ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
ΠΜΣ «Μεθοδολογία Βιοϊατρικής Έρευνας, Βιοστατιστική και Κλινική Βιοπληροφορική»
Διεύθυνση: Εργαστήριο Βιομαθηματικών, Τμήμα Ιατρικής, Πανεπιστήμιο Θεσσαλίας, Παπακυριαζή 22, Λάρισα 41222
URL: http://biomath.med.uth.gr • email: biomath@med.uth.gr • Τηλ: 6939 040581 (κιν), 2410 565064 (σταθ)

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Μέθοδοι εύρεσης του σφάλματος δημοσίευσης στην μετά-ανάλυση»

«Methods to detect publication bias in meta-analysis»

ΑΘΑΝΑΣΙΟΣ ΤΣΙΤΣΙΦΛΗΣ (M060620012)
ΜΑΘΗΜΑΤΙΚΟΣ
ΑΝΑΛΥΤΗΣ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΔΙΑΔΙΚΑΣΙΩΝ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ:
ΔΑΡΔΙΩΤΗΣ ΕΥΘΥΜΙΟΣ (ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ, ΙΑΤΡΟΣ)
ΣΤΕΦΑΝΙΔΗΣ ΙΩΑΝΝΗΣ (ΚΑΘΗΓΗΤΗΣ, ΙΑΤΡΟΣ)
ΔΟΞΑΝΗ ΧΡΥΣΟΥΛΑ (ΙΑΤΡΟΣ)

ΛΑΡΙΣΑ, 2021

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Η μετά-ανάλυση αποτελεί μία ισχυρή και ευρέως διαδεδομένη τεχνική στατιστικής ανάλυσης και επεξεργασίας ευρημάτων από διαφορετικές έρευνες και συμβάλλει στη λήψη αποφάσεων στην αποδεικτική ιατρική. Ωστόσο, η πιθανότητα δημοσίευσης μίας έρευνας από ένα επιστημονικό περιοδικό συνδέεται συχνά με τη στατιστική σημαντικότητα των αποτελεσμάτων της. Πιο σημαντικά ευρήματα είναι πιθανότερο να δημοσιευτούν, προκαλώντας έτσι το λεγόμενο "σφάλμα δημοσίευσης" στη μετά-ανάλυση δημοσιευμένων μελετών. Το σφάλμα δημοσίευσης είναι ένα σοβαρό πρόβλημα των συστηματικών ανασκοπήσεων και των μετά-αναλύσεων που μπορεί να επηρεάσει την εγκυρότητα και την γενίκευση των συμπερασμάτων. Ο εντοπισμός του συγκεκριμένου σφάλματος είναι μία κρίσιμη διαδικασία καθώς, στην περίπτωση που αγνοηθεί, μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα των συστηματικών ανασκοπήσεων.

Στην παρούσα εργασία γίνεται έρευνα των μεθόδων εντοπισμού του σφάλματος δημοσίευσης σε δημοσιεύσεις και στην παγκόσμια βιβλιογραφία. Γίνεται αναζήτηση σε βάσεις δεδομένων όπως PubMed και NCBI (https://pubmed.ncbi.nlm.nih.gov/), Cochrane library (https://www.cochranelibrary.com/), ScienceDirect (https://www.sciencedirect.com/), JAMA (https://jamanetwork.com/journals/jama) καθώς και στη μηχανή αναζήτησης Google (https://www.google.gr/).

Από την έρευνα προέκυψαν απόψεις σχετικές με την διαχείριση του σφάλματος δημοσίευσης, οι οποίες μπορούν να διακριθούν σε 2 κατηγορίες: σε εκείνες που βασίζονται στη μελέτη του γραφήματος "χωνί" και στα μοντέλα επιλογής. Γίνεται παρουσίαση των μεθόδων της κάθε κατηγορίας, παρουσιάζονται τα αποτελέσματα τους και ακολουθεί συζήτηση.

Λέξεις κλειδιά: Μετά-ανάλυση, σφάλμα δημοσίευσης, γράφημα χωνί, μοντέλα επιλογής, γραμμική παλινδρόμηση, συσχέτιση κατάταξης.

ABSTRACT

Meta-analysis has become a powerful and widely used tool to integrate findings from different studies and inform decision making in evidence-based medicine. However, the chance of a study being published by a scientific journal is frequently associated with the statistical significance of its results: more significant findings are more likely to be published, causing publication bias in meta-analysis of published studies. Publication bias is a serious problem in systematic reviews and meta-analyses, which can affect the validity and generalization of conclusions. Detecting publication bias is a critical problem because such bias may lead to incorrect conclusions of systematic reviews.

In this study we looked for methods that detect publication bias in international literature. A search has been made in databases like PubMed and NCBI (https://pubmed.ncbi.nlm.nih.gov/), Cochrane library (https://www.cochranelibrary.com/), ScienceDirect (https://www.sciencedirect.com/), JAMA (https://jamanetwork.com/journals/jama) as well as in search engines like Google (https://www.google.gr/)

Currently, approaches to dealing with publication bias can be distinguished into two categories: funnel-plot-based methods and selection models. For each category, the main methods are presented followed by a discussion about their results.

Keywords: Meta-analysis, publication bias, funnel plot, selection models, linear regression, rank correlation.

A. INTRODUCTION

A randomized controlled trial (RCT) compares a treated group with a control group in a parallel group study design. Often, for specific comparison between two treatments, multiple clinical trials are conducted for the same outcome. However, the study results are never completely homogenous and it is difficult to make inferences regarding the effectiveness and safety of the treatment. Usually, based on the results, is hard to draw a safe conclusion regarding the effectiveness of a specific treatment. We need to provide an overall estimate that shows the magnitude of effectiveness [1].

Similarly, genetic association studies (GAS) assess the association between disease status and genetic variants (gene polymorphisms) in a population and often, for specific disease, multiple GAS are conducted for the same variant. However, the study results are never completely homogenous and it is difficult to make inferences whether a variant is responsible for developing the disease [1].

In cases like these, meta-analysis can play a role. Meta-analysis is a technique that allows us:

i. to estimate the overall (pooled) difference between two treatments for a specific outcome after synthesizing the outcomes of multiple trials,
ii. to explore the sources of heterogeneity across studies and
iii. to investigate the existence of publication bias [1].

The pooled estimate of treatment difference can be systematically influenced by the selection of studies for inclusion in the meta-analysis. Then, in the meta-analysis, bias may be introduced in two different ways [1]:

a. by including studies which have themselves produced biased estimates of the treatment difference. This bias can be due to methodological flaws (such as inappropriate allocation of patients to treatment groups),
b. by excluding eligible studies because the relevant data are not available. This is the publication bias which is introduced when the meta-analysis is restricted to the synthesis of results obtained only from trials which have been published.

The earliest study providing evidence of publication bias of which we are aware was performed in 1959 [2]. Sterling reviewed four prominent psychology journals and found that, of the articles reporting results involving hypothesis testing, more than 95% reported statistically significant ("positive") findings. Dr. Sterling updated his study in 1988 and reviewed the same four psychology journals and four medical journals, for 1986 and 1987. There was little indication that the situation had changed in the psychology journals over the 40-year period covered, with about 95% of the articles that performed hypothesis testing reporting statistically significant results. The situation in the medical journals was found to be very similar, with about 85% of the articles reporting statistically significant results. Although it is possible that 85% to 95% of all studies undertaken have results that reject a prior null hypothesis (H_0), this possibility seems remote. Since Sterling's original report, evidence has accumulated to show that studies reaching publication are a select sample of all studies conducted [3, 4].

Another evidence of publication bias with respect to clinical trials was reported in a study by John Simes [5]. He compared the results of two separate meta-analyses of chemotherapy clinical trials: one meta-analysis used only data from studies that had been published, and the other used data from studies listed in a cancer trials registry, some of which had never been reported. Simes found that when just published trials were considered, use of a combined chemotherapeutic regimen for ovarian cancer was statistically significantly superior to use of a single alkylating agent. When the results of all registered trials were considered, however, including those that remained unpublished, no statistically significant advantage of the combination chemotherapy was observed [6].

In another study [7], authors of published reports of randomized trials were surveyed to learn which other trials they had performed, the results of these trials, and their publication status. Fifty-five percent of the published trials, compared with only 15% of the unpublished trials, had statistically significant results favoring a new therapy. Most of the unpublished trials had not detected any differences between treatments (44%) or had only shown a trend that had not achieved “statistical significance” (35%).

Since the 1990s, there has been a growing awareness that publication bias represents a major threat to the validity of conclusions from medical research [8]. In 2005, the International Committee of Medical Journal Editors introduced the requirement of a prospective trial registration in a public registry as a condition for publication [9]. The seventh revision of the Declaration of Helsinki adopted by the World Medical Association in 2008 contained the requirement of a prospective registration of clinical trials and called disclosure of all results an author’s duty [10]. It is unclear how these initiatives affected the level of research underreporting in the medical literature.

Often, the decision to submit or accept a paper for publication in a journal with the results of a trial, is influenced by the significance of results. For example, studies with statistically significant effects and positive treatment outcomes are more likely to be published [11, 12], resulting in a biased estimate of the effect of treatment in the meta-analysis. Both the decision to submit a study for publication and the probability that a journal will accept it for publication are associated with the study results [13]. In practice, the studies which are less likely to appear in the published literature tend to be the less conclusive ones (because of smaller sample sizes or less statistical precision) or those where the treatment effect is small. Although searching for relevant unpublished studies is important and may sometimes alleviate publication bias, identifying such studies may be difficult [14]. Hence, we need methods to detect publication bias, based on the data in the available studies.

B. METHODS

Currently, approaches to dealing with publication bias can be distinguished into two categories: funnel-plot-based methods and selection models [15].

On the first category of approaches to detecting publication bias is based on a funnel plot, which usually presents effect sizes plotted against their standard errors or precisions (the inverse of standard errors) [11, 16, 17]. In the presence of publication bias, the funnel plot is expected to be skewed. One may intuitively assess publication bias by examining the asymmetry of the funnel plot; however, the visual examination is usually subjective. Various statistical tests have been proposed for publication bias in the funnel plot, such as Begg's rank test [13] and Egger's regression test [18] and its extensions [14, 20, 21, 22]. The rank test examines the correlation between the effect sizes and their corresponding sampling variances; a strong correlation implies publication bias. Egger's test regresses the standardized effect sizes on their precisions; in the absence of publication bias, the regression intercept is expected to be zero. Note that this regression is equivalent to a weighted regression of the effect sizes on their standard errors, weighted by the inverse of their variances; the weighted regression's slope, instead of the intercept, is expected to be zero in the absence of publication bias [20]. The weighted regression version of the test is popular among meta-analysts, probably because it directly links the effect sizes to their standard errors without the standardization process. In addition, another attractive method is the trim and fill method, which not only tests for publication bias but also adjusts the estimated overall effect size [23,24]. Although these publication bias tests have been widely used in meta-analysis applications, they may suffer from inflated type I error rate or poor power in certain simulation settings [22, 25, 26, 28, 29].

The second category of methods for publication bias is based on selection models. These approaches typically use the weighted distribution theory to model the selection (i.e., publication) process and develop estimation procedures that account for the selection process [30, 31, 32, 33, 34]. The selection models are usually complicated, limiting their applicability. Moreover, they incorporate weight functions in order to correct publication bias, but strong and largely untestable assumptions are often made [34]. Therefore, the validity of their adjusted results may be doubtful, and these methods are usually employed as sensitivity analysis.

Besides detecting publication bias using selection models and funnel-plot-based methods, it is also important to quantify publication bias using measures that permit comparisons between different meta-analyses. A candidate measure is the intercept of the regression test [18]. However, as a measure of asymmetry of the collected study results, the regression intercept lacks a clear interpretation; for example, it is difficult to provide a range guideline to determine mild, moderate, or substantial publication bias based on the regression intercept. Due to this limitation, meta-analysts usually report the p-value of Egger's regression test, but not the magnitude of the intercept.

The commonly used quantitative methods rely on the assumption that publication bias is one of the key reasons for an association between an effect estimate and its measure

of precision; thus, all the commonly reported methods (e.g., Egger's and Peter's regression tests) are statistical tests of this association. Although these methods do not detect publication bias per *se*, they aim to detect the presence of this association (between effect estimate and precision) also known as study asymmetry in the funnel plot [25]. All these methods depend on p-value thresholds (usually <0.1) of a statistical test to define the presence of asymmetry; thus, they also depend on the number of studies included in the meta-analysis.

Funnel plot:

The funnel plots are a commonly used graphical method to detect publication bias. Funnel plots are scatter plots of the effect estimates of each study against some measure of precision. In the absence of publication bias, it is expected that these plots will resemble a symmetrical inverted funnel where smaller studies will scatter widely at the bottom of the plot (because of random variation), with the spread narrowing with increasing size of the study and thus its precision. However, the use of the funnel plot for the evaluation of publication bias related asymmetry is often subjective, as it involves a personal interpretation of how the studies are scattered in the plot, without an objective or quantitative measure. As such, several P value-driven statistical methods have been developed to complement the funnel plot in detecting publication bias [18, 21, 22, 13].

The name "*funnel plot*" arises from the fact that precision of the estimated intervention effect increases as the size of the study increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias, the plot should approximately resemble a symmetrical (inverted) funnel. This is illustrated in Figure 1, in which the effect estimates in the larger studies are close to the true intervention odds ratio of 0.4.

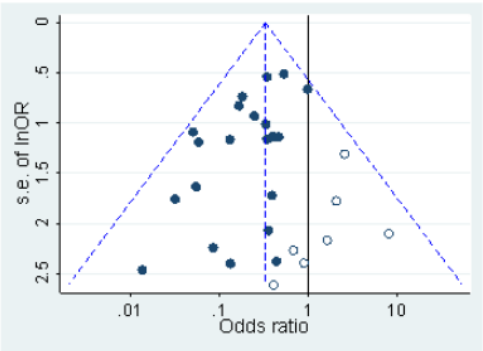

Figure 1: Symmetrical plot in the absence of bias

If there is bias, for example because smaller studies without statistically significant effects (shown as open circles in Figure 1) remain unpublished, this will lead to an asymmetrical appearance of the funnel plot with a gap in a bottom corner of the graph (Figure 2). In this situation the effect calculated in a meta-analysis will tend to overestimate the intervention effect [18, 35]. The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

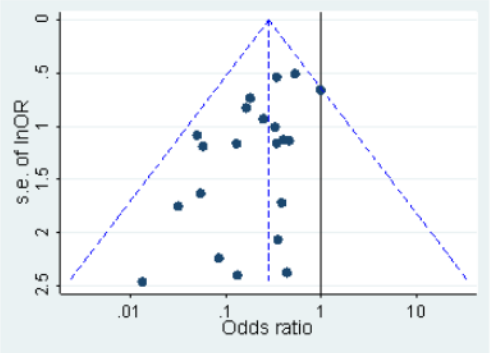

Figure 2: Asymmetrical plot in the presence of reporting bias

Funnel plots were first used in educational research and psychology, with effect estimates plotted against total sample size [11]. It is now usually recommended that the standard error of the intervention effect estimate be plotted, rather than the total sample size, on the vertical axis [16]. This is because statistical power of a trial is determined by factors in addition to sample size, such as the number of participants experiencing the event for dichotomous outcomes, and the standard deviation of responses for continuous outcomes. For example, a study with 100,000 participants and 10 events is less likely to show a statistically significant intervention effect than a study with 1000 participants and 100 events. The standard error summarizes these other factors. Plotting standard errors on a reversed scale places the larger, or most powerful, studies towards the top of the plot. Another potential advantage of using standard errors is that a simple triangular region can be plotted, within which 95% of studies would be expected to lie in the absence of both biases and heterogeneity. Funnel plots of effect estimates against their standard errors (on a reversed scale) can be created using proper software (e.g., RevMan). A triangular 95% confidence region based on a fixed-effect meta-analysis can be included in the plot, and different plotting symbols allow studies in different subgroups to be identified [36].

Publication bias need not lead to asymmetry in funnel plots. In the absence of any intervention effect, selective publication based on the p-value alone will lead to a symmetrical funnel plot in which studies on the extreme left or right are more likely to be published than those in the middle. This could bias the estimated between-study heterogeneity variance [36].

Ratio measures of intervention effect (such as odds ratios and risk ratios) should be plotted on a logarithmic scale. This ensures that effects of the same magnitude but opposite directions (for example odds ratios of 0.5 and 2) are equidistant from 1.0. For outcomes measured on a continuous (numerical) scale (e.g., blood pressure, depression score) intervention effects are measured as mean differences or standardized mean

differences, which should therefore be used as the horizontal axis in funnel plots. So far as we are aware, no empirical investigations have examined choice of axes for funnel plots for continuous outcomes. For mean differences, the standard error is approximately proportional to the inverse of the square root of the number of participants, and therefore seems an uncontroversial choice for the vertical axis [36].

A test for funnel plot asymmetry (small study effects) formally examines whether the association between estimated intervention effects and a measure of study size (such as the standard error of the intervention effect) is greater than might be expected to occur by chance. For outcomes measured on a continuous (numerical) scale this is reasonably straightforward. Using an approach proposed by Egger et al. [18], we can perform a linear regression of the intervention effect estimates on their standard errors, weighting by 1/(variance of the intervention effect estimate). This looks for a straight-line relationship between intervention effect and its standard error. Under the null hypothesis of no small study effects (e.g., in Figure 1) such a line would be vertical. The greater the association between intervention effect and standard error (e.g., in Figure 2), the more the slope would move away from vertical. Note that the weighting is important to ensure the regression estimates are not dominated by the smaller studies.

In particular, a regression line is fitted: $\mathbf{y_i=a+bx_i}$, for i=1,2,…,n where **n** is the number of studies, $\mathbf{y_i}$ is the standardized estimate of $\boldsymbol{\theta_i}$ ($\theta_i = \ln(OR_i)$, $y_i = \theta_i\sqrt{w_i}$, $w_i = \frac{1}{vi}$, $v_i=var(\theta_i)$) and $\mathbf{x_i}$ is the precision ($x_i = \sqrt{w_i}$).

A test of publication bias would be a test whether the intercept **a** is equal to zero. The intercept **a** and slope **b** can be obtained by performing a typical least-squares regression of $y_i = \theta_i\sqrt{w_i}$, on $x_i = \sqrt{w_i}$) (using SPSS). Then, in testing whether the intercept **a** is zero (0), the statistic $t = \frac{a}{SE(a)}$ is compared against the 5% point of the t-distribution with $n-1$ df. Thus, if **t** is less than the 5% point of the t-distribution with $n-1$ df, there is no indication of publication bias [1].

Begg's rank correlation method [37, 13] uses Kendall's test to evaluate the association between the standardized effect estimates (T_i*) and their variance of the treatment effect (v_i). Define the standardized effect sizes of k studies to be combined to be:

$$\overline{T}_i^* = (T_i - \overline{T}_\bullet)/(\widetilde{v}_i^*)^{1/2},$$

$$T_\bullet = \left(\sum_{j=1}^{k} v_j^{-1} T_j\right) \Big/ \sum_{j=1}^{k} v_j^{-1},$$

where ***i*** is the *i*th study and the usual fixed-effect estimate of the pooled effect is:

and $\widetilde{v}_i^*$:

$$\tilde{v}_i^* = v_i - \left(\sum_{j=1}^{k} v_j^{-1} \right)^{-1},$$

the variance of $(T_i - \overline{T}_\bullet)$.

The test is based on deriving a rank correlation between T_i* and $\tilde{v}_i^*$, achieved by comparing the ranks of the two quantities. Having assigned each T_i* and a rank $\tilde{v}_i^*$ (largest value gets rank 1 and so on), it is then necessary to evaluate all *k(k-1)/2* possible pairs of the k studies. Defining the number of all possible pairings in which one factor is ranked in the same order as the other as P, (that is the ranks of both T_i* and are higher or lower for one study compared to the other) and the number in which the ordering is reversed as Q, (that is the rank of T_i* is higher than that of the paired $\tilde{v}_i^*$ study while the rank of is lower, or vice versa). A normalized test statistic is obtained by calculating the quantity:

$$Z = (P - Q)/[k(k-1)(2k+5)/18]^{½},$$

which is the normalized Kendall rank correlation test statistic and can be compared to the standardized normal distribution. Any effect-size scale can be used as long as it is assumed distributed asymptotic normal. Thus, if Z is less than the 5% point of the normal distribution with n-1 df, there is no indication of publication bias [14].

Selection models:

Various selection models proposed for the modelling of publication bias in meta-analysis. They are arranged broadly in order of sophistication, starting with the simplest models. Latter sections describe Bayesian methodology to implement weight functions. These methods often use weight functions defined previously from a classical perspective. Although the Bayesian methods are often more difficult to implement, they do have advantages over their classical counterparts [38].

The simple model of Hedges [39, 40] follows on from work conducted by Lane and Dunlap [41] who carried out a simulation to investigate the degree to which an effect size is over-estimated when only statistically significant results are considered. Hedges dichotomized studies into those in which significant results, defined by some prespecified level of **a** (taken as 0.05), were obtained and hence published, and those which produced non-significant results and were not published.

Only a fixed effect meta-analysis of continuous outcomes using the standardized mean difference scale was considered. In some instances, e.g., where relevant studies have been identified and reports retrieved, but effect sizes are not reported if outcome measures are not significant, this model may be realistic. If a less stringent censoring rule is present, implying some non-significant results have been obtained, this model can still be used, but the effect size must be calculated using just the results from the significant studies.

The weight function takes the value 1 if the test is significant (at the 5% or any other specified level), and 0 otherwise for each study in the meta-analysis [37, 39]:

$$w(x) = \begin{cases} 1 & \text{if } x > C_\alpha(v) \\ 0 & \text{if } x < C_\alpha(v) \end{cases}$$

where $C_\alpha(v)$ is the critical value of the **a**-level test for the study and v is the standard error of the studies' effect size. This weight function is represented graphically in Figure 3:

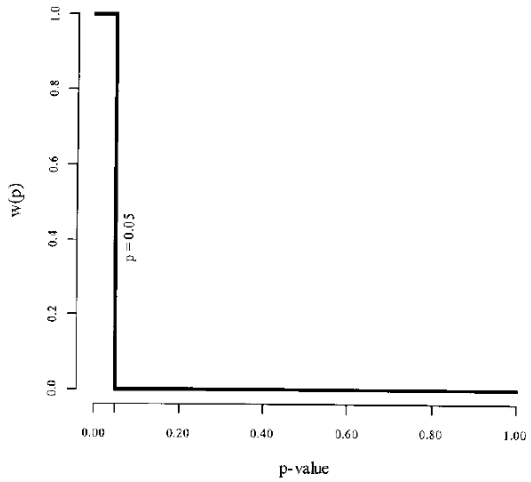

Figure 3: Simple weight function proposed by Hedges

Estimation of model parameters for this model, based on a vote counting approach is initially presented [40]. It treats positive and negative results as independent realizations of a Bernoulli process and the adjusted treatment effect can be estimated using a modification of binomial theory [39, 40]. However, this method is limited in its application because it is only valid when a relatively large number of studies exist, and which are all approximately the same size. Due to these drawbacks, Hedges considers this approach as most useful for providing quick approximate estimates rather than serving as the analytic tool for definitive analysis [39].

A further technical examination of the problem of interpreting the sample mean and variance from a normal distribution when they are reported conditional upon rejection of the hypothesis that the mean is zero is provided by Hedges [42]. This method has been described as appealing when most of all the published studies are significant, and inappropriate if most of the studies are non-significant [3]. The method is strongly dependent on the nature of the distribution of the p-values in the range 0.00 to 0.05, and accuracy is open to question. The assumption that all non-published results have a mean effect of zero has been considered too simplistic [43] as there is evidence to suggest that the mean is displaced in the direction of the mean of the published studies. An attempt to address this limitation has been made by trying to estimate the mean effect size in the population of all studies, both published and unpublished [43].

Champney [44] considered the same situation as Hedges, where it is assumed, all studies reporting significant results are published and all studies with non-significant results are not. Champney extends Hedges work by developing a method of adjustment which is based on a random effects model [45]. Maximum likelihood estimates are derived

using both grids searching techniques and the EM algorithm [46]. This work suggests that publication bias may have substantial effects on estimation of the between-study variance even when the estimate of the mean is not strongly affected [43, 39, 40].

Iyengar and Greenhouse [43] expand on the model of Hedges by presenting more sophisticated study censoring schemes. Two different families of weight functions are discussed:

$$w(x\,|\beta,q) = \begin{cases} \dfrac{|x|^{\beta}}{t(q,0.05)^{\beta}}, & \text{if} \quad |x| \le t(q,0.05) \\ 1, & \text{otherwise} \end{cases}$$

$$w(x\,|\gamma,q) = \begin{cases} e^{-\gamma}, & \text{if} \quad |x| \le t(q,0.05) \\ 1, & \text{otherwise} \end{cases}$$

In both cases, $P(|T_0| \geq t(q,0.05)) = 0.05$, where T_0, has a central t-distribution with q degrees of freedom. Both these functions imply that all studies statistically significant at the 0.05 level will be published as the weight functions will take the value one over these values. When β and γ are zero, the weight functions indicate no selection bias, and they approach the weight function of Hedges $w(x)$ as β and γ approach infinity [39].

Rust et al. [47] describes a weight function for estimating publication bias based on a measure of effect size, rather than the statistical significance of the associated p-value used in all the models described above. This paper has largely been ignored by other researchers in the medical meta-analysis field, perhaps due to the economic context in which it was published, although the model developed is generalizable to the treatment estimates found in medicine and other related disciplines [38]. The model assumes that publication bias involves a fixed censorship threshold, beyond which no censorship occurs. This censorship threshold is estimated in the model rather than specified a priori. This implies that if the reported effect size is greater than the threshold value then the probability of publication is one, and when it is smaller than the threshold the probability of publication is a fixed value, which is estimated from the model, together with the other model parameters, via maximum likelihood methods. A limitation of this model is that the method as proposed does not adjust for differences in the sample size across the studies, however, the authors note that this would be relatively easy to incorporate.

Weight functions are implemented from a Bayesian perspective:

Bayesian methods can be considered as an alternative to the classical approach to statistical analysis. The name originates from the Reverend Thomas Bayes (1702–1761), who in papers published posthumously [48], outlined a different system for making statements regarding probabilities and random phenomena. At the heart of this alternative system was an equation which forms the basis of all modern Bayesian theory. This is now commonly referred to as Bayes' theorem. Bayesian methods have become more frequently used in several areas of healthcare research, including meta-analysis, over the last few years [49, 50]. Though much of this increase in their use has been directly because of advances in computational methods, it has also been partly due to their more appealing

nature and specifically the fact that they overcome some of the difficulties encountered by other methods traditionally used.

The Bayesian approach can be summarized as follows: opinions are expressed in probabilities, data are collected, and these data change the prior probabilities, through the operation of Bayes' theorem, to yield posterior probabilities [51]. Opinions are expressed in probabilities which implies that the subjective beliefs of the researchers (or possibly experts from the field/panel consensus), prior to conducting the analysis, form the starting point for the analysis. These prior beliefs are than combined with the data, in the form of a likelihood function, to produce a posterior distribution, which takes both subjective and objective evidence into account. It is in the incorporation of subjective beliefs that the Bayesian approach differs greatest from the classical viewpoint, which only considers objective evidence. However, another key difference between the Bayesian and Classical approach is the role that the likelihood function plays. In the classical approach, the likelihood function defines the support for various values of the parameter of interest, conditional upon the observed data. In the Bayesian approach since both the data and model parameters are considered random, the conditioning may be reversed, and thus the Bayesian considers the likelihood function to measure the plausibility of the observed data condition upon the parameters of the model [52].

Several approaches which assess publication bias in meta-analysis using methods other than adjustment via the use of selection models, have also been developed. In Rosenthal's "*file drawer*" method, there is a consideration of how many new studies averaging a null result are required to bring the overall treatment effect to non-significance [53]. It was developed by Rosenthal [53, 54] and it could be seen as estimating the number of studies filed away by researchers without being published. The method is based on combining the normal z-scores (the Z_i) corresponding to the p-values observed for each study. The overall z-score can be calculated by the formula:

$$Z = \sum_{i=1}^{k} Z_i \Big/ \sqrt{k}$$

where k is the number of studies in the meta-analysis. This sum of z-scores is a z-score itself and the combined z-scores are considered significant (i.e., the outcome measured in the studies is significant) if $Z > Z_{a/2}$ (the a/2 percentage point of a Standard Normal Distribution). The number of unpublished studies (fail-safe N) with an average observed effect of zero that there would need to be in order to reduce the overall z-score to non-significance is determined. Define k_0 to be the additional number of studies required such that:

$$k_o > -k + \left(\sum_{i=1}^{k} Z_i \right)^2 \Big/ \left(Z_{\alpha/2} \right)^2$$

Gleser and Olkin [55] have developed a methodology that attempts to estimate the number of missing studies from a meta-analysis. Two general methods are presented, each allowing the number of unpublished studies (N) to be estimated using the *p-values* reported in the published studies. Duval and Tweedie [24] suggest this is largely do the lack

of robustness of Gleser and Olkin's method both to isolated negative values (leading to the zero estimates) and the heavy dependence on the null hypothesis.

Eberly and Casella [56] present a model for estimating the total number of studies carried out, both published and unpublished, which is dependent on the probability of publication. A simple selection model is used, where all studies significant at level a are published, while non-significant studies are published with probability p. This method differs from those using weight functions.

The "*trim and fill*" method of Duval and Tweedie [23, 24] formalizes the use of funnel plots and estimates and adjusts for the numbers and outcomes of missing studies. An iterative rank-based algorithm estimates how many studies are missing. This number of studies are "*trimmed*" from the asymmetric outlying part of the funnel (i.e., those with the largest effect size estimates): these can broadly be thought of as studies which have no counterpart on the other side of the funnel plot (i.e., this is the truncation that is picked up by "*eye-balling*" a funnel plot). Then the symmetric remainder are used to estimate the "*true center*" of the funnel using standard meta-analysis techniques. The "*trimmed*" studies are then replaced and their "*missing counterparts*" imputed or "*filled*": these are mirror images of the "*trimmed*" studies with the mirror axis placed along the adjusted pooled estimate. This last stage is necessary for the variance of the pooled estimate to be calculated correctly.

This approach assumes studies are suppressed and not published under a scenario where it is the magnitude of the effect size, and not the p-value which determines the chance of publication (the size of the studies is not taken into consideration). The key assumption of the method is that it is the most extreme negative studies (i.e., those with the smallest outcome estimates) which have not been published [38].

C. RESULTS

Although funnel plot asymmetry has long been equated with publication bias [3, 11], the funnel plot should be seen as a generic means of displaying small-study effects – a tendency for the intervention effects estimated in smaller studies to differ from those estimated in larger studies [25]. Small-study effects may be due to reasons other than publication bias [18, 25].

Differences in methodological quality are an important potential source of funnel plot asymmetry. Smaller studies tend to be conducted and analyzed with less methodological rigour than larger studies [57]. Trials of lower quality also tend to show larger intervention effects [58]. Therefore, trials that would have been 'negative', if conducted and analyzed properly, may become 'positive' (Figure 4):

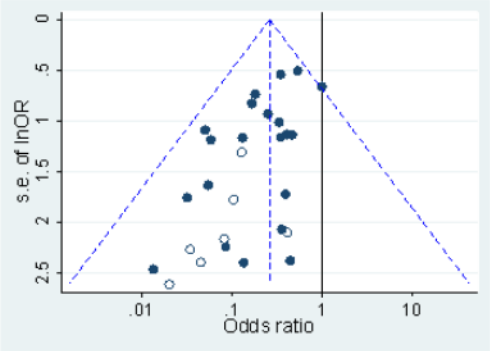

Figure 4: Asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates

True heterogeneity in intervention effects may also lead to funnel plot asymmetry. For example, substantial benefit may be seen only in patients at high risk for the outcome which is affected by the intervention and these high-risk patients are usually more likely to be included in early, small studies [59, 60]. In addition, small trials are generally conducted before larger trials are established and in the intervening years standard treatment may have improved (resulting in smaller intervention effects in the larger trials). Furthermore, some interventions may have been implemented less thoroughly in larger trials and may, therefore, have resulted in smaller estimates of the intervention effect [61]. Finally, it is of course possible that an asymmetrical funnel plot arises merely by the play of chance. Terrin et al. have suggested that the funnel plot is inappropriate for heterogeneous meta-analyses, drawing attention to the premise that the studies come from a single underlying population given by the originators of the funnel plot [11, 26].

The value of the funnel plot has not been systematically examined, and symmetry (or asymmetry) has generally been defined informally, through visual examination. Unsurprisingly, funnel plots have been interpreted differently by different observers [18].

Egger's test has highly inflated Type 1 errors in some circumstances when binary outcomes are considered, at least in part because of the correlation between estimated (log) odds ratios and their standard errors [14, 62, 22]. This has led to a number of modified

tests being developed for binary outcomes [14, 21, 22, 27] and more generally [19]. A comparative evaluation of these modified tests is required before guidance can be given on which test is optimal in a given situation. Egger approach, is known to be intrinsically biased because: (i) the independent variable is subject to sampling variability; (ii) the standardized treatment effect is correlated with its estimated precision; and (iii) for binary data, the independent regression variable is a biased estimate of the true precision, with larger bias for smaller sample sizes [63, 64, 65].

Begg's test is known to have low power for meta-analyses that include few studies [25]. From comparisons between this and the linear regression test described later, it seems that the linear regression test is more powerful than the rank correlation test, though results of the two tests can sometimes be discrepant. In addition, it has been shown that Begg's method has low power with continuous (normally distributed) data, particularly when the number of studies is small [13, 14]. This approach is attractive due to its conceptual and computational simplicity, although concerns have been raised about its possible lack of power [13].

In selection models, Hedges performed simulations examining the applicability of such models for meta-analysis and concluded that the procedure does a 'remarkably good job' of estimating the selection model. However, when the standard errors of the treatment effect estimates from individual studies are large, the estimated selection model may be too uncertain to be useful for some purposes. This is because the data contain relatively little information about the weights in the selection model [38].

When the number of studies is small, two problems arise. There may be numerical problems in obtaining stable estimates of the model parameters. More importantly, the standard errors of estimates will be large, perhaps so large as to make any specific inferences impossible or meaningless. For example, a result might be highly statistically significant in a conventional analysis and the estimates after correction for selection might be essentially identical. However, the corrected estimate might be far from statistically significant because the standard error was so much larger [20].

D. DISSCUSSION

As with any modelling exercise, the eventual selection of a 'final model' is a difficult task, and one which in the meta-analysis literature has received little attention. This is partly because of the relative lack of use of regression models generally, both Bayesian and classical. That having been said, one aspect of model selection that has received considerable attention and aroused heated debate is the choice between fixed and random effect models [66]. The results suggest that no single method is sufficient for assessing evidence of publication bias, and that such methods may also offer insight into potential sources of heterogeneity, which may in turn guide the design of future studies [67].

If a meta-analysis dataset displays heterogeneity that can be explained by including study-level covariates, these should be accounted for before an assessment of publication bias is made because their influence can distort a funnel plot and thus any statistical methods based on it, as discussed earlier. One way of doing this is to extend the regression models to include study-level covariates. Evaluating the performance of such an approach is ongoing work. Alternatively, if covariates are discrete, separate assessments can be made for each grouping. Splitting the data in this way will reduce the power of the tests considerably, however [37].

For funnel plot, a proposed enhancement [68] is to include contour lines corresponding to perceived 'milestones' of statistical significance (P = 0.01, 0.05, 0.1 etc). This allows the statistical significance of study estimates, and areas in which studies are perceived to be missing, to be considered. Such 'contour-enhanced' funnel plots may help review authors to differentiate asymmetry due to publication bias from that due to other factors. For example, if studies appear to be missing in areas of statistical non-significance (see Figure 1 for an example) then this adds credence to the possibility that the asymmetry is due to publication bias. Conversely, if the supposed missing studies are in areas of higher statistical significance (see Figure 2 for an example), this would suggest the cause of the asymmetry may be more likely to be due to factors other than publication bias (Table 1). If there are no statistically significant studies then publication bias may not be a plausible explanation for funnel plot asymmetry [69].

Sources of asymmetry in funnel plots

Selection bias
- Publication bias
- Location biases:
  - English language bias
  - Citation bias
  - Multiple publication bias

True heterogeneity
- Size of effect differs according to study size:
  - Intensity of intervention
  - Differences in underlying risk

Data irregularities
- Poor methodological design of small studies
- Inadequate analysis
- Fraud

Artefactual
- Choice of effect measure

Chance

Table 1

In interpreting funnel plots, systematic review authors thus need to distinguish the different possible reasons for funnel plot asymmetry listed in Table 1. Knowledge of the intervention, and the circumstances in which it was implemented in different studies, can help identify true heterogeneity as a cause of funnel plot asymmetry. There remains a concern that visual interpretation of funnel plots is inherently subjective. Therefore, we now discuss statistical tests for funnel plot asymmetry, and the extent to which they may assist in the objective interpretation of funnel plots. When review authors are concerned that small study effects are influencing the results of a meta-analysis, they may want to conduct sensitivity analyses in order to explore the robustness of the meta-analysis' conclusions to different assumptions about the causes of funnel plot asymmetry [36].

The Egger's test [18], the rank correlation test [13] and related tests [14, 21, 22, 27, 29] investigate whether small study effects are present. Methods to perform a meta-analysis when the size of the study and the size of effect are inversely associated have also been developed [70, 71]. Small study effects indicate a possible presence of publication bias [72, 25]. For example, when small studies with positive results are more likely to be published than small studies with negative results and all large studies have the same probability to be published, the size of the study and the size of effect are inversely associated. However, the presence of small study effects also has other possible causes including differences in the design of small and large studies and a poorer methodological quality in smaller studies [25, 72].

The "trim and fill" method estimates how many studies with the most negative results are missing from a meta-analysis and adjusts for the fact that they are missing [23, 24]. Other weight functions that have been considered in selection models include continuous functions of the p-value (e.g., negative-exponential) [32, 73], step functions with more than one cut point [32, 74, 75], and nonparametric functions [30, 76]. Most selection models that have been developed use maximum likelihood estimation [30, 31, 43, 74, 75, 76].

Some authors have argued that visual interpretation of funnel plots is too subjective to be useful. Terrin et al. found that researchers had only a limited ability to correctly identify funnel plots from meta-analyses subject to publication bias [77].

Other authors have proposed more sophisticated methods that avoid strong assumptions about the association between study P value and publication probability [30, 31]. These methods can be extended to estimate intervention effects, corrected for the estimated publication bias [78]. However, they require a large number of studies so that a sufficient range of study P values is included. A Bayesian approach in which the number and outcomes of unobserved studies are simulated has also been proposed as a means of correcting intervention effect estimates for publication bias [79]. Recent work has examined the possibility of assessing robustness over a range of weight functions, thus avoiding the need for large numbers of studies [75]. The complexity of the statistical methods, and the large number of studies needed, probably explain why selection models have not been widely used in practice [36].

Although these models are more complex than previous ones, Hedges still considers them to be unrealistic, as factors other than the p-value, such as the size and study design also play an important role in the decision to publish, both from the researchers' and journal editors' perspectives. Well-designed studies with null results will generally be published, but small poorly designed ones with very small p-values may not. It is when the p-value is in an intermediate range that the decision to publish may be greatly influenced by the p-value [38].

Selection models can be used to adjust a meta-analysis for suspected selected publication. Many different weight functions have been suggested to define the probabilities that studies are published. These methods are still in the experimental stage of development and, like the method of "trim and fill", can be used to assess the robustness of meta-analyses results to publication bias. Selection models are quite sophisticated and there is currently a lack of software to implement them. The weight functions suggested are often based on strict assumptions or very limited empirical data, which may be overlooked because of the complexity of these methods. Such methods have been used very rarely in practice, which can be attributed in part to their complexity. These models perform best when a large number of studies are being meta-analyzed (illustrative applications in the social sciences have used meta-analyses of up to several hundred studies). This is a severe limitation since many meta-analyses in medicine and related fields often include only a handful of studies [38].

It is not always clear how to interpret such results. Indeed, it is difficult even to state how large the number of studies must be in order to avoid the problems; this can depend to a great degree on the idiosyncrasies of the particular data set. However, it may be possible to combat these problems to a degree by reducing the number of parameters to be estimated. This can be accomplished by specifying a simpler weight function where there are fewer cut-points defining a smaller number of p-value ranges where the probability of selection is equal. While this eases the burden of estimating the weight function, results may vary to a considerable degree based on the choice of interval boundaries. Under such circumstances, it is better to regard the model as a tool for sensitivity analysis, and to consider variation in results with different cut-point specifications. Moreover, it is important in such cases to compare conclusions from the weight function model with results from other approaches, such as the "trim-and-fill" methods and the model with a priori specification of weights [20].

E. REFERENCES

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

2 STERLING T. D., (1959), "P*ublication decisions and their possible effects on inferences drawn from tests of significance-or vice versa*", J. Am. Stat. Assoc. 54: 30-34.

3 Begg C. B., Berlin J. A. (1988), "*Publication bias: A problem in interpreting medical data*", Journal of the Royal Statistical Society; Series A; 151(3):419-463.

4 Dickersin K., "*The existence of publication bias and risk factors for its occurrence*", Journal of the American Medical Association (1990): 263(10):1385-1389.

5 SIMES R. J. (1986), "*Publication bias: The case for an international registry of clinical trials.*", J. Clin. Oncol. 4 1529-1541.

6 Dickersin K., Yuan-I Min, "*Publication Bias: The Problem That Won't Go Away*", Annals of the New York Academy of Science, December (1993), Volume 703, Issue 1, p. 135-148.

7 DICKERSIN K., S. CHAN, T. C. CHALMERSH, S. SACKS & H. SMITH JR. (1987), "*Publication bias and clinical trials*", Controlled Clin. Trials 8: 343-353.

8 Chalmers I., "*Underreporting research is scientific misconduct.*", Journal of the American Medical Association 1990; 263(10):1405–1408.

9 International Committee of Medical Journal Editors. Available from: http://icmje.org/. [Accessed on 13 May 2014].

10 World Medical Association Declaration of Helsinki - ethical principles for medical research involving human subjects. Available from: http://www.wma.net/en/30publications/10policies/b3/17c.pdf. [Accessed on 13 May 2014].

11 Light RJ, Pillemer DB. (1984), "*Summing Up: The Science of Reviewing Research.*", Harvard University Press: Cambridge MA.

12 Ioannidis JP, Cappelleri JC, Lau J., "*Issues in comparisons between meta-analyses and large trials.*", Journal of the American Medical Association 1998; 279(14):p. 1089-1093.

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

14 Petra Macaskill, Stephen D. Walter, Les Irwig, "*A comparison of methods to detect publication bias in meta-analysis*", STATISTICS IN MEDICINE Statist. Med. 2001; 20:641-654, Copyright (2001) John Wiley & Sons, Ltd.

15 Lifeng Lin and Haitao Chu, "*Quantifying Publication Bias in Meta-Analysis*", Biometrics. September 2018; 74(3): 785–794. doi:10.1111/biom.12817.

16 Sterne JAC, Egger M., "*Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis.*", Journal of Clinical Epidemiology. 2001; 54:1046–1055. [PubMed: 11576817].

17 Sterne JAC, Egger M, Davey Smith G., "Investigating and dealing with publication and other biases in meta-analysis.", BMJ. 2001; 323:101–105. [PubMed: 11451790].

18 Egger M, Davey Smith G, Schneider M, Minder C., "*Bias in meta-analysis detected by a simple, graphical test.*", BMJ. 1997; 315:629–634. [PubMed: 9310563].

19 Stanley, Tom D. (2005), "*Beyond Publication Bias.*" Journal of Economic Surveys 19(3): 309–45.

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

21 Harbord RM, Egger M, Sterne JAC., "*A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints.*", Statistics in Medicine. 2006; 25:3443–3457. [PubMed: 16345038].

22 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L., "*Comparison of two methods to detect publication bias in meta-analysis.*", JAMA. 2006; 295:676–680. [PubMed: 16467236].

23 Duval S, Tweedie R., "*A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis.*", Journal of the American Statistical Association. 2000a; 95:89–98.

24 Duval S, Tweedie R., "*Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis.*", Biometrics. 2000b; 56:455–463. [PubMed: 10877304].

25 Sterne JAC, 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; 53:1119–1129. [PubMed: 11106885.

26 Terrin N, Schmid CH, Lau J, Olkin I., "*Adjusting for publication bias in the presence of heterogeneity.*", Statistics in Medicine. 2003; 22:2113–2126. [PubMed: 12820277].

27 Schwarzer Guido, Gerd Antes, and Martin Schumacher (2007), "*A Test for Publication Bias in Meta-Analysis with Sparse Binary Data.*" Statistics in Medicine 26(4): 721–33.
28 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L., "*Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity.*", Statistics in Medicine. 2007; 26:4544–4562. [PubMed: 17476644].
29 Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. Statistics in Medicine. 2008; 27:746–763. [PubMed: 17592831].
30 Dear KBG, Begg CB., "*An approach for assessing publication bias prior to performing a meta-analysis.*", Statistical Science. 1992; 7:237–245.
31 Hedges LV., "*Modeling publication selection effects in meta-analysis.*", Statistical Science. 1992; 7:246–255.
32 Silliman NP., "*Hierarchical selection models with applications in meta-analysis.*", Journal of the American Statistical Association. 1997a; 92(439):926–936.
33 Silliman NP., "*Nonparametric classes of weight functions to model publication bias.*", Biometrika. 1997b; 84:909–918.
34 Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR., "*Empirical assessment of effect of publication bias on meta-analyses.*", BMJ. 2000; 320:1574–1577. [PubMed: 10845965].
35 Villar J, Carroli G, Belizan JM., "*Predictive ability of meta-analyses of randomised controlled trials.*", Lancet 1995;345:772–6.
36 Julian PT Higgins and Sally Green, "*Cochrane Handbook for Systematic Reviews of Interventions*", https://handbook-5-1.cochrane.org/index.htm#chapter_10/10_4_1_funnel_plots.htm, Version 5.1.0, 10.4.1 Funnel plots.
37 Begg CB., Publication bias. In "*The Handbook of Research Synthesis*", Cooper H, Hedges LV (eds). Russell Sage Foundation: New York, 1994; Chapter 25.
38 Alexander J Sutton, Fujian Song , Simon M Gilbody and Keith R Abrams, "*Modelling publication bias in meta-analysis: a review*", Statistical Methods in Medical Research 2000; 9: 421–445.
39 Hedges LV., "*Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistically insignificant mean differences.*", Journal of Educational Statistics 1984; 9: 61–85.
40 Hedges LV, Olkin I., "*Statistical methods for meta-analysis.*", London: Academic Press, 1985.
41 Lane DM, Dunlap WP., "*Estimating effect-size bias resulting from significance criterion in editorial decisions.*", British Journal of Mathematical and Statistical Psychology 1978; 31: 107–12.
42 Hedges LV., "*Estimating the normal mean and variance under a publication selection model.*", in: Gleser LJ, Perlman MD, Press SJ, Sampson AR eds. "*Contributions to probability and statistics: essays in honor of Ingram Olkin.*", New York: Springer, 1989: 447–58.
43 Iyengar S, Greenhouse JB., "*Selection models and the file drawer problem [with discussion].*", Statistical Science 1988; 3(1): 109–117.
44 Champney TF., "*Adjustments for selection: publication bias in quantitative research synthesis.*", Doctoral thesis, University of Chicago, 1983.
45 DerSimonian R, Laird N., "*Meta-analysis in clinical trials.*", Controlled Clinical Trials 1986; 7: 177–88.
46 Dempster AP, Laird NM, Rubin DB., "Maximum likelihood from incomplete data via the EM algorithm.", Journal of the Royal Statistical Society Part B 1977; 39: 1–38.
47 Rust RT, Lehmann DR, Farley JU., "*Estimating publication bias in meta-analysis.*", Journal of Marketing Research 1990; 27: 220–26.
48 Bayes TR., "*An essay towards solving the doctrine of chances.*", Philos Trans R Soc Lond 1763;53:370.
49 Breslow NE., "*Biostatisticians and Bayes (with discussion).*", Statist Sci 1990;5:269–98.
50 Berry DA, Stangl DK., "*Bayesian biostatistics.*", New York: Marcel Dekker, 1996.
51 Phillips LD, "*Bayesian statistics for social scientists.*", London: Nelson, 1973.
52 Clayton DG, Hills M., "*Statistical models in epidemiology.*", Oxford: Oxford University Press, 1993.
53 Rosenthal R., "*The file drawer problem and tolerance for null results.*", Psychological Bulletin 1979; 86: 638–41.
54 Rosenthal R., "*Combining the results to independent studies.*", Professional Psychology 1978; 17: 136–37.
55 Gleser LJ, Olkin I., "*Models for estimating the number of unpublished studies.*", Statistics in Medicine 1996; 15: 2493–507.
56 Eberly LE, Casella G., "*Estimating the number of unseen studies.*", Technical report #1308– MA. Biometrics Unit, Cornell University, Ithaca, NY, 1996.

57 Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J., "*How important are comprehensive literature searches and the assessment of trial quality in systematic reviews?*", Empirical study. Health Technology Assessment 2003; 7: 1.
58 Schulz KF, Chalmers I, Hayes RJ, Altman DG., "*Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.*", JAMA 1995; 273: 408-412.
59 Davey Smith G, Egger M., "*Who benefits from medical interventions? Treating low risk patients can be a high risk strategy.*", BMJ 1994; 308: 72-74.
60 Glasziou PP, Iriwg LM., "*An evidence based approach to individualising treatment.*", BMJ 1995; 311: 1356-1359.
61 Stuck AE, Rubenstein LZ, Wieland D., "*Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity.*", Letter. BMJ 1998; 316: 469-471.
62 Schwarzer Guido, Gerd Antes, and Martin Schumacher (2002), "*Inflation of Type 1 Error Rate in Two Statistical Tests for the Detection of Publication Bias in Meta-Analyses with Binary Outcomes.*" Statistics in Medicine 21(17): 2465–77.
63 Irwig L, Macaskill P, Berry G, Glasziou P., "*Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased (Letter).*", British Medical Journal 1998; 316(7129):470.
64 Draper NR, Smith H., "*Applied Regression Analysis*", 2nd edn. Wiley: New York, 1981.
65 Agresti A., "*Categorical Data Analysis.*", Wiley: New York, 1990.
66 AJ Sutton, KR Abrams, DR Jones, TA Sheldon, F Song, "*Systematic reviews of trials and other studies*", Health Technology Assessment 1998; Vol. 2: No. 19.
67 Marcus R. Munafo, Taane G. Clark, Jonathan Flint, "*Assessing publication bias in genetic association studies: evidence from a recent meta-analysis*", Science Direct, Psychiatry Research 129 (2004) 39– 44, doi:10.1016/j.psychres.2004.06.011.
68 Peters J, Sutton AJ, Jones DR, Abrams KR, Rushton L., "*The contour enhanced funnel plot: an aid to interpreting funnel asymmetry.*", Journal of Clinical Epidemiology (in press, 2008).
69 Ioannidis JP, Trikalinos TA., "*The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey.*", Canadian Medical Association Journal 2007; 176: 1091-1096.
70 Moreno SG, Sutton AJ, Thompson JR, Ades AE, Abrams KR, Cooper NJ., "*A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity.*", Statistics in Medicine 2012; 31(14):1407–1417.
71 Rucker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M., "*Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis.*", Biostatistics 2011; 12(1):122–142.
72 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;343:d4002.
73 Preston C, Ashby D, Smyth R., "*Adjusting for publication bias: modelling the selection process.*", Journal of Evaluation in Clinical Practice 2004; 10(2):313–322.
74 Hedges L, Vevea J., "*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.
75 Vevea JL, Woods CM., "*Publication bias in research synthesis: sensitivity analysis using a priori weight functions.*", Psychological Methods 2005; 10(4):428–443.
76 Rufibach K., "*Selection models with monotone weight functions in meta-analysis.*", Biometrical Journal 2011; 53(4):689–704.
77 Terrin N, Schmid CH, Lau J., "*In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias.*", Journal of Clinical Epidemiology 2005; 58: 894-901.
78 Vevea JL, Hedges LV., "*A general linear model for estimating effect size in the presence of publication bias.*", Psychometrika 1995; 60: 419-435.
79 Givens GH, Smith DD, Tweedie RL., "*Publication bias in meta-analysis: a Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate.*", Statistical Science 1997; 12: 221-250.