TESS-PSST), are found to be better at detecting publication selection bias than the conventional alternatives. Specifically, they have higher power to detect publication selection than Egger-type tests, they accommodate heterogeneity and low average rate of false positives. Code in R is reported in the Supporting Information of the article.

Generally, for tests examining funnel plot asymmetry, regression-based methods to estimate the effect of intervention should be used only when there are sufficient studies (at least 10) to allow appropriate estimation of the regression line, as stated in Cochrane Handbook for Systematic Reviews of Interventions (2020) [7].

All these methods assess the risk of publication bias by looking at small-study effects, capturing its mechanism indirectly. They assume that publication bias is driven by

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

effect sizes, depending on their sample size, considering that studies with higher standard error (thus a lower precision) have higher average effect sizes than larger studies. This results to publication of only small studies with significant results, while the others remain in the "file drawer" [2].

- The 3rd category of techniques are the so-called selection methods which assess and adjust for publication bias relating to the size, direction and statistical significance of study results. They model any kind of process through which publication bias may have affected the results.

The idea behind all selection models is to specify a distribution which predicts, based on simple or highly sophisticated hypotheses, how it is that some study is published ("selected"), depending on its results, usually the study's p-value. The selection model can be seen as a function that returns the probability of publication for different values of p. This function can be used to derive a corrected estimate of the true effect size [2].

The statistical model underlying any kind of selection method consists of two components:

a) The data (effect size) model which describes how the data is generated in the absence of publication bias. It is described by the function $f(x_k)$, identical to the random-effects model. It assumes that the observed effect sizes θ_k are normally distributed around an average effect μ and deviate from μ due to sampling error and between-study heterogeneity variance τ^2. Knowing μ, τ^2, a study's standard error, and that effect sizes are normally distributed, the function $f(x_k)$ predicts how likely it is to observe some effect size x_k, assuming that there is no publication bias.
b) The selection model describes the publication process, using a wide variety of forms [47].

Yet, when there is publication bias, this effect size distribution, and thus $f(x_k)$ itself, is an incorrect representation of reality. Due to selective publication, some studies are over-represented, presumably those with surprisingly high effect sizes and small samples. There is therefore needed to derive a more "realistic" version of $f(x_k)$, which incorporates the fact that some results had a greater chance of being included than others; that they were given a higher "weight".

This is achieved through a weight function $w(p_k)$. The weight function tells us the selection probability of a study k, depending on its p-value. Based on this, we can define an adapted version of $f(x_k)$, which also incorporates the publication bias mechanism. This function $f^*(x_k)$ is symbolized by this formula:

$$f^*(x_k) = \frac{w(p_k) f(x_k)}{\int w(p_k) f(x_k) dx_k}$$

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

The weight function $w(p_k)$ in this equation represents our assumed selection model and it is often implemented as a step function [2].

When $w(p_k)$ is a step function, this means that values p_k which fall into the same interval are selected with the same probability. This interval-specific selection probability is denoted with ω_i and can differ from interval to interval. The size of the segments is determined by several cut-points (denoted with a_i). The number of cut-points, as well as their exact value, can be chosen by researchers. For example, when $w(p_k)$ contains four segments (and thus four cut-points), it can be defined so:

$$w(p_k) = \begin{cases} \omega_1 & \text{if} \quad 0 \leq p_k \leq a_1 \\ \omega_2 & \text{if} \quad a_1 \leq p_k \leq a_2 \\ \omega_3 & \text{if} \quad a_2 \leq p_k \leq a_3 \\ \omega_4 & \text{if} \quad a_3 \leq p_k \leq a_4 \quad (\text{where} \quad a_4 = 1) \end{cases}$$

For any value of p_k, the function above returns a specific selection probability ω_i, based on the p-value interval into which this value falls. Now a selection model is defined with actual values filled in for the cut-points a_i and selection probabilities ω_i [2].

When a selection model is defined based on a step function, usually only the cut-points a_i are specified. These are the only fixed parameters in the model, while the selection probabilities $\omega=\omega_1, \omega_2, \ldots, \omega_c$ are estimated from the data. Based on the formula in the equation, the selection model can then be fitted to data. This involves using maximum likelihood procedures or Bayesian approaches to jointly estimate ω, as well as a corrected estimate of μ and τ^2 which takes the disparate selection probabilities ω into account. The resulting corrected estimate of μ then represents the true average effect size when controlling for the assumed publication bias mechanism [47][48].

When the selection model is fitted, ω_i is not estimated as an absolute selection probability, but in terms of its relative likelihood of selection. This entails giving the first interval in the step function a reference value of 1, while all other values of ω_i represent the likelihood of selection in relation to this reference group.

Of course, the corrected estimate of the true average effect μ will only be accurate when the selection model itself is appropriate. A rough indication of this is a significant likelihood ratio test (LRT) of the selection model parameters. The test is based on the null hypothesis that there is no selection, and that the relative selection likelihood is identical for all intervals. It should be noted, however, that this significance test has been found to frequently produce anti-conservative results. This means that its results should be interpreted cautiously [2][49].

The first model for study selection was proposed by Hedges in 1984 [50]. It assumes that a) effect sizes are homogenous across studies and effect size estimates are normally distributed with unknown variance and b) only studies with statistically significant results are published. These are the assumptions for the data model and

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

selection model respectively, working as a simple one-parameter likelihood function for an extreme , “worst-case” form of bias.

Iyengar and Greenhouse in 1988 [51] generalized Hedges’ approach to allow for a less strict selection model for the publication process that comprises the publication of studies with results both significant and non-significant.

They considered two weight functions for the selection model, a one-parameter function that implies the relative likelihood that a nonsignificant study is published increases as those results approach statistical significance and a one-parameter step function that implies that the relative likelihood is constant. The data model was also expanded to accommodate effect sizes that are heterogeneous across studies.

Next step involved random-effects formulations of outcome (data) model and left-continuous step functions for selection weights defined by ranges of p-values incorporated via inverse-probability weighting. The intervals have to be specified and a reasonable choice is to create two intervals, such that statistically significant and non-significant studies are treated differently.

This model with two intervals is sometimes also referred to as the three-parameter selection model [2][47], because three parameters are estimated: the true effect μ, the between-study heterogeneity variance τ^2 and the relative likelihood on the second interval ω_2, which represents the probability that a non-significant result is selected for publication (the relative weight specifying how much less likely a statistically non-significant study is published compared to a significant study). It is also applicable when a small number of studies are included in the meta-analysis. The selmodel function in the {metafor} package in R can be used also for the three-parameter model (the same applies for various kinds of selection models).

Then, a vast corpus of selection methods was introduced modeling the selection process as a function of one-sided p-values, which is used because it preserved information not only about the statistical significance of the results but also for their direction. They are mainly differing on how the weights of the studies are computed, using complex multiparameter weight functions that can approximate any functional form [5][47][48][52].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

A summary of several selection models is shown in the following table [47]:

Article(s)	Data model	Selection model
Hedges (1984)	Effect sizes are modeled as homogeneous across studies. Effect size estimates are modeled as normally distributed with unknown variance (i.e., so that individual study *t* statistics are modeled as noncentral *t* distributed).	Only studies with results that are statistically significant are published.
Iyengar and Greenhouse (1988)	As in Hedges (1984). In the discussion and rejoinder, the data model was conceptually expanded to accommodate heterogeneous effect sizes as in Hedges (1992).	Studies with results that both are and are not statistically significant are published but with different relative likelihoods. The relative likelihood is modeled via one of two simple one-parameter functions.
Dear and Begg (1992); Hedges (1992)	Effect sizes are modeled as heterogeneous across studies via a normal distribution with common mean and common variance. Effect size estimates are modeled as normally distributed with known variance.	Studies with results that both are and are not statistically significant are published but with different relative likelihoods. The relative likelihood is modeled via complex multiparameter functions.
Vevea and Hedges (1995)	Effect sizes are modeled as heterogeneous across studies via a normal distribution with mean that is a linear function of study-level moderators and common variance. Effect size estimates are modeled as normally distributed with known variance.	As in Hedges (1992).
Copas (1999); Copas and Li (1997); Copas and Shi (2001)	As in Hedges (1992).	Studies with results that both are and are not statistically significant are published but with different relative likelihoods. The relative likelihood is modeled via a linear function that depends on the estimate of the effect size and its standard error.
Simonsohn et al. (2014)	As in Hedges (1984).	As in Hedges (1984).
van Assen et al. (2015)	As in Hedges (1984).	As in Hedges (1984).

Note: The estimation strategy employed by all but the last two articles is maximum likelihood. Simonsohn, Nelson, and Simmons (2014) and van Assen, van Aert, and Wicherts (2015) employed a distance-based estimation strategy based on the Kolmogorov–Smirnov statistic and, in the most recent implementation (van Aert, Wicherts, & van Assen, 2016), the Irwin-Hall distribution, respectively.

Ioannidis and Trikalinos developed TES (test for excess significance) comparing the expected number of significant studies to the observed number of significant studies. This method only works for homogeneous effects and it is recommended by the authors performing the test within subgroups defined by study-level covariates if heterogeneity is present [53].

A distinct type of these method is using p-values, the “p-curve”, the “p-uniform” and its extension “p-uniform*” tests, based on the assumptions identical to those of the original Hedges’ approach, but on alternative estimation strategies (Kolmogorov-Smirnov statistic as the distance metric for p-curve, whereas p-uniform uses a moment estimator based on Irwin-Hall distribution). They involve testing whether the distribution of published p-values significantly deviates from a uniform distribution [2][20].

P-curve tests for right skewness of a uniform shape of p-curve (under the assumption that the true treatment effect is null) as evidence of true treatment effect or left-skewed in case of selective publishing or p-hacking [54].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

The P-uniform test assumes that the distribution of p-values on the true effect size is uniform and a test for deviation from the uniform distribution is a test for publication bias [55].

P-uniform and p-curve have shown to yield accurate estimates in the presence of publication bias and homogeneous true effect size and outperformed the trim-and-fill method, but there is disagreement over their validity when there is between-study heterogeneity, publication of non-significant results and questionable research practices [15].

An extension to p-uniform is p-uniform* introduced by Van Aert and Van Assen in 2018, solving the problem of heterogeneity and including also statistically non-significant studies. The method implicitly assigns different weights to statistically significant and non-significant studies, by considering the likelihood of a study getting published given its statistical (non)significance. An important assumption of p-uniform* is that all statistically significant studies are assumed to be equally likely published and the same holds for all statistically non-significant studies. Even though a simulation study has shown that it is an improvement over p-uniform, researchers are warning to be cautious when publication bias is expected to be extreme with only statistically significant studies in meta-analysis [56][20].

An alternative approach was developed by Copas and colleagues in a series of papers [57][58][59][60]. In their models, study selection depends separately on effect size and its standard error, resulting in more flexible methods, enabling to characterize a wider variety of selection mechanism and model the directionality of the study outcomes on publication probability. In practice, it is not always possible to estimate all of the parameters of these models simultaneously and even when it is, there may be little information about the key parameters. Consequently, as with other selection methods, it is recommended for sensitivity analysis.

Recent work improving identifiability of Copas' model may increase its use in practice, but still investigators need to specify a range of parameter values, for which knowledge of selection models for missing values is important.

Bayesian methods are also have been used for the estimation of the parameters in the models, since 1987 when Bayarri and DeGroot introduced a Bayesian method similar to Hedges's early approach [61]. Several methods have been carried out but in practice, they are too complicated to implement, and there is a lack of accessible software [48]. Sometimes the use of the non-informative priors may be problematic (especially with small number of studies) and the results are sensitive to the choice of prior distribution. It is proposed that sensitivity to the choice of prior distribution should always be assessed [47][48][52].

A recent Bayesian test is based on Copas selection model was introduced, the Robust Bayesian Copas Selection Model [62], which can be implemented with heavy-tailed distributions for the random-effects in the Copas selection model (namely, higher probabilities of extreme values that can have a significant impact on the total), thereby

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

providing robustness against deviations from normality for the study-specific effects. The writers also introduced the D measure for quantifying the magnitude of publication bias, where D quantifies the amount of dissimilarity between a standard random-effects meta-analysis and a meta-analysis done with the Copas selection model.

This method seems to perform well in a variety of simulations and real data settings and the authors also provided an R package, RobustBayesianCopas, for its implementation.

In general, selection methods explicitly specify the data and selection model and they possess some advantages over the methods included in the 2nd category.

- They allow identifiability to be assessed (whether it is possible, even in principle, to estimate the model parameters).
- Maximum Likelihood Estimation strategy has strong theoretical properties, yields standard errors and confidence intervals and allows for hypothesis tests of model parameters
- They can be extended to accommodate very general data and/or selection models, e.g. heterogeneous effect sizes, study-level moderators and other features in the data model as well as different forms of publication bias in the selection model.

Selection methods are relied upon assumptions which are considered idealistic in practice, as real-life data models and selection models are far more complicated, sequential and iterative and involve not only researchers but also editors and reviewers. Both models cannot be well estimated without a large amount of data (that is, a large number of studies) and the results are highly sensitive to the data model and, particularly, the selection model assumed.

For these reasons, mainly, it is advocated that selection models should be used for sensitivity analysis (exploring the range of estimates that result from different forms of and severity of publication bias) [7] [15] [47] [48].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Discussion

It is well recognized that publication bias, that is papers with interesting and statistically significant results, rather than papers with inconclusive outcomes, tend to be accepted for publication, exerts a deleterious effect on meta-analytic results, as their magnitude may be falsely manifested as important (along with other reporting biases and Questionable Research Practices).

Various statistical methods have been applied in order to detect these biases and reduce the risk of distortions. Even after more than 40 years of research, there is evidence that no publication bias method consistently outperforms all the others. It is common that different methods yield wildly different results [63]. For this reason, it is recommended to use several methods for publication bias evaluation, being difficult to know which one is best for our data and if its results are robust [2].

It is also unknown the exact extent to which selective reporting has affected the results of the meta-analysis. A recent meta-meta-analysis spanning several disciplines suggested that publication bias may be milder than expected in meta-analyses published in PLoS One, top medical journals and top psychology journals [64]. Another meta-meta-analysis led to the conclusion that evidence for publication bias in the studied subsets is weak, but suggestive of mild publication bias [65].

In a large meta-epidemiological study of treatment effects from Cochrane Database of Systematic Reviews, 19% of the large meta-analyses showed evidence for small-study effects, but only 3.9% showed evidence for publication bias [66].

Nonetheless, all authors advice for routinely assessment of publication bias in every meta-analysis, applying a set of different methods, as each method assess publication bias in a different way with possible different results, with careful interpretation of the gained information [7][67].

In the era of computing, there is relatively limited availability of software to implement the forementioned methods. The advances in statistical computing have made many of these methods more accessible to researchers and they are released in the open-source statistical software R. Similar but fewer programs can be found in STATA.

Recent versions of STATA (17) [68], SPSS (28) [69] provide user-friendly point-and-click options for meta-analysis and assessment of publication bias via different methods. Also, JASP is an open-source project, designed with the user in mind, which provides relevant modules [70].

In the following Table, a list of packages for R and STATA is presented.

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

Package	Methods
R	
`meta`	Funnel plot,[68] Asymmetry tests (Begg,[74] Egger,[69] Macaskill,[91] Harbord,[92] Peter[89]), Trim-and-Fill[77,78]
`metafor`	Funnel plot, asymmetry tests (Begg, Egger), Trim-and-Fill, Fail-safe *N* (Rosenthal,[63] Orwin,[64] Rosenberg[65])
`metamisc`	Asymmetry tests (Egger, Macaskill, Peters, Debray[196] [for survival data])
`metasens`	Copas sensitivity analysis,[112,130] Copas and Jackson upper bound,[132] Rucker's limit meta-analysis[106]
`PubBias`	Ioannidis and Trikalinos excess significance test[182]
`puniform`	p-uniform[184,189]
`PublicationBias`	Mathur and Vanderweele sensitivity analysis for PB[195]
`publipha`	Hedges[117] selection model
`selectMeta`	Iyengar and Greenhouse,[115] Dear and Begg,[116] Rufibach[123] selection models
`weightr`	Vevea and Hedges[118] selection model
`xmeta`	Test for PB under Copas' model,[194] Multivariate Egger's test,[164] Galaxy plot,[166] Bivariate trim-and-fill[167]
Stata	
`meta bias`	Asymmetry tests (Egger, Harbord, Peters, Begg)
`meta funnel plot`	Funnel plot
`meta trim fill`	Trim-and-Fill

Detailed R coding for a variety of methods can also be tracked in a) Doing Meta-Analysis in R: A Hands-on Guide [2] (available also as e-book, accessible at https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/), b) the book Questionable Research Practices in Clinical Psychology [20] and c) the book The Handbook of Research Synthesis and Meta-Analysis, 3rd edition [52].

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

There is still need for developing newer and better statistical methods to deal with publication bias in meta-analysis, especially when heterogeneity and/or non-independence among outcomes are encountered [15][71]. The increasing field of Network Meta-Analysis also necessitates novel methods for handling publication bias [15][72][73].

As in Medicine the preventive approach is always preferable, the same applies to publication bias. Practices as preregistration of studies in existing study registries (a more detailed catalogue of them can be found in WHO's site: https://www.who.int/observatories/global-observatory-on-health-research-and-development/resources/databases/databases-on-processes-for-r-d/clinical-trials), use of more powerful study designs, along with statistical consultation and training in statistical reasoning, open data and more rigorous search of 'grey literature' [3][5][7](even there is some critic on this later subject [74]) can mitigate the risk of publication bias and improve the validity of the meta-analytic results in a much better way than any statistical approach.

Taking in account that "Publication Bias begins at home", as stated in an Editorial by Ellen Weber [75], the editors' choices, the Questionable Research Practices, the social-economic context of research (e.g. desire of and competitive condition of academic career, funding policies, pressure from pharmaceutical companies), it seems wise to rethink the quote of Professor Douglas Altman [76]:

"As the system encourages poor research it is the system that should be changed.

We need less research, better research, and research done for the right reasons .

Abandoning using the number of publications as a measure of ability would be a start."

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

References

1.Glass GV. Primary, secondary, and meta-analysis of research. Educational Researcher. 1976; 5 (10):3–8

2. Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL and London: Chapman & Hall/CRC Press, 2021

3. Αρχαγγελίδη Ο. Το φαινόμενο της μεροληψίας δημοσίευσης στις επιστήμες υγείας. Παρουσίαση στατιστικών μεθόδων για την αντιμετώπισή του. Αρχεία Ελληνικής Ιατρικής 2013, 30(3): 340-354

4. Bernard R, Weisssgerber TL et al. fiddle: a tool to combat publication bias by getting research out of the file drawer and into the scientific community. Clinical Science 2020; 134: 2729-2739

5. Rothstein HR, Sutton AJ, Borenstein M., "*Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.*", John Wiley & Sons; Chichester, UK, 2005

6. Elias Zintzaras, Lecture notes from Postgraduate Program (MSc) "*Research Methodology in Biomedicine, Biostatistics and Clinical Bioinformatics*", University of Thessaly, 2020-21

7. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022

8. The James Lind Library: https://www.jameslindlibrary.org/essays-essay/2-7-dealing-with-biased-reporting-of-the-available-evidence/

9. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. J Am Stat Assoc. 1959;54(285):30-34.

10. Smart RG. The importance of negative results in psychological research. Can Psychol. 1964;5(4):225

11. Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. Am Stat. 1995;49(1): 108-112

12.. Greenwald AG. Consequences of prejudice against the null hypothesis. Psychol Bull. 1975;82(1):1-20

13. 9.Dickersin K, Chalmers I. Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. J R Soc Med. 2011 Dec;104(12): 532-8

14. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull. 1979;86(3):638-641

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

15. Marks-Anglin A, Chen Y. A historical review of publication bias. Res Synth Methods. 2020 Nov;11(6):725- 742

16. Orwin RG. A fail-safe N for effect size in meta-analysis. J Edu Stat. 1983;8(2):157-159

17. Rosenberg MS. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution. 2005;59(2):464-468

18. Gleser LJ, Olkin I., "Models for estimating the number of unpublished studies.", Statistics in Medicine 1996; 15: 2493–507

19. Eberly L., Casella G. Bayesian Estimation of the Number of Unseen Studies in a Meta-Analysis. Journal of Official Statistics, 1999;15(4): 477-494

20. O' Donohue, W., Masuda, A., Lilienfeld, S. (eds) Avoiding Questionable Research Practices in Applied Psychology. Springer, Nature Switzerland AG, 2022

21. Light RJ, Pillemer DB. Summing up: the science of reviewing research Harvard University press: Cambridge, MA, 1984, xiii + 191 pp. Edu Res. 1986;15(8):16-17

22. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol. 2001 Oct; 54(10): 1046-55

23. Sterne, Jonathan A. C.; Harbord, Roger M. Funnel Plots in Meta-analysis. The Stata Journal: Promoting communications on statistics and Stata, 2004; 4(2): 127–141

24. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ. 2006;333(7568): 597-600

25. Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 2000;53(5):477-484

26. Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ. 2007;176(8):1091-1096

27. Peters J., Sutton A. et al. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. Journal of Clinical Epidemiology 61 (2008) 991-996

28. Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. Int J Evid Based Healthc. 2018 Dec;16(4):195-203

29. Marcel A. L. M. van Assen, Olmo R. van den Akker, Hilde E. M. Augusteijn, Marjan Bakker, Michèle B. Nuijten, Anton Olsson-Collentine, Andrea H. Stoevenbelt, Jelte M. Wicherts, and Robbie C. M. van Aert. The Meta-Plot. Zeitschrift für Psychologie, 2023; 231(1): 65-78

30. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50(4): 1088-1101

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

31. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP., "Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials.", British Medical Journal 2011 Jul 22;343:d4002

32. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315 (7109):629-634

33. https://training.cochrane.org/resource/identifying-publication-bias-meta-analyses-continuous-outcomes

34. Pustejovsky, JE, Rodgers, MA. Testing for funnel plot asymmetry of standardized mean differences. *Res Syn Meth*. 2019; 10: 57– 71

35. Doleman, B, Freeman, SC, Lund, JN, Williams, JP, Sutton, AJ. Funnel plots may show asymmetry in the absence of publication bias with continuous outcomes dependent on baseline risk: presentation of a new publication bias test. *Res Syn Meth*. 2020; 11: 522– 534

36. Furuya-Kanamori L, Xu C, Lin L, et al. P value–driven methods were underpowered to detect publication bias: analysis of Cochrane review meta-analyses. J Clin Epidemiol. 2020; 118:86-92

37. Lin L, Chu H, Murad MH, et al. Empirical comparison of publication bias tests in meta-analysis. J Gen Intern Med. 2018;33 (8):1260-1267

38. Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc. 2000;95(449):89-98

39. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785–94

40. Zhu Q, Carriere K. Detecting and correcting for publication bias in meta-analysis–a truncated normal distribution approach. Stat Methods Med Res. 2018;27(9):2722-2741

41. Weinhandl ED, Duval S. Generalization of trim and fill for application in meta-regression. Res Synth Methods. 2012;3(1): 51-67

42. Moreno SG, Sutton AJ, Ades A, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol. 2009;9(1):2

43. Rücker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for smallstudy effects via a limit meta-analysis. Biostatistics. 2011;12(1): 122-142.

44. Stanley TD, Doucouliagos H. Meta-regression approximations to reduce publication selection bias. Res Synth Methods. 2014; 5(1):60-78

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

45. Stanley, T D. 2017. “Limitations of PET-PEESE and other meta-analysis methods.” Social Psychological and Personality Science 8 (5): 581–91

46. Stanley T., Doucouliagos H., Ioannidis J., Carter E. Detecting publication selection bias through excess statistical significance Res Synth Methods. 2021 Nov;12(6):776-795

47. McShane B. B., Böckenholt U., Hansen K. T. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspectives on Psychological Science 2016; 11(5): 730–749

48. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med. 2015 Jan 30;34(2):343-60

49. Hedges LV, Vevea JL. Estimating Effect Size Under Publication Bias: Small Sample Properties and Robustness of a Random Effects Selection Model.” *Journal of Educational and Behavioral Statistics* 1996; 21 (4): 299–332

50. Hedges LV. Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistically insignificant mean differences. J Edu Stat. 1984;9(1):61-85.

51. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. Stat Sci. 1988;3(1):109-117

52. Cooper, Harris M., Hedges, Larry V., Valentine, Jeff C. Handbook of research synthesis and meta-analysis, 3rd edition. Russell Sage Foundation, New York, 2019

53. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials. 2007;4(3):245-53

54. Simonsohn U, Nelson LD, Simmons JP. P-curve and effect size: correcting for publication bias using only significant results. Perspect Psychol Sci. 2014;9(6):666-681

55. van Assen MALM, van Aert R, Wicherts JM. Meta-analysis using effect size distributions of only statistically significant studies. Psychol Methods. 2015;20(3):293-309. 185

56 van Aert, R. C. M., & van Assen, M. A. L. M. (2018, October 2). Correcting for Publication Bias in a Meta-Analysis with the P-uniform* Method. https://doi.org/10.31222/osf.io/zqjr9

57 Copas, J. B., & Li, H. Inference for non-random samples. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1997; 59: 55–95

58. Copas, J. B. What works? Selectivity models and meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 1999; 162:95–109

59. Copas, J. B., & Shi, J. Q. A sensitivity analysis for publication bias in systematic reviews. Statistical Methods in Social Research, 2001;10: 251–265

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

60 Copas, J. B. A likelihood-based sensitivity analysis for publication bias in meta-analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2013; 62: 47–66

61. Bayarri, M. J., and Morris H. DeGroot. “Bayes Analysis of Selection Models.” Journal of the Royal Statistical Society, Series D (The Statistician). 1987; 36(2/3): 137–46

62. Bai R, Lin L, Boland MR, Chen Y. A robust Bayesian Copas selection model for quantifying and correcting publication bias. 2020. arXiv preprint arXiv:2005.02930 (https://doi.org/10.48550/arXiv.2005.02930)

63. Lin L., Chu H. et al. Empirical Comparison of Publication Bias Tests in Meta-Analysis. J Gen Intern Med. 2018 Aug; 33(8): 1260–1267

64. Mathur M., VanderWeele TJ. Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers Res Synth Methods. 2021 March ; 12(2): 176–191

65. van Aert RCM, Wicherts JM, van Assen MALM. Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. PLoS One. 2019 Apr 12;14(4):e0215052

66. Schwab S, Kreiliger G, Held L. Assessing treatment effects and publication bias across different specialties in medicine: a meta-epidemiological study. BMJ Open 2021;11:e045942

67. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Res Synth Methods. 2021 Mar;12(2):248-259

68. https://www.stata.com/features/meta-analysis/

69. https://www.ibm.com/docs/en/spss-statistics/saas?topic=features-meta-analysis

70. https://jasp-stats.org/features/

71. Nakagawa S, et al. Methods for testing publication bias in ecological and evolutionary meta-analyses. Methods Ecol Evol. 2022;13:4–21

72. Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Stat Med. 2013;32(1): 51-66

73. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Stat Med. 2014;33(30):5399-5412

74. Schmucker CM, Blumle A, Schell LK, Schwarzer G, Oeller P, Cabrera L, et al. (2017) Systematic review finds that study data not published in full text articles have unclear impact on meta-analyses results in medical research. PLoS ONE 12(4): e0176210

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218

75. Weber EJ. Publication bias begins at home. Emerg Med J 2019;36:518–519

76. Altman D G. The scandal of poor medical research *BMJ* 1994; 308 :283

Institutional Repository - Library & Information Centre - University of Thessaly
06/07/2024 06:15:11 EEST - 3.144.100.218