Publication bias in meta-analysis

Shengping Yang, PhD, Gilbert Berdine MD

I am trying to perform a meta-analysis on obesity and the risk of one of the obesity-linked cancers. A meta-analysis can combine data from multiple individual studies in order to increase statistical power, but, if not properly performed, there could be biases in the analysis, which subsequently result in distorted estimations of the statistical effects and the interpretation of those effects. I am wondering how to evaluate and avoid such biases, especially publication bias, in a meta-analysis.

As an important analytical tool, meta-analysis has been widely used in many areas of scientific research. A meta-analysis often includes data from several or more studies on the same or similar research topic. Each included study may represent a small subset of a general population, so that results from a meta-analysis have a broader application to the general population.[1] Therefore, a meta-analysis is a cost-effective approach to address problems that would not be addressed in a single clinical or observational study. On the other hand, data used in a meta-analysis often come from studies that have been completed, and thus researchers who are performing a meta-analysis have no control of the study designs, inclusion/exclusion criteria, data collections and analyses for the individual studies. In reality, a researcher might not be able to access all the studies associated with a research topic, which could very likely introduce biases in a meta-analysis. Due to these limitations, there are several potential pitfalls that researchers should be aware of and avoid in a meta-analysis. We will focus on publication bias in this article.

1. PUBLICATION BIAS

Publication bias, by its name, refers to the failure to publish the results of certain studies based on the direction, nature, or strength of the study findings. Publication bias frequently occurs in academic publications and can be generalized to include outcome-reporting bias, time-lag bias, gray-literature bias, full-publication bias, language bias, citation bias, and media-attention bias.[2] It has been reported that more than 20% of completed studies may not be published for various reasons,[3] including publication bias. For example, studies with a small sample size as well as those with non-significant or negative results are less likely to be published, especially in journals with a high impact.[4,5] Meanwhile, studies with non-significant results tend to be published much later than those with significant results. In addition, studies conducted outside English-speaking countries are less likely to be published in peer-reviewed journals in English. Consequently, the results from published studies may be systematically different from those of unpublished studies, and this translates into challenges for a meta-analysis.

2. ASSESS PUBLICATION BIAS IN A META-ANALYSIS

Ideally, studies included in a meta-analysis represent random samples from a target population. However, due to publication bias, data from small/non-significant studies are less likely to be available/accessible in literature, and a meta-analysis without including those studies might end up with biased findings. To evaluate the risk of such a bias, an assessment of publication bias is often performed.

2.1 FUNNEL PLOT

One of the most widely used methods for assessing publication bias is a funnel plot. A funnel plot is a scatterplot of the treatment effects estimated from individual studies against measurements of study precisions. Because the precision of effect estimate of a study is positively associated with the sample
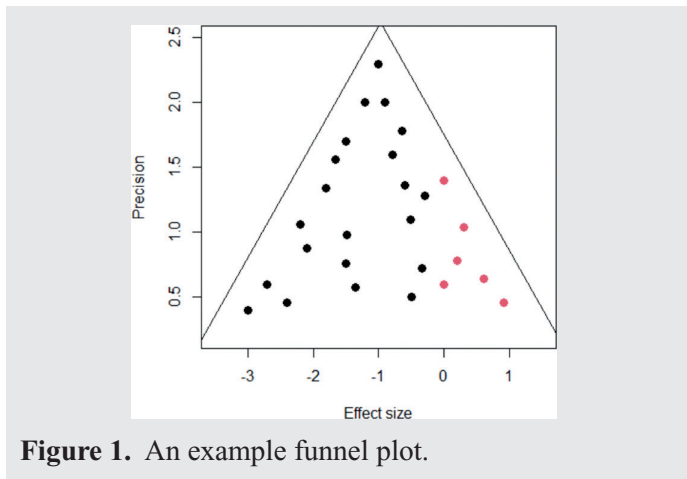
Figure 1. An example funnel plot.

size of that study, larger studies with better precision will be on the top of a funnel plot, and smaller studies will be at the bottom. In addition, because smaller studies often tend to represent more specific and homogeneous subpopulations, the effect estimates from smaller studies often have a wide range and are less accurate in terms of the general population. If all the studies included are random samples from the same population, the plot is expected to resemble a symmetrical, inverted funnel that is narrow on the top and more spread out at the bottom (Figure 1). On the other hand, if there is publication bias, then studies, especially smaller studies, with non-significant results will not be included because of unavailability in the literature, and the funnel plot will have a gap on one side. For example, in the example plot, studies represented by the circles in red (right bottom) will not be included (Figure 1). Note that in the latter situation, the results from a meta-analysis will overestimate the true effect sizes, and more substantial bias will produce more pronounced overestimation.

It is also worth noting that publication bias is not the only cause of funnel plot asymmetry. In fact, it is more appropriate to regard asymmetry in a funnel plot as a measurement of small study effect, i.e., studies with a small sample size are often more likely to have different, often wider, range of effect sizes, compared to studies with a larger sample size.

Although a funnel plot is a valuable method for evaluating potential publication bias, studies have

shown that many researchers might not be able to visually identify publication bias using such a plot,[6] and the same plot can be interpreted differently by different researchers.

2.2 Tests for assessing funnel plot asymmetry

The Begg's and Egger's tests are the two widely used tests for assessing funnel plot asymmetry.

2.2.1 The Begg's rank test

The test proposed by Begg and Mazumbar[7] was developed based on the Spearman correlation between adjusted effect sizes and their variances. The deviation of the correlation from zero is an indication of the funnel plot asymmetry. In other words, in the presence of publication bias, a positive correlation between the effect size and variance of the estimate is expected because both the effect size and variance are larger for smaller studies.

2.2.2 The Egger's test

Egger et al[8] proposed to evaluate the degree of funnel plot asymmetry by examining the intercept from the regression of standard normal deviate against precision. Specifically, the standard normal deviate (SND), defined as the odds ratio divided by its standard error, is regressed against the estimate's precision: $SND = a + b \times$ precision. For smaller studies, both the precisions and $SNDs$ are small due to larger standard errors; while for larger studies, the precisions are large, and if the treatment effects are large, then the $SNDs$ are also large. Therefore, for studies that represent randomly selected samples from a population, the regression line will scatter about a line that runs through the origin, with the slope b indicating the size and direction of effect. Otherwise, if there is asymmetry, then the regression line will be away from the origin.

In general, the Egger's test has better power than the Begg's test, although the power for both tests is considered low, especially when the effect sizes are heterogeneous among individual studies. In addition, because both tests were developed based on the

differences between large and small studies, neither test would perform well when the included studies are either all large or small.

3. AVOID PUBLICATION BIAS IN A META-ANALYSIS

Although publication bias cannot be completely avoided, attempts have been made to minimize the risk of publication bias.

3.1 PROSPECTIVE REGISTRATION

Many efforts have been made to promote prospective registration for clinical studies. For example, the International Committee of Medical Journal Editors mandated that beginning July 2005, all clinical trials be registered at or before the enrollment of the first participant as a condition of consideration for publication,[9] to promote transparency and research integrity, as well as to prevent selective reporting and publication bias in clinical studies.

3.2 SEARCH FOR UNPUBLISHED RESULTS

In general, to avoid publication bias, a thorough literature search is crucial. Besides published results, unpublished results can be identified by exploring sources, such as meeting abstracts, PhD dissertations, supplementary materials of a published article, as well as by contacting authors and companies/organizations directly involved in a study. In addition, a literature search should not be restricted to studies according to language of publication to avoid possible language bias. Note that the inclusion of unpublished results should be performed with care. Very often, small and unpublished studies are more likely to have poor study design and insufficient analytic and scientific rigor. Therefore, unpublished results should be thoroughly examined before being included to avoid introducing bias caused by poor study quality.

3.3 IMPROVE PUBLICATION GUIDELINES

Very often, the primary considerations for accepting an article for publication are the novelty or importance of the research, as well as the significances of the findings. To reduce publication bias, some journals, especially open-access journals, started using scientific and technical quality as one of the acceptance criteria rather than significance of findings,[10,11] and this improvement in publication guidelines makes it easier for authors to submit more non-significant or negative results for publications.

In summary, the problem of publication bias is not trivial in a meta-analysis. In the presence of publication bias, a meta-analysis might report distorted, often over-estimated results. The assessment of publication bias can start with a visual examination of a funnel plot, followed by a formal test of asymmetry. However, when there is evidence of asymmetry, publication bias might not be the only explanation. A thorough literature search on both published and unpublished results may partially mitigate the risk of publication bias. Endeavors should be made to maximize the inclusion of unpublished results, which are often generated from smaller studies, while attention should be paid to ensure these studies are methodologically sound. The majority of journals prioritize their acceptance of articles that are novel and that have significant findings, and there is a trend that more journals are focusing on scientific and technical quality for acceptance, especially among open-access journals. A holistic approach is necessary to address publication bias.

REFERENCES

1. Danos DM. Toward a transparent meta-analysis. The Southwest Respiratory and Critical Care Chronicles 2020;8(33): 60–62.
2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: An updated review of related biases. Health Technology Assessment 2010;14(8): iii–193.
3. Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. Circ Cardiovasc Qual Outcomes 2012;5:238–40.
4. Dickersin K, Min YI. Publication bias: the problem that won't go away. Ann N Y Acad Sci 1993;703:135–46.
5. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. Lancet 1991;337:867–72.
6. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clinical Epidemiology 2005;58(9): 894–901.
7. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50(4): 1088–101.
8. Egger M, Smith GD, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. British Medical Journal 1997;315(7109):629–634.
9. van Enst WA, Ochodo E, Scholten RJ, et al. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. BMC Med Res Methodol 2014;23;14:70.
10. Krumholz HM, Ross JS, Gross CP, et al. A historic moment for open science: the Yale University Open Data Access project and Medtronic. Ann Intern Med 2013;158:910–1.
11. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. JAMA 2013;309:1355–6.