A Comparison of Methods to Detect Publication Bias for Meta-analysis of Continuous Data

Nik Ruzni Nik Idris
Department of Computational and Theoretical Sciences, Kulliyyah of Science,
International Islamic University Malaysia, Kuantan 25200, Pahang, Malaysia

Abstract: Publication bias is a serious problem in meta-analysis. Various methods have been developed to detect the presence of publication bias in meta-analysis. These methods have been assessed and compared in many dichotomous studies utilizing the log-odds ratio as the measure of effect. This study evaluates and compares the performance of three popular methods, namely the Egger’s linear regression method, the Begg and Mazumdar’s rank correlation method and the Duval and Tweedie’s trim and fill method, on meta-analysis of continuous data. The data comprised simulated meta-analyses with different levels of primary studies in the absence and presence of induced publication bias. The performances of these methods were measured through the power and type I error rate for the tests. The results suggest the trim and fill method to be superior in terms of its ability to detect publication bias when it exists, even in presence of only 5% unpublished studies. However, this method is not recommended for large meta-analysis as it produces high rate of false-positive results. Both linear regression and rank correlation method performed relatively well in moderate bias but should be avoided in small meta-analysis as their power is very low in this data.

Key words: Publication bias, trim and fill method, linear regression method, rank correlation method, power of test, type 1 error

INTRODUCTION

One of the common concerns in meta-analysis pertains to the issue of publication bias. A bias may be present in estimation of parameters in almost every research design (Yin et al., 2011; Al-Khasawneh, 2010; Zhang et al., 2009). In meta-analysis, a publication bias may be encountered if it is based on integration of results obtained from studies which have been published. The bias occur when published studies are associated with their outcomes, namely, studies which produced large effects or significant results are more likely to get published. If a publication bias operates in the same direction for all studies, biased estimates which seemingly precise and accurate will be produced. The conclusions based on these results will appear convincing although they could be seriously misleading. (Begg and Berlin, 1988).

The simplest method to detect publication bias is through subjective inspection of funnel plot (Light et al., 1994). A funnel plot provides graphical illustration in a form of a scatter plot of the estimate of the treatment effects from each study against the measure of its precision, which is typically the inverse of the corresponding standard error. When estimating the effect size, the method assumes smaller less precise studies are subject to more random variation than larger studies. In the absence of underlying bias, the plot will look like an inverted funnel. In presence of publication bias, some smaller studies with smaller effect will be missing resulting in an asymmetrical funnel plots. Visual assessment of the graphical plot is however very subjective. Statistical tests based on funnel plots such as the Egger’s linear regression test (Egger et al., 1997), Begg and Mazumdar’s Rank Correlation method (Begg and Mazumdar, 1994), Duval and Tweedie’s Trim and Fill method (Duval and Tweedie, 2000) were developed to provide more objective examination of publication bias. Despite advancement of a more complex methods (Dear and Begg, 1992; Givens et al., 1997), simpler and more practical methods remain widely used and have been implemented in a number of software packages. These methods have been assessed and compared in many studies (Begg and Mazumdar, 1994; Sterne et al., 2000; Sutton et al., 2000; Macaskill et al., 2001). In their study, Sterne et al. (2000) compared the performance of Egger’s linear regression method and Begg and Mazumdar’s rank correlation. The study uses odds ratio as the outcome measure, where only dichotomous meta-analyses with at least five primary studies were considered. The study concluded that the regression test has higher power compared to the rank correlation test, however, it produced very high false-positive rates in cases where there is large treatment effects, or when there are few events per trial. An empirical study by Sutton et al. (2000) examined the impact of publication bias on collection of meta-analyses.
using the trim and fill method. Similarly, the study uses
odds ratio but included meta-analyses with minimum
number of ten primary studies only. In this study, they
found the effect on overall inferences were small, i.e., after
adjusting for publication bias, the conclusions was
change in only about 10% of the studies. A more recent
study (Peters et al., 2006) found that if odds ratios is
used, funnel plot asymmetry could be attributed by the
correlation between the natural log of odds ratios and it
standard error and this has the effect of inflating the type
I error for test.

The above studies are based on dichotomous data
which utilizes the odds ratio as the measure of effect.
Some of the problem encountered may not be of much
concern in continuous data. This paper evaluates and
compares the performance of three methods commonly
used to detect publication bias in meta-analyses of
continuous data. Specifically, it compares the performance
of the Egger’s linear regression method, the Egger and
Mazumdar’s Rank correlation method and the Duval and
Tweedie’s Trim and Fill method. The data comprised
simulated meta-analysis with different levels of primary
studies in absence and in presence of induced publication
bias. Performance of the three methods is measured
through the power and the estimated type I error
generated from the tests.

MATERIALS AND METHODS

Meta-analysis data used in this study were simulated
using R. Simulation study is often used in many research
design in diverse area of research as it allows the control
and manipulation of research design factors
and the incorporation sampling error into the
analyses (Boroni and Clause, 2011; Islam and Zhou,
2009; El-Messousi et al., 2007; Prakash and Karunanithi,
2009). The treatment effects and their corresponding
standard errors were simulated so that no publication bias
was present. This is achieved by combining, for each
meta-analysis, three randomly generated data sets; small
studies with small-size effects, large studies with a
medium-size effect and small studies with large-size
effects. The characteristics and assigned values used in
this simulation study, namely the size of treatment effects,
their variances and the number of primary studies, were
based, in part, on the meta-analyses of the effects of
TUNA treatments on the patients with benign prostatic
hyperplasia. This meta-analysis consists of treatment
effects and the corresponding variances ranges from 3.0
to 7.0 and 0.5 to 3.0, respectively.

Publication bias of varying degree was then induced
on the simulated meta-analysis based on the assumption
that the statistical significance of a study is predictive of
publication status (Dickersin, 2005). Large studies are
likely to achieve statistical significance, even if their
effects are relatively small, small studies, on the other
hand, will reach statistical significance only if they yield
large effects. Thus studies with least significant effect size
are more likely to be subjected to publication bias.

Based largely on what was done in work discussed
earlier, three levels of primary studies, N, were used, small
(N = 10), medium (N = 30) and large (N = 50, 100) and three
levels of percentage of missing publications, x%, were
induced, namely high where more than 50% of the studies
were excluded and medium and low which corresponds to
30% and 5-10% of the excluded publications, respectively.
The power to detect the publication bias for all three
methods were estimated based on the proportion of
p-value <0.05 in presence of induced publication bias.
Type I error was estimated on the basis of the proportion
of p-value <0.05 when the meta analysis were initially
generated with no publication bias. An ideal test would be
those that produced type I error approximately close to
the nominal value and a good power to detect publication
bias when it exists.

Each of the sixteen meta-analyses with different
combinations of N and x% was replicated 10,000 times.
The maximum standard error of estimates for the type I
error and power in the simulations were 1.5 and 1.8%,
respectively.

RESULTS

For all three methods, the power increases with
increasing number of primary studies in a meta-analysis,
N. The trim and fill method demonstrated superior power
among the three methods. The power of test for this
method was reasonably good even in small meta-analysis
(N = 10) with very little bias presence (x = 5%). Both the
linear regression and rank correlation methods could not
detect the bias at this level and their power remained
mostly below 5%. Similar trend was observed when the
percentage of missing publications is at 10% (Fig. 1).

In presence of moderate to severe publication bias
(x>30%), the trim and fill method performed very well, with
power remained above 80% across all N. The linear
regression method performed relatively well in this
scenario, attaining power at above 80%, when the number
of primary studies were larger than 30 (Fig. 1). The rank
correlation did not perform as well even in severe bias
(x>50%), unless the number of primary studies in meta-
analysis are very large (N>80).

In terms of type I error however, the trim and fill
method performed poorly, consistently exceeding the
appropriate rate of 5% across the different degree of
publication bias and for all level of N (Fig. 2). The rate of
Fig. 1(a-d): Power of trim and fill method, linear regression method and rank correlation method to detect publication bias missing studies (a) x% = 5, (b) x% = 10, (c) x% = 30 and (d) x% = 50

Fig. 2(a-d): Probability of Type 1 error for trim and fill method, linear regression method and rank correlation method. Legend: Missing studies (a) x% = 5, (b) x% = 10, (c) x% = 30 and (d) x% = 50

incorrectly rejecting the null hypothesis ranges from 9% to 30%, increasing with the number of primary studies. On the other hand, in linear regression and rank correlation methods, the Type 1 error rate is too low. For rank
correlation, the type I error rate ranges between 0.05 to 1% across the N for all level of publication bias and between 0.01 to 0.04% for the linear regression method.

DISCUSSION

This study examines the performance of three popular methods to detect publication bias in meta-analysis of continuous data, namely, Egger’s linear regression, Begg and Mazumdar’s rank correlation and the Duval and Tweedie’s trim and fill method. The performance of these methods was assessed through the power and the type I error generated from the tests.

Some of the results from this study support the findings from earlier studies which are based on dichotomous data. As expected, the power of the tests increased with increasing numbers of primary studies included in a meta-analysis and increasing percentage of unpublished studies.

The results suggest the trim and fill method to be superior in terms of its ability to detect publication bias when it exists. The method is able to detect the presence of very low bias, i.e., when only 5% of the publications are not reported. However, the trim and fill method produced type I error that is consistently exceeding the appropriate level of 5%, particularly in moderate to large size meta-analysis. This may raise concern over its tendency in giving off false-positive results.

This finding is consistent with earlier study (Terrin et al., 2005) on dichotomous data which suggested that if the trim and fill method is used on heterogeneous sets of binary data, it may produce false-positive conclusions. One advantage of this limitation is that a non-significant test from trim and fill method clearly indicates the presence publication bias has been ruled out.

In terms of power, the linear regression method is found to be more powerful than the rank correlation method. This is particularly significant in the presence of moderate publication bias. The differences reduce with increasing number of primary studies. The power of both methods was very poor in presence of low bias and in small meta-analysis. The power of the linear regression test in the small meta-analysis was consistently below 60% and those of rank correlation stayed below 30%.

It has been claimed that in dichotomous data, the regression method may give false-positive results in the absence of bias (Sutton et al., 2000). Additionally, findings of earlier study (Peters et al., 2006) suggested serious concerns over the practical use of Egger’s linear regression test on dichotomous data when the log-odds ratio is used as the measure of effect as it produces inflated type I error. In contrast, for continuous data meta-analysis, both this method and the rank correlation method produced false-positive rates which were incorrectly low in the absence of underlying bias. In fact non-significant test using either one of these method should be interpreted with cautious.

Nonetheless, the power to detect asymmetry for linear regression method is reasonably good when the bias is present at moderate level. The test attained approximately 80% power when the number of studies is in the neighborhood of 20. The rank correlation method requires more than 30 primary studies to achieve the same level power within the same degree of bias.

A publication bias is in fact another form of missing data. In any data analysis, if missing data is not properly handled, it could be one of the main sources of bias in parameter estimates. There are variety of studies on missing data from diverse research area (Nath and Bhattacharjee, 2012; Uysal, 2007; Salha and Ahmed, 2009). In this study, two factors were examined, namely the number of primary studies included in meta analysis, N and the degree of publication bias present, x %. It was suggested (Terrin et al., 2005; Peters et al., 2006) that heterogeneity level of the dichotomous data has some influence on the power and the type I error produced by the tests. Further work may be extended to include this factor in meta-analysis of continuous data.

Obviously there is no best method for detecting the publication bias that can be use for all type of data. It is recommended that the number of studies be taken into consideration when choosing methods to detect the bias. In continuous data, a significant trim and fill test must be interpreted with caution due to high likelihood of false positive conclusion. At the same time if the linear regression or the rank correlation method is used, the non-significant test should not be taken as confirmation that bias is absent. In fact, the linear regression and rank correlation method should be avoided for small size meta analysis as they would not be able to detect publication bias when the percentage of missing publication is less than 20%. On the other hand, the trim and fill method should be avoided in large Meta analysis due to its tendency to conclude publication bias even when it does not exist.

REFERENCES


