Medicine®

The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses

Linyu Shi, MS, Lifeng Lin, PhD*

Abstract

Publication bias is a type of systematic error when synthesizing evidence that cannot represent the underlying truth. Clinical studies with favorable results are more likely published and thus exaggerate the synthesized evidence in meta-analyses. The trim-and-fill method is a popular tool to detect and adjust for publication bias. Simulation studies have been performed to assess this method, but they may not fully represent realistic settings about publication bias. Based on real-world meta-analyses, this article provides practical guidelines and recommendations for using the trim-and-fill method. We used a worked illustrative example to demonstrate the idea of the trim-and-fill method, and we reviewed three estimators (R_0, L_0, and Q_0) for imputing missing studies. A resampling method was proposed to calculate P values for all 3 estimators. We also summarized available meta-analysis software programs for implementing the trim-and-fill method. Moreover, we applied the method to 29,932 meta-analyses from the *Cochrane Database of Systematic Reviews*, and empirically evaluated its overall performance. We carefully explored potential issues occurred in our analysis. The estimators L_0 and Q_0 detected at least one missing study in more meta-analyses than R_0, while Q_0 often imputed more missing studies than L_0. After adding imputed missing studies, the significance of heterogeneity and overall effect sizes changed in many meta-analyses. All estimators generally converged fast. However, L_0 and Q_0 failed to converge in a few meta-analyses that contained studies with identical effect sizes. Also, P values produced by different estimators could yield different conclusions of publication bias significance. Outliers and the pre-specified direction of missing studies could have influential impact on the trim-and-fill results. Meta-analysts are recommended to perform the trim-and-fill method with great caution when using meta-analysis software programs. Some default settings (e.g., the choice of estimators and the direction of missing studies) in the programs may not be optimal for a certain meta-analysis; they should be determined on a case-by-case basis. Sensitivity analyses are encouraged to examine effects of different estimators and outlying studies. Also, the trim-and-fill estimator should be routinely reported in meta-analyses, because the results depend highly on it.

Abbreviations: CI = confidence interval, FDA = Food and Drug Administration, OR = odds ratio.

Keywords: meta-analysis, publication bias, systematic review, trim-and-fill method

1. Introduction

Publication bias has been a major threat to the validity of the conclusions of meta-analyses and systematic reviews.[1–4] It occurs when the studies in a meta-analysis are selectively published based on their results (e.g., the significance of their P values, the magnitude of their effect estimates, or their sample

sizes).[5] Studies with less significant results or smaller sample sizes are often more likely suppressed from publication, either by journal editors or authors themselves who may lack enthusiasm for publishing such studies.[6] Consequently, if publication bias appears in a meta-analysis, the synthesized effect estimates may be exaggerated in an artificially favorable direction. For example, Turner et al[7] identified a total of 74 studies of antidepressant agents that were registered in the US Food and Drug Administration (FDA); among them, 23 were not published. Overall, the effect sizes in the published studies increased by 32% compared with those in the FDA.

The best method to deal with publication bias is to retrieve related unpublished results as in Turner et al.[7] However, this method is often time-consuming and may be infeasible in many meta-analyses from the practical perspective. Also, the quality of the unpublished results without peer reviews may be questionable. Therefore, various statistical methods have been alternatively used to assess publication bias.[8–12] Among them, the trim-and-fill is one of the most popular methods over the past 20 years.[13–15] Based on a search on Google Scholar on 10 January 2019, Figure 1 shows the number of publications containing the exact phrase "trim-and-fill" year by year since the introduction of this method in 2000. The histogram presents a sharply increasing trend, especially after 2010.
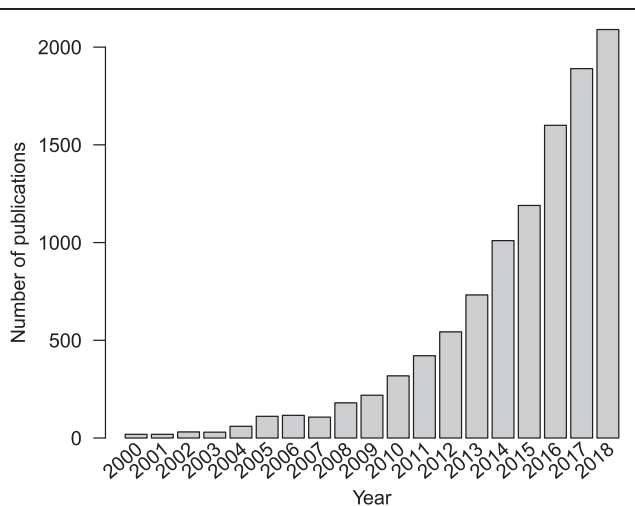
Figure 1. Histogram of publications that used the trim-and-fill method from 2000 to 2018.

Compared with other statistical methods (such as selection models[8]), the trim-and-fill method is relatively intuitive and efficient to detect and adjust for potential publication bias. It is a nonparametric approach based on examining the funnel plot's asymmetry. The funnel plot is widely and frequently used in meta-analyses for assessing publication bias[16]; it is a scatter plot with studies' effect sizes on the horizontal axis and their standard errors (or other measures of precision, e.g., sample sizes) on the vertical axis.[17–19] The funnel plot is supposed to be symmetrical if no publication bias appears.[9] Missing studies suppressed by publication bias in a meta-analysis usually lead to a noticeable asymmetrical funnel plot. Unlike other popular methods for detecting publication bias (such as various regression tests[9,20]), the trim-and-fill method not only indicates the significance of publication bias but also provide bias-adjusted results.[21] Therefore, this method attracts many evidence users in practical applications and is very effective to perform sensitivity analyses, especially when extracting unpublished results is infeasible and can be only approximated by statistically imputed missing studies.

The aims of this article are 2-folded. The trim-and-fill method is essentially a delicate statistical approach which involves non-trivial computing procedures, and most meta-analysts rely on user-friendly statistical programs (e.g., R, Stata, and SAS) to implement it. However, the implementation contains many important steps for determining the direction and magnitude of publication bias, and the statistical programs often provide default options for the steps which may be overlooked or even misused by their users. This article provides practical guidelines for appropriately and accurately using the trim-and-fill method.

In addition, the existing literature has examined the performance of the trim-and-fill method via extensive simulation studies,[14,22–24] which suggested that the method may be used with caution in the presence of substantial heterogeneity between studies in a meta-analysis. However, it is often difficult to justify the appropriateness of the simulation settings in clinical practice, because publication bias could be induced by many factors, including the studies' effect size magnitudes, P values, and sample sizes. In fact, the assumptions about the suppressed (missing) studies are dramatically different in different statistical methods

for publication bias: the trim-and-fill method assumes that the studies with the most extreme effect size magnitudes in an unfavorable direction are suppressed, while many other methods assume that the suppression depends only on P values or sample sizes.[8,9,20,25] Consequently, it is critical to evaluate the properties of the trim-and-fill method among real-world meta-analyses, which may be more practical and informative than simulated meta-analyses. This article applies the trim-and-fill method to a large collection of meta-analyses published in the *Cochrane Database of Systematic Reviews* and summarizes its overall performance. The findings offer useful recommendations for implementing the trim-and-fill method in future meta-analyses.

2. Methods

2.1. An illustrative example

Figure 2(A) shows the funnel plot of a meta-analysis by Andersen et al[26] (pages 55 and 56) on comparing the effects between antibiotics and placebo on preventing infection among patients with simple appendicitis after appendix surgeries. This meta-analysis combined 26 studies, containing a total of 2610 participants who took antibiotics after appendectomy and 2707 who took placebo. The endpoint was wound infection, which can be measured as a binary outcome. Regardless the original statistical analyses performed by Andersen et al,[26] here we used the odds ratio (OR) as the effect size based on the 2 × 2 table reported by each study and analyzed it on a logarithmic scale. A larger effect size indicated a worse wound infection, while a smaller effect size indicated a better condition of wound. Therefore, smaller effect sizes in the negative direction were likely favored in the publication process, and studies with larger effect sizes might be suppressed in the positive direction.

The distribution of the published studies displayed in the funnel plot is consistent with the foregoing observation. The lower right area in the funnel plot seems to contain some missing studies that tend to report more serious wound infections with larger standard errors (implying smaller sample sizes). Such missing studies have less desirable clinical results and thus are likely suppressed. Because of the potentially missing studies, the funnel plot looks fairly asymmetrical and strongly indicates publication bias. The overall log OR is estimated as −0.928 with 95% confidence interval (CI) [−1.172, −0.685]; that is, the OR is 0.395 with 95% CI [0.310, 0.504].

2.2. Fundamental idea of the trim-and-fill method

The trim-and-fill method aims at estimating potentially missing studies due to publication bias in the funnel plot and adjusting the overall effect estimate. The fundamental assumption of the trim-and-fill method is that the studies with the most extreme effect sizes, either on the left or on the right side, are suppressed. The direction of the missing studies depends on the expectation from general stakeholders (including patients, physicians, decision makers, and sponsors) on a case-by-case basis. For example, the studies in the foregoing illustrative example favor the negative direction of the OR, because they indicate reduced wound infections. Consider another example where the effect size remains to be the OR and the outcome is smoking cessation[27]; the OR in the positive direction may be favored in this case, because it indicates a higher smoking cessation rate and supports the effectiveness of interventions.
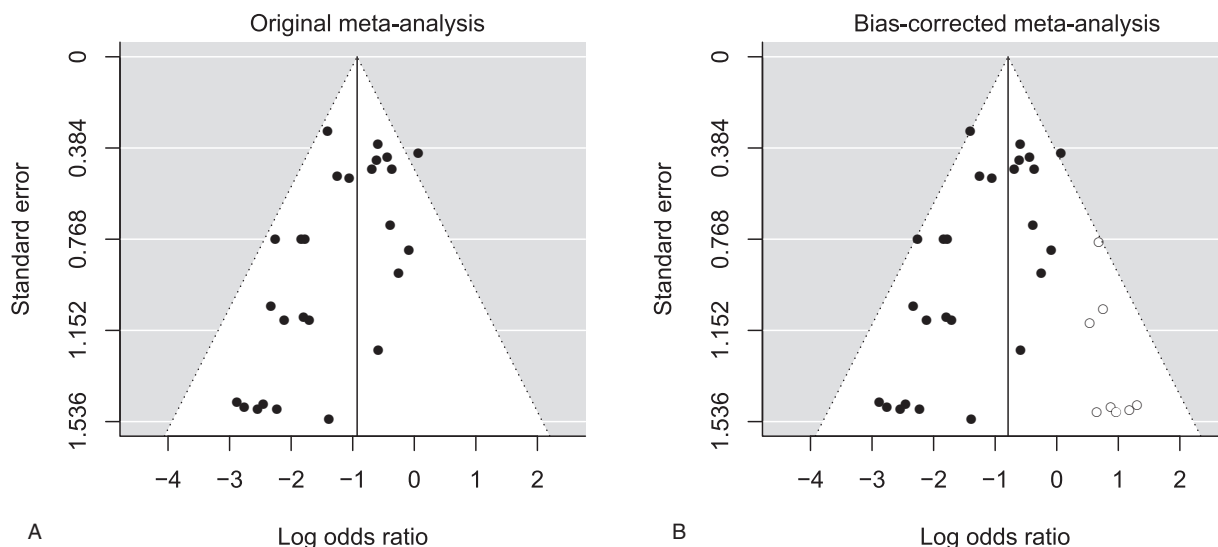
Figure 2. Funnel plots of the meta-analysis by Andersen et al[26] before (panel A) and after (panel B) applying the trim-and-fill method. The closed dots indicate the observed studies, and the open dots indicate the missing studies imputed by the trim-and-fill method (based on the estimator L_0). The dashed lines that create a triangular area indicate the 95% confidence limits (under the fixed-effect setting), and the vertical solid line represents the overall effect size.

The idea of the trim-and-fill method is to first *trim* the studies that cause a funnel plot's asymmetry so that the overall effect estimate produced by the remaining studies can be considered minimally impacted by publication bias, and then to *fill* imputed missing studies in the funnel plot based on the bias-corrected overall estimate.

In practice, without the true overall effect size, it is infeasible to identify the studies to be trimmed at the first step. An iterative algorithm is consequently applied to deal with this problem. Specifically, at the initial step, based on the estimated overall effect size in the original meta-analysis without any adjustment for publication bias, we can use a certain estimator (see Table 1) to estimate the number of missing studies. However, because the result may be subject to publication bias, we need to trim some studies (the number of which equals to the estimated number of missing studies) with effect sizes in the favorable direction (opposite to that of missing studies), so that the funnel plot becomes more symmetrical and the remaining studies are less influenced by publication bias. Then, we use the trimmed meta-analysis to obtain an updated overall effect estimate and continue to estimate the number of missing studies until it converges (i.e., remaining the same in 2 consecutive steps). Finally, the imputed missing studies are filled in the funnel plot to adjust for publication bias. Section 2.3 provides more details of the iterative algorithm.

In the current literature, the algorithm's convergence and the number of iterations that lead to the convergence are largely unknown. This article will present findings of these important properties of the trim-and-fill method. Moreover, in this article and in many realistic applications, studies collected in a meta-analysis is often considered heterogeneous,[28,29] and the random-effects meta-analysis model, such as the DerSimonian–Laird method,[30] is used to incorporate between-study heterogeneity in the overall effect estimate.[31]

Figure 2(B) presents the funnel plot with the missing studies imputed by the trim-and-fill method. Eight missing studies are filled in the plot, and the overall log OR becomes −0.790 with 95% CI [−1.024, −0.557].

2.3. Implementation

Table 1 gives the notation used in a meta-analysis and the trim-and-fill method, including 3 estimators R_0, L_0, and Q_0 for imputing the missing studies. The overall effect size θ is unknown and of primary interest; therefore, the foregoing 3 estimators of the number of missing studies cannot be directly calculated and need to be updated step by step using the iterative algorithm.

The original trim-and-fill method by Duval and Tweedie[13] tentatively assumes that k_0 studies with the most extreme effect sizes in the negative direction are suppressed due to publication bias. If the missing studies are in the positive direction as in the illustrative example in Section 2.1, one may simply invert the

Table 1	
Notation used in the trim-and-fill method assuming that studies with effect sizes in the negative direction are suppressed.	
Notation	**Definition**
n	The total number of observed (published) studies in a meta-analysis
y_i	The reported effect size of study i
s_i	The reported standard error of study i
θ	The true overall effect size
$\hat{\theta}$	The estimated overall effect size
τ^2	The between-study variance due to heterogeneity
$\hat{\tau}^2$	The estimated between-study variance
k_0	The true number of missing studies due to publication bias
$\hat{k}_0$	The estimated number of missing studies
x_i	The centralized study-specific effect sizes, $x_i = y_i - \theta$
r_i^*	The rank of the $|x_i|$ among all n studies
γ	The length of the rightmost run of ranks associated with positive values of the x_i
T_n	Wilcoxon rank test statistic, $T_n = \sum_{x_i > 0} r_i^*$
R_0	$R_0 = \gamma^* - 1$; it is truncated as 0 if $R_0 < 0$
L_0	$L_0 = \frac{4T_n - n(n+1)}{2n - 1}$; it is truncated as 0 if $L_0 < 0$ and rounded to the nearest integer
Q_0	$Q_0 = n - \frac{1}{2} - \sqrt{2n^2 - 4T_n + \frac{1}{4}}$; it is truncated as 0 if $Q_0 < 0$ and rounded to the nearest integer

direction of the original effect sizes (i.e., $\tilde{y}_i = -y_i$ and $\tilde{s}_i^2 = s_i^2$) so that the direction of missing studies becomes negative. Then, the standard trim-and-fill method can be directly applied to the inverted data $(\tilde{y}_i, \tilde{s}_i^2)$ and the final results are back-transformed by inverting the direction again.

In practice, the direction of potentially missing studies may differ case by case, and this must be pre-specified before performing the trim-and-fill method. False positive conclusions about publication bias may arise if this direction is wrongly specified. Many meta-analysis software programs use Egger regression[9] to suggest such a direction as the default. A positive intercept from Egger regression indicates the meta-analysis result tend to be biased toward the right side of the funnel plot, and the missing studies are in the negative direction.[11] In contrast, a negative intercept implies bias toward the left side, and the missing studies are likely in the positive direction. Although Egger regression provides a convenient way to determine the direction of missing studies, its decision may be opposite to that based on stakeholders' belief in some case, leading to invalid results from the trim-and-fill method. Therefore, for each specific meta-analysis, researchers should judge the direction of potentially missing studies by accounting for stakeholders' expectation, instead of relying only on Egger regression.

As its name suggests, the trim-and-fill method consists of 2 main steps. The first step aims at trimming k_0 studies in the opposite direction of missing studies so that the trimmed meta-analysis is less affected by publication bias. The iterative algorithm is applied to estimate the number of missing studies k_0 as well as the overall effect size θ; their estimates are denoted as $\hat{k}_0$ and $\hat{\theta}$, respectively. We use the illustrative example in Section 2.1 to show the process of this iterative algorithm. Egger regression shows that the missing studies tend to be on the right, which agrees with the funnel plot in Figure 2(A). Using the random-effects model, the initial overall effect size in the original meta-analysis is estimated as $\hat{\theta}^{(0)} = -0.928$. At the first iteration, based on the estimator L_0, the number of missing studies is estimated as $\hat{k}_0(1) = 7$ for the centralized data $x_i^{(1)} = y_i - \hat{\theta}^{(0)}$. Consequently, 7 studies with the most negative effect sizes on the left side are trimmed; based on the remaining 19 studies, the estimated overall effect size is updated as $\hat{\theta}^{(1)} = -0.808$. The estimated number of missing studies is further updated as $\hat{k}_0(2) = 8$ after using $\hat{\theta}^{(1)}$ to centralize the studies. We continue to trim 8 studies with the most negative effect sizes and obtain updated estimate $\hat{\theta}^{(2)} = -0.790$, which leads to $\hat{k}_0(3) = 8$, equaling to $\hat{k}_0(2)$ in the previous iteration; therefore, the estimated overall effect size remains to be $\hat{\theta}^{(3)} = -0.790$. The algorithm converges at the third iteration, and the final estimate of the number of missing studies is $\hat{k}_0 = \hat{k}_0(3) = 8$.

The second major step is to fill $\hat{k}_0 = 8$ missing studies in the funnel plot. Specifically, the estimate of the overall effect size $\hat{\theta}^{(3)} = \hat{\theta}^{(2)} = -0.790$ in the last iteration that achieves the convergence is used as the axis of symmetry, and we project 8 studies with the most negative effect sizes from the left side to the right side in the funnel plot as in Figure 2(B). Applying the random-effects model to the filled meta-analysis with the observed 26 studies and the imputed 8 missing studies, the final bias-adjusted estimate of the overall effect size is $\hat{\theta}_{tf} = -0.790$ with 95% CI [−1.024, −0.557].

Compared with the original meta-analysis, the estimated overall log OR is closer to zero, because the imputed missing studies are added in the positive direction. Also, its

95% CI shrinks slightly, possible because more studies are included in the meta-analysis and thus the result becomes more precise.

2.4. Calculation of P value

Besides adjusting for publication bias, most meta-analyses report P values to show the significance of the bias. The significance level for publication bias is often set to .1 (instead of the commonly-used .05[32]), because most statistical tests have low power to detect publication bias.[25] The P value of the trim-and-fill method is usually calculated based on the R_0 estimator, because it follows the negative binomial distribution under the null hypothesis of no publication bias. It is infeasible to derive the theoretical (closed-form) null distributions of L_0 and Q_0;[13] alternatively, we propose to use the resampling method to calculate the P values for all 3 estimators.

Referring to the notation in Table 1, the process of calculating the resampling-based P values is as follows. First, under the null hypothesis, we estimate the overall effect size as $\hat{\theta}$ and the between-study variance as $\hat{\tau}^2$ in the original meta-analysis. Also, we calculate the three estimators for the trim-and-fill method as R_0, L_0, and Q_0. Second, we generate a total of B (say, $B = 10,000$) resampled meta-analyses under the null hypothesis. For the bth resampled meta-analysis, we sample n within-study variances from those of the original meta-analysis, $s_{i\,i=1}^{2\,n}$, with replacement, and we denote them as $s_{(b)i\,i=1}^{2\,n}$. Also, we sample the effect sizes from $y_{(b)i} \sim N(\hat{\theta}, s_{(b)i}^2 + \hat{\tau}^2)$ under the null hypothesis. Third, for each resampled meta-analysis, we obtain the three trim-and-fill estimators and denote them as $R_{(b)0}$, $L_{(b)0}$, and $Q_{(b)0}(b = 1, \ldots, B)$. Finally, the resampling-based P values of the three estimators are calculated as

$$P_{R_0} = (B+1)^{-1} \sum_{b=1}^{B} [I(R_{(b)0} \geq R_0) + 1];$$

$$P_{L_0} = (B+1)^{-1} \sum_{b=1}^{B} [I(L_{(b)0} \geq L_0) + 1];$$

$$P_{Q_0} = (B+1)^{-1} \sum_{b=1}^{B} [I(Q_{(b)0} \geq Q_0) + 1],$$

where $I(\cdot)$ is the indicator function, and the constant one is artificially added to both numerator and denominator in each proportion to avoid calculating the P values as zero.

In the illustrative example in Section 2.1, based on $B = 10,000$ resampling iterations, the P values of R_0, L_0 and Q_0 are <.01, .02 and .03, respectively, all indicating significant publication bias. Moreover, the theoretical P value of R_0 (based on the negative binomial distribution under the null hypothesis) is also <.01.

2.5. Software programs

Many software programs are available to perform the trim-and-fill method. Table 2 provides a summary of several commonly-used programs, including the function trimfill() in the two R core packages for meta-analysis, "metafor"[33] (version 2.0–0) and "meta"[34] (version 4.9–2), the Stata (version 15) command metatrim,[35,36] the SAS (version 9.4) macro PUB_BIAS,[37] and the commercial program Comprehensive Meta-Analysis (CMA, version 3.0).[38] Of note, the Reviewer Manager (RevMan, version 5),[39] which is the software specifically used for

Table 2

Commonly-used software programs for performing the trim-and-fill method.

Software program	Estimator for the number of missing studies k_0
Function trimfill() in the R package "metafor"[*]	R_0, L_0 (default), and Q_0
Function trimfill() in the R package "meta"[†]	R_0, L_0 (default)
Stata command metatrim[‡]	R_0, L_0 (default), and Q_0
SAS macro PUB_BIAS[§]	R_0 (default)
Comprehensive Meta-Analysis[‖]	L_0 (default)

[*] https://cran.r-project.org/package=metafor.
[†] https://cran.r-project.org/package=meta.
[‡] https://www.stata.com/support/faqs/statistics/meta-analysis/.
[§] https://analytics.ncsu.edu/sesug/2006/PO04_06.PDF.
[‖] https://www.meta-analysis.com/.

preparing and maintaining Cochrane reviews, cannot perform the trim-and-fill method.[40]

All programs except the SAS macro use L_0 as the default trim-and-fill estimator; R_0 is the only option in the SAS macro PUB_BIAS. The R package "metafor" and Stata command metatrim can implement all three estimators, while Q_0 is not available in the R package "meta". The Stata macro metatrim refers to R_0, L_0, and Q_0 as "run", "linear", and "quadratic", respectively. In addition, the direction of missing studies is usually determined using Egger regression in these programs by default, while CMA is a menu-driven program and asks users to choose the direction of missing studies.

2.6. Empirical evaluation

To comprehensively assess the performance of the trim-and-fill method among realistic meta-analyses, we applied the method to a large collection of meta-analyses from the *Cochrane Database of Systematic Reviews*, which offers leading sources of evidence on healthcare-related topics.

We collected all Cochrane reviews from 2003 Issue 1 to 2018 Issue 5, and downloaded their data iteratively using the R package "RCurl"[41] in May 2018. We selected meta-analyses containing at least 5 studies from all reviews and classified them into 2 groups based on their outcomes (binary or non-binary). For each meta-analysis with binary outcomes, we used the log OR as the effect size, regardless the choice of effect size in its original review. When zero counts existed in a study's 2×2 table, the continuity correction of 0.5 was added to all 4 data cells to adjust the (log) OR and its variance.[39] If both groups of the study had no events, the (log) OR was not estimable and was removed from our analysis.

We used the package "metafor"[33] (version 2.0–0) in R (version 3.5.1) to perform the trim-and-fill method for each eligible meta-analysis based on all three estimators R_0, L_0, and Q_0, and obtained the number of iterations to achieve the trim-and-fill algorithm's convergence. Also, we estimated the number of missing studies, the original and bias-adjusted overall effect sizes, and the P value of the Q test for heterogeneity in each meta-analysis.

Moreover, using some convenience samples, we compared the P values of all three estimators and investigated the effects of outlying studies on the trim-and-fill results. The outlying studies were identified using the diagnostics proposed by Viechtbauer and Cheung[42] under the random-effects setting. Specifically, a

residual of each study in a meta-analysis was calculated and was expected to approximately follow the standard normal distribution; the study was considered outlying if its residual was larger than three in absolute magnitude. In addition, we carefully explored the potential issues occurred when performing the trim-and-fill method. We summarized the method's overall empirical performance and provided practical recommendations.

No ethical approval and patient consent were required in our study, because this article focused on statistical methods for meta-analyses, and all analyses were performed based on published data in the literature.

3. Results

3.1. Overall empirical performance of the trim-and-fill method

3.1.1. The estimated number of missing studies. In total, our analysis included 18,562 meta-analyses with binary outcomes and 11,370 with non-binary outcomes. The upper panels in Figures 3 and 4 show the frequencies of the estimated number of missing studies $\hat{k}_0$ based on the three estimators among meta-analyses with binary and non-binary outcomes, respectively. Recall that $\hat{k}_0 = 0$ implied no publication bias. The R_0 estimator ranged from 0 to 38 for binary outcomes and from 0 to 18 for non-binary outcomes; it detected publication bias in much less meta-analyses than L_0 (ranging from 0 to 34 for binary outcomes and from 0 to 24 for non-binary outcomes) and Q_0 (ranging from 0 to 62 for binary outcomes and from 0 to 113 for non-binary outcomes). The R_0 estimator was zero in 11,147 (60.1%) meta-analyses with binary outcomes, while 6099 (32.9%) ones had $L_0 = 0$ and $Q_0 = 0$. Among meta-analyses with non-binary outcomes, 7495 (65.9%) had $R_0 = 0$ and 5168 (45.5%) had $L_0 = 0$ and $Q_0 = 0$. The R_0 estimator tended to detect less missing studies than L_0 and Q_0 for both outcome types.

The L_0 and Q_0 were estimated as zero among the same meta-analyses, possibly because their mathematical formulas were similar in the absence of publication bias ($k_0 = 0$). However, in the presence of publication bias ($k_0 > 0$), L_0 tended to detect less missing studies than Q_0. In general, L_0 detected 1 or 2 missing studies in many meta-analyses. As shown in Figure 3(C) and 4(C), the distributions of Q_0 had heavy right tails; Q_0 detected at least ten missing studies in much more meta-analyses than the other 2 estimators.

3.1.2. The number of iterations to achieve the trim-and-fill algorithm's convergence. The lower panels in Figures 3 and 4 present the number of iterations to achieve the convergence of the trim-and-fill algorithm using the three estimators among the Cochrane meta-analyses with binary and non-binary outcomes, respectively. All 3 estimators converged fast in most cases; the frequencies generally had a decreasing trend as the number of iterations increased. The R_0 estimator converged slightly faster than L_0, while both converged within four iterations in around 98% meta-analyses for both outcome types. Furthermore, L_0 tended to converge faster than Q_0.

3.1.3. Changes of heterogeneity and overall effect size. Table 3 presents the changes of the significance of heterogeneity (based on the Q test) and overall effect sizes after applying the trim-and-fill method to the Cochrane meta-analyses. Of note, the significant level was set to .05 here, which was different from that for publication bias tests. For most meta-analyses with binary outcomes, their heterogeneity remained non-significant after
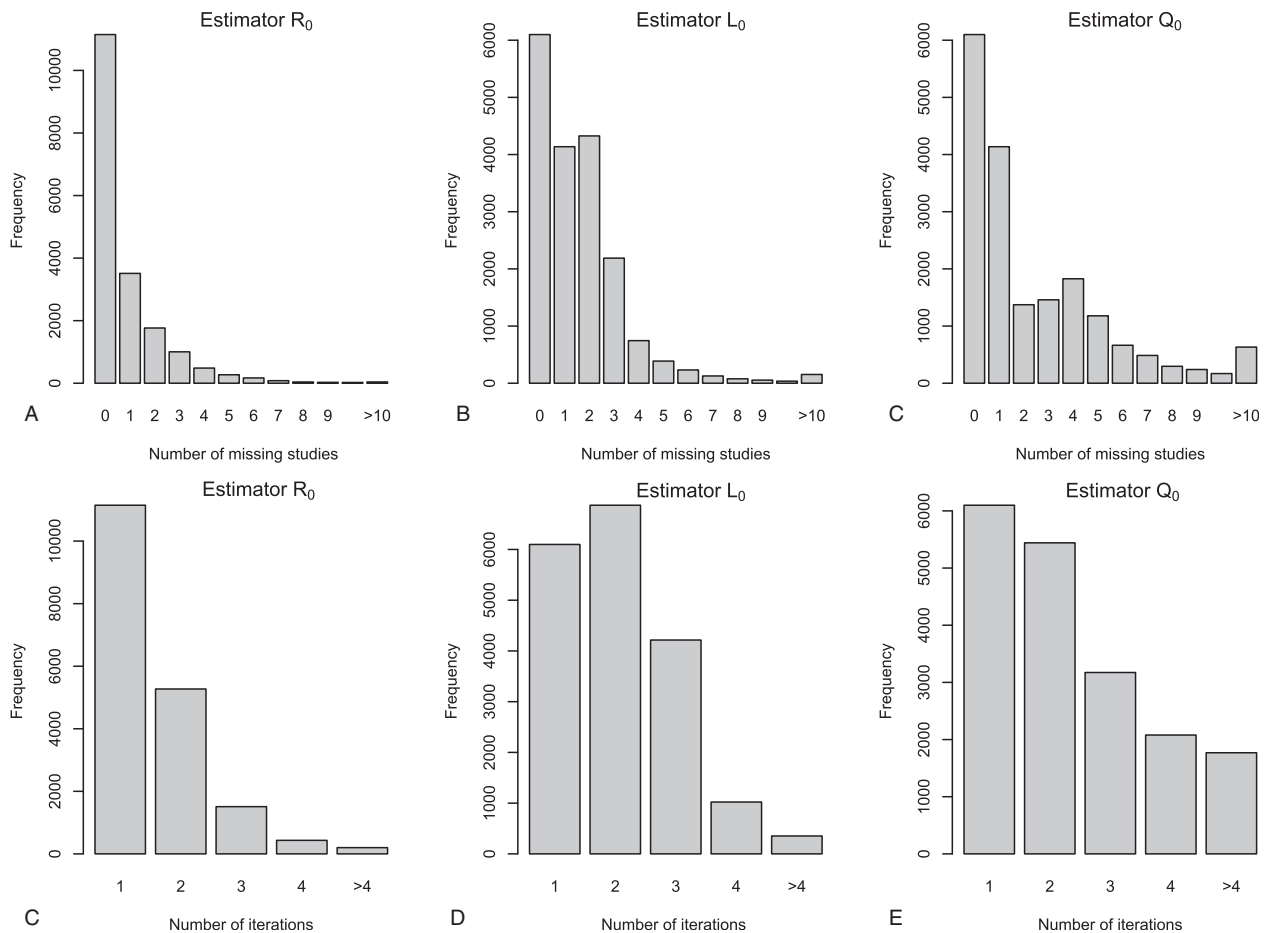
Figure 3. Histograms of the estimated number of missing studies (upper panels) and the number of iterations to achieve the convergence of the trim-and-fill algorithm (lower panels) based on the three estimators R_0, L_0 and Q_0 among Cochrane meta-analyses with binary outcomes.

using the trim-and-fill method, while around 20% meta-analyses remained significantly heterogeneous. A noticeable proportion of meta-analyses (about 5–15%) were not significantly heterogeneous before using the trim-and-fill method, but their heterogeneity became significant after adding imputed missing studies. The heterogeneity in only few meta-analyses (much less than 1%) was originally significant and became non-significant when using the trim-and-fill method. These were possibly because adding imputed missing studies likely extended the distribution range of a meta-analysis and thus made the whole set of studies more heterogeneous. For example, referring to the illustrative example in Figure 2, the log OR mostly ranged from −3 to 0 in the original meta-analysis, and the range extended to from −3 to 1.5 after incorporating the eight imputed missing studies. In addition, Q_0 changed the significance of heterogeneity in more meta-analyses than the other 2 estimators, possibly because it could detect more missing studies (Fig. 3) and thus lead to more heterogeneous bias-adjusted meta-analyses.

When the outcomes were non-binary, a larger proportion of meta-analyses was significantly heterogeneous compared with those with binary outcomes: around 50% of meta-analyses with non-binary outcomes remained significantly heterogeneous after using the trim-and-fill method. The other trends were similar to those for binary outcomes.

For both binary and non-binary outcomes, the significance of the estimated overall effect sizes $\hat{\theta}$ changed in a noticeable number of meta-analyses after using the trim-and-fill method. The corresponding proportions roughly ranged from 1% to 8%, depending on the estimators used to impute missing studies; Q_0 inverted the significance in more meta-analyses than the other 2 estimators. The inverted significance of $\hat{\theta}$ might be interpreted from two perspectives. On the one hand, the imputed missing studies were likely in an unfavorable direction and adding them into the meta-analysis might move $\hat{\theta}$ toward the null value, so that $\hat{\theta}$ might become non-significant. On the other hand, incorporating missing studies in the meta-analysis effectively increased the total sample size and thus the precision, so that the CI of $\hat{\theta}$ might shrink and $\hat{\theta}$ might become significant if the true effect size was away from the null value. Because the significance of $\hat{\theta}$ depended on various factors, it should be explored case by case.

3.2. Issues occurred when performing the trim-and-fill method
3.2.1. Case of failing to calculate the trim-and-fill estimator.
When we used Q_0 to implement the trim-and-fill method among the Cochrane meta-analyses with binary outcomes, 3784 (20.39%) produced errors in estimating the number of missing
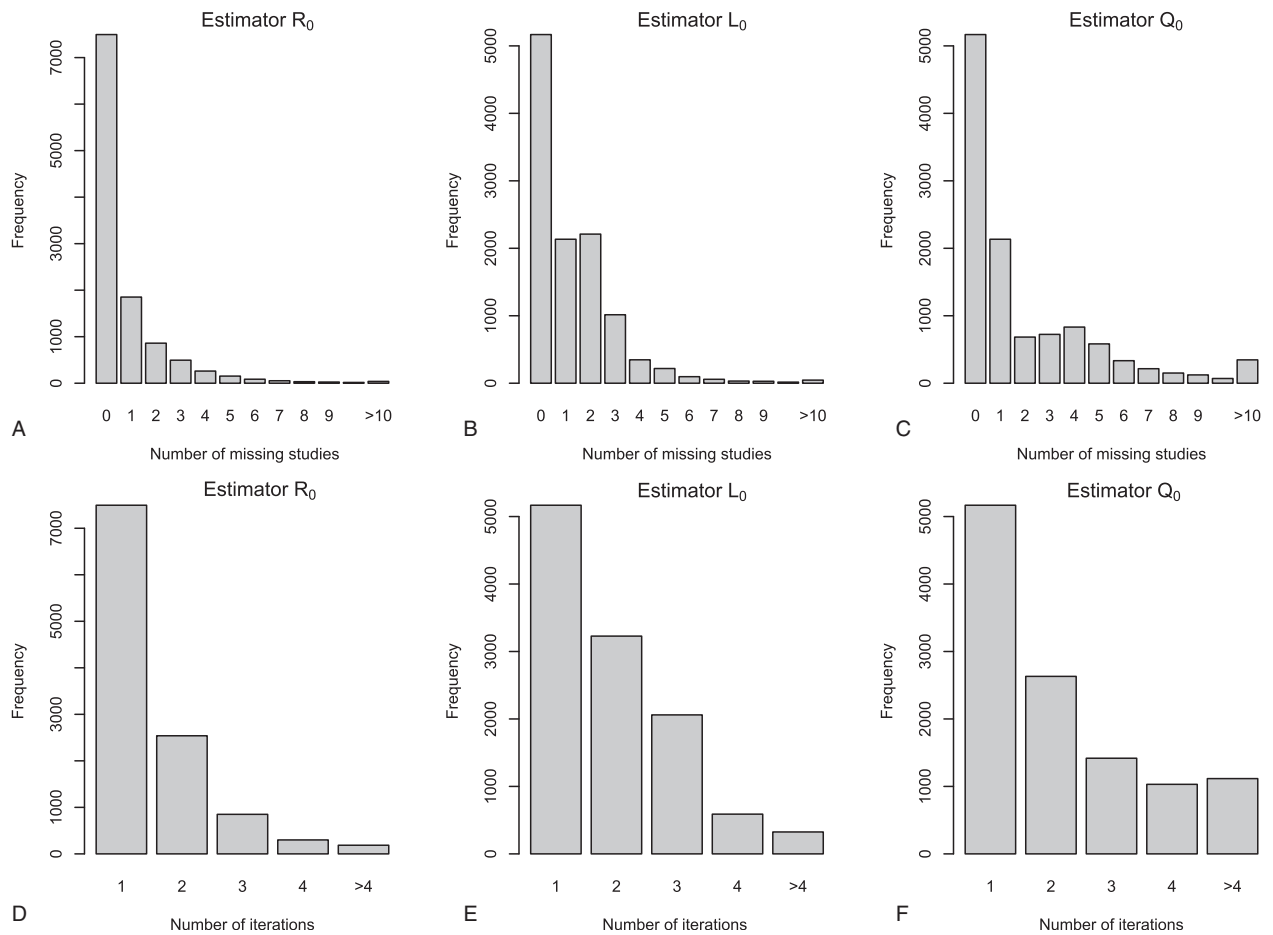
Figure 4. Histograms of the estimated number of missing studies (upper panels) and the number of iterations to achieve the convergence of the trim-and-fill algorithm (lower panels) based on the three estimators R_0, L_0 and Q_0 among Cochrane meta-analyses with non-binary outcomes.

Table 3

Changes of the significance of the heterogeneity (based on the Q test) and the estimated overall effect size ($\hat{\theta}$) before and after applying the trim-and-fill method to Cochrane meta-analyses.

Estimator	Non-significance unchanged	Changing from non-significance to significance	Significance unchanged	Changing from significance to non-significance
Among the 18,562 Cochrane meta-analyses with binary outcomes[*]:				
Heterogeneity (based on the Q test)				
R_0	13,630 (73.43%)	1,031 (5.55%)	3,890 (20.96%)	11 (0.06%)
L_0	13,046 (70.28%)	1,615 (8.70%)	3,880 (20.90%)	20 (0.11%)
Q_0	12,045 (64.89%)	2,615 (14.09%)	3,886 (20.94%)	15 (0.08%)
Overall effect size				
R_0	9,963 (53.67%)	311 (1.68%)	7,694 (41.45%)	594 (3.20%)
L_0	9,648 (51.98%)	625 (3.37%)	7,378 (39.75%)	910 (4.90%)
Q_0	8,720 (46.98%)	1,554 (8.37%)	6,900 (37.17%)	1,387 (7.47%)
Among the 11,370 Cochrane meta-analyses with non-binary outcomes[†]:				
Heterogeneity (based on the Q test)				
R_0	5,111 (44.95%)	628 (5.52%)	5,628 (49.50%)	3 (0.03%)
L_0	4,767 (41.93%)	972 (8.55%)	5,622 (49.44%)	9 (0.08%)
Q_0	4,373 (38.46%)	1,360 (11.96%)	5,623 (49.45%)	8 (0.07%)
Overall effect size				
R_0	4,144 (36.45%)	156 (1.37%)	6,581 (57.88%)	489 (4.30%)
L_0	3,835 (33.73%)	465 (4.09%)	6,564 (57.73%)	506 (4.45%)
Q_0	3,505 (30.83%)	793 (6.97%)	6,317 (55.56%)	749 (6.59%)

The significance level is .05.

[*] The L_0 and Q_0 estimators did not converge in one meta-analysis with binary outcomes.

[†] The Q_0 estimator did not converge in 6 meta-analyses with non-binary outcomes.

studies; specifically, R displayed NaN ("not a number") for Q_0. Similar issues occurred for non-binary outcomes; Q_0 produced NaN in 1880 (16.53%) meta-analyses with non-binary outcomes. Such unusual results were because these meta-analyses led to negative values inside the square root in the formula of Q_0; see Table 1. To avoid this issue, if the value inside the square root was negative, we slightly revised the formula of Q_0 by setting this negative value to 1/4, so that Q_0 was estimated as $n-1$; this was the maximum number of missing studies that could be estimated by the trim-and-fill method. After the revision, Q_0 was successfully calculated in all meta-analyses.

3.2.2. Case of the trim-and-fill method failing to converge.
In 2 Cochrane meta-analyses with binary outcomes and 6 with non-binary outcomes, the trim-and-fill algorithm did not converge when using L_0 and Q_0 due to different reasons. In these cases, the estimators continued to oscillate (e.g., calculated as 4, 5, 4, 5, 4, and so on) after sufficient iterations so that the algorithm could not converge.

First, L_0 in one meta-analysis with binary outcomes by Li et al[43] (page 42) did not converge. This meta-analysis included 12 studies comparing effects of intravenous magnesium with placebo on myocardial infarction, and the outcome was cardiogenic shock. Potential missing studies were on the right side because negative effect sizes indicated less cardiogenic shocks and thus better treatment effects. The L_0 estimator continuously oscillated between 4 and 5 after 2 trim-and-fill iterations, while R_0 and Q_0 converged to 2 and 11, respectively. The oscillation of L_0 was possibly because the effect sizes of several studies were fairly close to zero and 1 study had an overwhelming weight (>90%) in this meta-analysis. Therefore, the estimated overall effect size $\hat{\theta}$ differed little during the trim-and-fill iterations, and such slight differences caused two studies to continuously exchange their ranks r_i^* in the term T_n used to calculate L_0 (see Table 1). Consequently, L_0 was rounded to either 4 or 5 and did not converge.

Second, Q_0 in a few meta-analyses also did not converge. For example, in a meta-analysis with binary outcomes by Spooner et al[44] (page 48), Q_0 oscillated between 3 and 4, while both R_0 and L_0 converged to 2. This meta-analysis contained five studies on the effect of mast-cell substance on preventing exercise-induced bronchoconstriction or asthma among children. In another meta-analysis by Heiwe et al[45] (page 281), Q_0 also did not converge. This meta-analysis compared the effects of 4-to-6-month cardiovascular exercise with control group on a non-binary outcome (the maximum heart rate), and the effect size was the mean difference. Both meta-analyses potentially had missing studies on the right side in the funnel plot. In each of them, the 2 rightmost studies with the largest effect sizes had identical point estimates and standard errors, and thus they overlapped in the funnel plots. The oscillation of Q_0 was likely caused by intrinsic computational inaccuracy in R when the "metafor" package analyzing the 2 rightmost studies with identical effect sizes. For example, interestingly, the output of $1.2/0.2 - 6$ in R (version 3.5.1) was not exactly zero; instead, R returned a very tiny negative value. Although the computational inaccuracy was tiny, it impacted the calculation of studies' ranks r_i^* for Q_0 (see Table 1).

Similarly, Q_0 did not converge in other 5 meta-analyses with non-binary outcomes, because in each of them at least 2 studies had identical point estimates of their effect sizes (while their standard errors might differ) and their effect sizes were the most extreme ones (in either the negative or positive direction).

Therefore, these studies fell on the same vertical line in the funnel plot, and Q_0 could not converge also due to intrinsic computational inaccuracy in R.

These findings suggested that Q_0 might not converge if a meta-analysis contained at least 2 studies with identical point estimates of effect sizes. Among the Cochrane meta-analyses, 1308 (7.05%) with binary outcomes and 2294 (20.18%) with non-binary outcomes had at least two such studies. Furthermore, 512 (2.76%) and 151 (1.33%) meta-analyses with binary and non-binary outcomes had at least 2 studies with identical point estimates and also identical standard errors, respectively. Nevertheless, if the corresponding studies were not the most extreme ones in the meta-analysis, the identical effect sizes did not cause the issue of convergence.

3.2.3. Disagreement between P values based on different estimators.
Table 4 presents P values of all 3 trim-and-fill estimators in a convenience sample of three meta-analyses by Andersen et al,[26] Li et al,[43] and Spooner et al[44] which have been introduced before. Because L_0 or Q_0 oscillated in the last 2 meta-analyses, we stopped the trim-and-fill algorithm at the 100th iteration so that the P value could be calculated.

The P values of R_0 based on the theoretical distribution and the resampling method were fairly similar in all 3 meta-analyses; however, the P values of different estimators might lead to different conclusions about publication bias significance. For example, in the meta-analysis by Li et al,[43] both R_0 and L_0 indicated non-significant publication bias, while Q_0 implied significant bias (at the level 0.1). In the meta-analysis by Spooner et al,[44] R_0 detected significant publication bias, while the other 2 estimators did not.

3.2.4. Impact of outlying studies on the trim-and-fill method.
We used the meta-analysis by Khanna et al[46] (page 346) to illustrate the impact of outlying studies on the trim-and-fill method. This meta-analysis collected 8 studies comparing effects of aripiprazole and clozapine on adverse effects among patients who experienced at least one adverse effect; the outcome was binary. Based on the outlier detection method by Viechtbauer and Cheung[42] under the random-effects setting, the rightmost study had a residual larger than 3 and thus was considered outlying. All three trim-and-fill estimators were zero, implying no publication bias. However, if this outlying study was removed, both L_0 and Q_0 became 2, while R_0 remained to be zero.

4. Discussion

4.1. Main findings

This article has illustrated the use of the trim-and-fill method and investigated its performance and potential issues in realistic

Table 4

P values of the 3 trim-and-fill estimators in 3 Cochrane meta-analyses.

Meta-analysis	P value			
	R_0 estimator[*]	R_0 estimator[†]	L_0 estimator[†]	Q_0 estimator[†]
Andersen et al	<.01	<.01	.02	.03
Li et al	.22	.25	.15	.06
Spooner et al	<.01	.02	.12	.11

[*] Based on the theoretical negative binomial distribution.
[†] Based on the resampling method with 10,000 resampling iterations.

applications based on a large collection of Cochrane meta-analyses. The trim-and-fill algorithm can mostly converge fast within 4 iterations, while a few meta-analyses required much more iterations to achieve the convergence, especially when using the Q_0 estimator.

In the original papers by Duval and Tweedie,[13,14] both R_0 and L_0 are recommended. The L_0 estimator may be preferable when a meta-analysis contains over 25% missing studies because it likely has a smaller mean squared error. The R_0 estimator is preferable when a meta-analysis has many observed studies (n is large)[2]; also, it has the simplest formula and its null distribution is exactly the negative binomial distribution which can be used to calculate the theoretical P value of publication bias. Our study has shown that both L_0 and Q_0 detected missing studies in noticeably more meta-analyses than R_0, and the estimated number of missing studies by Q_0 was often larger than that by L_0.

In meta-analyses that contained studies with identical effect sizes, L_0 and Q_0 may fail to converge, while R_0 had no such issues. Also, although only R_0 had a closed-form null distribution to yield P values, we have illustrated that the resampling method can be used to calculate P values for all 3 estimators. Based on convenience samples of three meta-analyses, P values based on different estimators could inform difference extents of publication bias significance. In addition, we have shown that outlying studies could have great impact on the trim-and-fill method.

4.2. Strengths and limitations

Our study was based on a total of 29,932 Cochrane meta-analyses, which are considered high-quality evidence for aiding healthcare-related decisions. We extracted the Cochrane meta-analyses at the subgroup-specific level, so these meta-analyses were not too heterogeneous. The trim-and-fill method was performed under the random-effects setting to properly account for heterogeneity.[31] Using the realistic meta-analyses, instead of simulated meta-analyses as in several existing studies,[14,22–24] our findings could offer practical guidelines and recommendations for the trim-and-fill method, because simulations may not fully induce the true mechanism of publication bias.

Nevertheless, our analysis had several limitations. First, because the database of meta-analyses used in this article was huge, it was infeasible to identify the true status of publication bias and the direction of potentially missing studies in each meta-analysis. Egger regression was used to determine the direction of missing studies, but it might be inaccurate. Second, we used only the R package "metafor" to implement the trim-and-fill method. Different programs may have slightly different implementation details, for example, when rounding the trim-and-fill estimators[13,14]; therefore, the results produced by different programs may not be identical. Third, our analysis depended on the assumptions made by the trim-and-fill method to assess publication bias; however, they may be violated in some cases. The trim-and-fill method is based on the funnel plot, while the funnel plot's asymmetry may be attributable to some other factors besides publication bias.[47] We could not rule out the confounders that might cause the asymmetry of funnel plots in our large database of meta-analyses.

4.3. Practical implications

Although the trim-and-fill method is attractive and popular, its implementation requires extensive statistical coding, so nearly all meta-analysts rely on certain statistical programs to perform the method. Our study indicates that cautions are needed when meta-analysts perform the trim-and-fill method and report its results.

First, as shown in Figures 3 and 4 and Table 3, the results of the trim-and-fill method depend greatly on the selected estimator (R_0, L_0, or Q_0) for imputing missing studies. In the current literature, the specific estimator is rarely reported. We recommend meta-analysts to try all three estimators as sensitivity analyses because different estimators may be advantageous in different situations, and to clearly specify the estimators used in their analyses.

Second, heterogeneity and publication bias may mutually impact each other. On the one hand, as shown in Table 3, many meta-analyses became heterogeneous after incorporating imputed missing studies by the trim-and-fill method. Publication bias could noticeably influence the estimate of heterogeneity.[48] On the other hand, existing studies have shown that substantial heterogeneity may seriously impair the power of the trim-and-fill method.[23] In the presence of both substantial heterogeneity and publication bias, meta-analysts should carefully examine whether the funnel plot's asymmetry is truly caused by publication bias or confounded by heterogeneity.[49] More sophisticated statistical methods (e.g., selection models) may be used to better model publication bias in such cases.[22]

Third, meta-analysts should be aware of the potential issues when performing the trim-and-fill method. For example, our empirical analysis has shown that some trim-and-fill estimators may not converge or cannot be calculated using some meta-analysis programs. These issues are likely due to some intrinsic computational inaccuracy of the programs when the meta-analysis contains studies with identical effect sizes. Meta-analysts might want to use different programs to implement the trim-and-fill method and check if the issues continue to occur.

Fourth, all 3 trim-and-fill estimators can yield P values using the resampling method, and these P values may inform different extents of publication bias significance. Again, it is critical to report the estimator used in the meta-analysis, so that the conclusion about publication bias can be replicable.

Fifth, outlying studies should be carefully investigated when using the trim-and-fill method. The studies with extreme outlying effect sizes may greatly influence the detection of publication bias. Different directions of the outliers may mislead meta-analysis results to different extents, and the statistical power of the trim-and-fill method may be sharply decreased due to outliers. Suppose missing studies are truly in the negative direction. If the outliers are in the opposite direction, then the estimated overall effect size $\hat{\theta}$ (before adjusting publication bias) may be too positive, and the trim-and-fill method may wrongly estimate the number of missing studies. On the other hand, if the outliers and the potentially missing studies happen to be in the same detection, the outliers may mask the missing studies, and the trim-and-fill method may fail to detect publication bias. Sensitivity analyses that include and exclude potential outliers may be considered to assess the impact of the outliers on the meta-analysis conclusions.[50]

Last but not least, the direction of missing studies is essential when performing the trim-and-fill method. Some meta-analysis programs (such as the R package "metafor" used in our analysis) apply Egger regression to automatically determine this direction, and some (such as CMA) may simply use the negative direction as the default. However, neither option is guaranteed to be correct for a specific meta-analysis. Egger regression is known to produce inflated false positive rates for (log) OR,[20,25,51] so its estimated

intercept may not accurately reflect the true direction of missing studies. In practice, meta-analysts must incorporate stakeholders' preferences to decide the direction. If the direction is determined wrongly, the trim-and-fill results will change greatly and are completely invalid. For example, in the illustrative example in Section 2.1, missing studies are expected to be in the positive direction, because negative effect sizes imply less wound infections and are favored by stakeholders. If we set the direction of missing studies to be negative and re-apply the trim-and-fill method to the illustrative example, none of the three estimators implies significant publication bias.

4.4. Future work

The Cochrane meta-analyses used in our study do not have gold standards to ascertain publication bias, and it is infeasible to obtain the true magnitude and direction of publication bias in each Cochrane meta-analysis. Our future work includes applying the trim-and-fill method to meta-analyses in which unpublished studies can be retrieved from certain sources,[7,52] such as clinical trial registries including ClinicalTrials.gov by the US National Institutes of Health and the International Clinical Trials Registry Platform by the World Health Organization. Such meta-analyses permit the comparison between the actual unpublished studies with the imputed missing studies by the trim-and-fill method. Also, the current trim-and-fill method is applicable only to univariate meta-analyses. Methodological work is highly needed to generalize it to multivariate and network meta-analyses of multiple outcomes and multiple treatments.[53]

Author contributions

Conceptualization: Lifeng Lin.
Data curation: Lifeng Lin.
Formal analysis: Linyu Shi.
Methodology: Linyu Shi, Lifeng Lin.
Software: Linyu Shi.
Supervision: Lifeng Lin.
Validation: Lifeng Lin.
Visualization: Linyu Shi.
Writing – original draft: Linyu Shi.
Writing – review & editing: Lifeng Lin.
Lifeng Lin orcid: 0000-0002-3562-9816.

References

[1] Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990;263:1385–9.
[2] Rothstein HR, Sutton AJ, Borenstein M. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chichester, UK: John Wiley & Sons; 2005.
[3] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.
[4] Gurevitch J, Koricheva J, Nakagawa S, et al. Meta-analysis and the science of research synthesis. Nature 2018;555:175–82.
[5] Bax L, Moons KG. Beyond publication bias. J Clin Epidemiol 2011;64:459–62.
[6] Nissen SB, Magidson T, Gross K, et al. Publication bias and the canonization of false facts. ELife 2016;5:e21451.
[7] Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252–60.
[8] Sutton AJ, Song F, Gilbody SM, et al. Modelling publication bias in meta-analysis: a review. Stat Methods Med Res 2000;9:421–45.
[9] Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–34.
[10] Jin Z-C, Zhou X-H, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med 2015;34:343–60.
[11] Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics 2018;74:785–94.
[12] Lin L, Chu H, Murad MH, et al. Empirical comparison of publication bias tests in meta-analysis. J Gen Intern Med 2018;33:1260–7.
[13] Duval S, Tweedie R. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc 2000;95:89–98.
[14] Duval S, Tweedie R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56:455–63.
[15] Sutton AJ, Duval SJ, Tweedie RL, et al. Empirical assessment of effect of publication bias on meta-analyses. BMJ 2000;320:1574–7.
[16] Herrmann D, Sinnett P, Holmes J, et al. Statistical controversies in clinical research: publication bias evaluations are not routinely conducted in clinical oncology systematic reviews. Ann Oncol 2017;28:931–7.
[17] Light RJ, Pillemer DB. Summing Up: The Science of Reviewing Research. Cambridge, MA: Harvard University Press; 1984.
[18] Peters JL, Sutton AJ, Jones DR, et al. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008;61:991–6.
[19] Lin L. Graphical augmentations to sample-size-based funnel plot in meta-analysis. Res Synth Methods 2019;In press. DOI: 10.1002/jrsm.1340.
[20] Peters JL, Sutton AJ, Jones DR, et al. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006;295:676–80.
[21] Murad MH, Chu H, Lin L, et al. The effect of publication bias magnitude and direction on the certainty in evidence. BMJ Evid-Based Med 2018;23:84–6.
[22] Terrin N, Schmid CH, Lau J, et al. Adjusting for publication bias in the presence of heterogeneity. Stat Med 2003;22:2113–26.
[23] Peters JL, Sutton AJ, Jones DR, et al. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med 2007;26:4544–62.
[24] Moreno SG, Sutton AJ, Ades AE, et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med Res Methodol 2009;9:2.
[25] Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med 2001;20:641–54.
[26] Andersen BR, Kallehave FL, Andersen HK. Antibiotics versus placebo for prevention of postoperative infection after appendicectomy. Cochrane Datab Syst Rev 2005;3: Art. No.: CD001439.
[27] Hasselblad V. Meta-analysis of multitreatment studies. Med Decis Making 1998;18:37–43.
[28] Higgins JPT. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol 2008;37:1158–60.
[29] Ma X, Lin L, Qu Z, et al. Performance of between-study heterogeneity measures in the Cochrane Library. Epidemiology 2018;29:821–4.
[30] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
[31] Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549.
[32] Sellke T, Bayarri MJ, Berger JO. Calibration of p values for testing precise null hypotheses. Am Stat 2001;55:62–71.
[33] Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Software 2010;36:3.
[34] Schwarzer G. meta: An R package for meta-analysis. R News 2007;7:40–5.
[35] Palmer TM, Sutton AJ, Peters JL, et al. Contour-enhanced funnel plots for meta-analysis. Stata J 2008;8:242–54.
[36] Palmer TM, Sterne JAC. Meta-Analysis in Stata: An Updated Collection from the Stata Journal. 2nd ed.College Station, TX: Stata Press; 2016.
[37] Rendina-Gobioff G, D Kromrey J. PUB_BIAS: a SAS® macro for detecting publication bias in meta-analysis; 2006.
[38] Borenstein M, Hedges LV, Higgins JPT, et al. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons; 2009.
[39] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2008.
[40] Bax L, Yu L-M, Ikeda N, et al. A systematic comparison of software dedicated to meta-analysis of causal studies. BMC Med Res Methodol 2007;7:40.

[41] Temple Lang D. RCurl: General Network (HTTP/FTP/...) Client Interface for R. R package version 1.95-4.82016.

[42] Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010;1:112–25.

[43] Li J, Zhang Q, Zhang M, et al. Intravenous magnesium for acute myocardial infarction. Cochrane Datab Syst Rev 2007;2: Art. No.: CD002755.

[44] Spooner C, Spooner GR, Rowe BH. Mast-cell stabilising agents to prevent exercise-induced bronchoconstriction. Cochrane Database Syst Rev 2003;4: Art. No.: CD002307.

[45] Heiwe S, Jacobson SH. Exercise training for adults with chronic kidney disease. Cochrane Database Syst Rev 2011;10: Art.No.: CD003236.

[46] Khanna P, Suo T, Komossa K, et al. Aripiprazole versus other atypical antipsychotics for schizophrenia. Cochrane Database Syst Rev 2014;1: Art. No.: CD006569.

[47] Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.

[48] Jackson D. The implications of publication bias for meta-analysis' other parameter. Stat Med 2006;25:2911–21.

[49] Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. Canad Med Assoc J 2007;176:1091–6.

[50] Lin L, Chu H, Hodges JS. Alternative measures of between-study heterogeneity in meta-analysis: reducing the impact of outlying studies. Biometrics 2017;73:156–66.

[51] Lin L. Bias caused by sampling error in meta-analysis with small sample sizes. PLoS One 2018;13:e0204056.

[52] Eyding D, Lelgemann M, Grouven U, et al. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. BMJ 2010;341:c4737.

[53] Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. BMJ 2017;358:j3932.