

Effect Comparison between Egger's Test and Begg's Test in Publication Bias Diagnosis in Meta-Analyses: Evidence from a Pilot Survey

Xiuquan Shi^{1,*}, Chan Nie¹, Shangpeng Shi¹, Tao Wang¹, Huajun Yang¹, Yanna Zhou¹,
Xiaoqin Song²

¹Department of Epidemiology and Health Statistics, School of Public Health, Zunyi Medical University, Zunyi, PR China

²The First Affiliated Hospital of Zhengzhou University, Zhengzhou, PR China

Abstract: *To compare the effect and assess the consistence between Egger's test (ET) and Begg's test (BT), so as to choose the suitable method in publication bias (PB) diagnosis. We extracted PB-related information from 117 published meta-analysis papers through a thorough search in Medline database, and then compared the difference between ET and BT (while funnel plot [the most commonly used method to detect PB] as an important reference to judge the possible truth of PB) to detect the bias of the same data. Moreover, the correlation, agreement ratio (AR) and Kappa between ET and BT were also analyzed. We found that the rate of P-value less than 0.05 was 14.29% and 7.14% in ET and BT, respectively. Moreover, ET was commonly to get less P-value than BT ($t=-2.22$, $P=0.03$) to detect the bias of the same data. Though the results of ET and BT in most meta-analysis papers had a significant correlation ($r_s=0.68$, $P<0.001$), AR and Kappa were relatively low (AR=91.04%, Kappa=0.457, $P=0.001$). Moreover, ET and BT were not consistent to the PB diagnosis by the forest plot (AR=87.50% and 91.43%, respectively), and sometimes ET and BT were not consistent each other. From the results we conclude that ET is more sensitive in PB diagnosis than BT, but it is still not accurate. Thus, we recommend an associated application to improve the reliability of PB diagnosis.*

Keywords: *Publication bias, Bias diagnosis, Egger's test, Begg's test, Meta-analysis*

1. INTRODUCTION

Publication bias (PB) is an inevitable problem in the systematic review and meta-analysis; it is well-known as one most important type of biases termed reporting bias, so it is also one of the main threats to the validity of meta-analysis [1]. It mainly refers those studies with significant, positive results are easier to be published than those with non-significant or 'negative' results [2]. It usually has three main sources, which are authors, editors (reviewers) and grant supporters. The subsequent over-representation of positive studies in systematic reviews may lead the reviews biased towards a positive result.

It is fair to assume that most systematic reviews will be subject to reporting bias with some extent if we have evidence that the biases exist. It will cause misleading or even come up with a wrong conclusion if the studies we have failed to be published were systematically different from those have been published. Biases in meta-analysis are widely existed in medical research [3], so more attention should be paid on them.

Even so, van Enst et al. had found that most systematic reviews of diagnostic test mentioned the issue, but only with 41% measuring PB [4]. Ioannidis and Trikalinos explored PB problem in 1,669 Cochrane reviews, ultimately analyzed 6,873 meta-analyses, and they found a half of the meta-analyses had no statistically significant results. Moreover, they concluded that only 5% of the full set of Cochrane reviews would qualify for the use of asymmetry tests, and only 12% of those with a larger number of events and participants. Unfortunately, they found very little (only around 3-4%) concordance between different asymmetry tests [5]. It is believed that PB is underreported in systematic reviews published in high-impact-factor journals [6]. In some low-impact-factor journals and some non-peer reviewed journal, systematic reviews and meta-analysis studies without PB

assessments are even more serious. Moreover, the misapplication and misinterpretation of statistical tests are also an important problem. For example, sometimes tests for PB when there is clinical heterogeneity will get a false-positive result.

Although several statistical methods have been developed to detect and adjust for the PB since the beginning of 1980s, some of them are not well known and are not being used properly in both the statistical and clinical literature [1]. E.g. some researchers conclude that statistical tools for PB are under-utilized, but the extent to which PB is assessed was not part of their study. Another opinion is that PB itself may be decreasing over time, even if the method for exploring it is still problematic [7].

Although Sutton et al. reported some models to assess the potential effect of PB within a meta-analysis context [8], they mainly analyzed the published meta-analysis papers by trim and fill method. Some related papers were also mentioned some methods to detect the PB elsewhere, such as p-curve and p-uniform approaches to view as alternative implementations of the original Hedges selection method [9], but they were discussed only a part of methods, especially either focused on the most common used methods (such as Egger's test (ET) and funnel plot), nor they mentioned how to judge whether the PB was existed if the final results of different methods were not consistent.

Among these PB diagnosis methods, as the top two popular test methods, the difference between ET and Begg's test (BT) is especially been paid attention. Shi et al had compared the power difference and analyzed the reasons between ET and BT [10].

From many previous published papers, we found that the diagnosis results were different between ET and BT [11-14]. Most of them are that the P-value of Egger's test (PET) is less than the P-value of Egger's test (PBT) [11, 13-14], but sometimes the PBT is greater than the PET instead [12]. Is it just adventitious or inevitable? How high is the concordance rate between BT and ET? These questions remain unclear.

In this paper, we aimed to test the hypothesis that BT was as good as ET, and they were high consistent in PB diagnosis. We compared ET and BT (while funnel plot [the most commonly used method to detect PB] as an important reference to judge the possible truth of PB) to detect the bias of the same data from recent published meta-analysis papers; and we reported the difference of the agreement ratio (AR), and the kappa value between the results of ET and BT to assess their consistence.

2. MATERIALS AND METHODS

2.1. Data Materials

The source of data was from a systematic search. We carried out a thorough search in the Medline through PubMed search gate. The search strategy was as following: ((publication bias [Abstract]) AND meta-analysis [Abstract]) AND (("Egger's test" AND "Begg's test" AND "funnel plot") [All fields]), with "Full Fee Text", the update deadline was Dec.1, 2015. All papers with full text were downloaded, after discarding the duplication; we got 113 full pdfs texts. As several papers had two or more independent subgroups, we could extract the data separately, so the final sample size was 117.

From the selected reports, two reviewers were reviewed carefully then extracted the data and other essential information including the P-values of ET and BT, funnel plot, other results of PB detection, the final conclusion of whether the bias was existed by original authors; the P-values of heterogeneity, I-square, model selection; some general characteristic such as first author, published year, data source (which region/country; the journal's name, the research topics), included numbers of original studies in a meta-analysis paper. Both reviewers were using a standardized form and extracted all data independently. If they had different extracted data of a same paper, they should discuss and consult for the 3rd expert to reach the agreement.

2.2. Statistical Methods for Bias Diagnosis

2.2.1. ET

ET is also called linear regression method, which was suggested by Egger et al. ET was used to test the funnel-plot symmetry, in which a regression model is built, using the standardized estimate of the size effect as a dependent variable and the inverse of the standard error (1/SE) as an independent variable. If the intercept is significantly different from zero, the estimate of the effect is considered biased [15].

2.2.2. *BT*

BT is also named as rank correlation test which was firstly introduced by Begg et al. This method examines the relationship between the standardized treatment effect and the variance of the treatment effect using Kendall's tau [16].

2.2.3. *Other Methods*

Other methods such as Funnel plot (the most common method to detect and assess PB, it was as an important reference of bias diagnosis when the final conclusion was not clear or the results were inconsistent between ET and BT in our study). Cumulative meta-analysis by precision [17], capture-recapture method [18], and Bayesian approach [19]; sometimes even sensitive analysis [20] also can be used to reflect and evaluate the PB. As these methods only partly used and reported in some original meta-analysis papers, so we were only used them as assistant PB diagnosis methods to judge the final conclusion of bias in our study.

2.3. **Difference, Correlation Analysis between ET and BT**

The difference, correlation analysis between ET and BT and other related factors were also analyzed. We also analyzed the difference, correlation analysis between the number of left dot (some funnel plot also put the left dot in the bottom of the figure) and the number of right dot (some funnel plot also put the right dot in the upper of the figure) in the funnel plots in the same time to assistant to judge the PB conclusion in these meta-analysis papers.

2.4. **AR and Kappa Values**

In order to assess the consistence between ET and BT in PB diagnosis, the AR and Kappa values were calculated among the results of ET, the results of BT, and the final conclusion of whether the bias was existed by original authors' opinions.

2.5. **Statistical Software**

All of the statistical analyses were performed with SPSS (version 18.0; SPSS Corporation, Chicago, IL, USA). All the tests were two-sided. $P < 0.05$ was considered as statistically significant.

3. **RESULTS**

3.1. **General Characteristics**

Through the search, we found that the meta-analysis papers were mostly came from Asia, especially from China (98 papers in 113; nearly 90%); besides China, in other Asian countries, there were 3 papers from Iran, 3 papers from Japan, 2 papers from India, and 1 paper from South Korea. However, as the mainly developed countries, there were little meta-analysis papers which entered our database in Europe and the United States, only 2 papers from the United States, 1 paper from the UK, Netherlands, and Greece, respectively. Only 1 paper was from Africa [21].

About the topic of meta-analysis papers, we found that cancers accounted for about one half, followed by the heart & vascular diseases (also accounted for nearly 1/4), and the rest topic was relatively dispersed.

Besides ET and BT, some meta-analysis papers (nearly 1/3) we searched and downloaded had done a subgroup analysis; while part of meta-analysis papers did a sensitivity analysis. Besides these publication diagnosis method, there were a few papers did the meta-regression, very small amount (only 5 article conducted the trim and fill method and calculated the fail-safe number).

In the 117 meta-analysis samples' data, PET ranged from 0.01 to 0.99 with a median was 0.29; while PBT ranged from 0.01 to 1.00, with a median was 0.35. The number of included original studies ranged from 3 to 85, with a median was 12. All details were listed in Table 1.

Table1. Descriptive index, correlations and regression analyses between PET and PBT; funnel plots; and I-square

Index	N ^Δ	Mean ± SD	95% CI of mean	Median (Min-Max)	Inter-quartile range	P-value of Normality tests	Spearman correlations
Number of included studies	117	15.52±12.78	12.69-18.34	12(3-85)	8	<0.001	—
I-square	109	49.75± 30.72	1.95-57.55	51.80(0-98)	48.52	0.083	—
PET	112	0.38±0.33	0.30-0.47	0.29(0.01- 0.99)	0.59	<0.001	r _s =0.680, P<0.001.
PBT	70	0.44±0.33*	0.35-0.52	0.35(0.01-1)	0.61	0.001	
Number of left dot	81	6.321±4.38	5.35-7.29	6(1-31)	3	<0.001	r _s =0.557, P<0.001
Number of right dot	81	6.272±3.72	5.45-7.10	5(2-21)	3.80	<0.001	

^Δ Number less than 117 due to missing (some papers only offer the result either Egger's test or Begg's test or funnel plots);

*vs. "PET", $t=-2.22$, $P=0.030$ based on paired t test for their difference is normal distribution

We also extracted the information of heterogeneity tests, and we found that 64 papers had reported a less than 0.05 P-value, accounted for 59.81% from 107 papers, except 10 papers were missing (which means it had heterogeneity in that meta-analysis paper). Moreover, there was 53.21% (58/109) of heterogeneity I^2 were greater than 50% (which means it had heterogeneity). These results suggested that the heterogeneity cannot be neglected in half of the sampled meta-analysis papers.

3.2. Publication Bias Results

3.2.1. Bias Detecting by ET

From 117 meta-analysis papers, there were 5 articles did not show the PET, and the P-value were less than 0.05 in 16 papers (that means it had PB in that meta-analysis paper), which accounted for 14.29% in the 112 articles remain.

3.2.2. Bias Detecting by BT

From 117 meta-analysis papers, there were 47 articles did not use BT or did not offer the PBT, and the P-value were less than 0.05 in 5 papers (which suggested it had PB in that meta-analysis paper). That was to say, 7.14% papers were biased by the BT diagnosis.

3.2.3. Other Results and the Final Conclusion of Bias by the Authors of the Original Papers

Except funnel plot analysis of PB, only a small part of papers mentioned the results of using other PB detection methods such as trim and filling method, the fail-safe numbers, so we did not list their results. In summary, we found there were 7 papers with bias conclusion by the authors of original papers. The bias rate was only 5.98%, while the rest of the 110 articles considered as without bias in authors' opinion, though part of the shape of the funnel plots looked not having a good asymmetry.

3.3. Difference, Correlation between ET and BT

We found the difference between ET and BT was also significant. Compared to "PET", "PBT" was greater ($t=-2.22$, $P=0.030$) based on paired t test for their difference was normal distribution.

The correlation analysis showed that the Spearman correlation coefficient (r_s) between PET and PBT was 0.680 ($P<0.001$), and the Spearman correlations coefficient between the number of left dot and the number of right dot was $r_s=0.557$, $P<0.001$ (Table 1).

Moreover, we also did a funnel plot analysis to assistant to detect the PB besides ET and BT. The funnel plot analysis showed that the number of left dot ranged from 1 to 31, with a median was 6; the number of right dot ranged from 2-21, with a median was 5. We also found the number of right dot

Effect Comparison between Egger's Test and Begg's Test in Publication Bias Diagnosis in Meta-Analyses: Evidence from a Pilot Survey

was a little less than the number of left dot based on Wilcoxon signed ranks test, but it was not significant ($Z=0.28, P=0.777$) (Table 1).

3.4. AR and Kappa Values

Firstly, the AR (%), Kappa value was 87.50% and 0.310 ($P<0.001$) between PET and the final conclusion in original meta-analysis papers. Secondly, the AR (%), Kappa value were 91.43%, 0.354, and $P=0.003$ between PBT and the final conclusion in original meta-analysis papers. At last, the AR (%), Kappa value and P-value were 91.04%, 0.457, $P=0.001$ between PET and PBT, respectively (Table 2 and Table 3).

Table2. The association, AR, kappa and difference between Egger's test, Begg's test and the final bias conclusion

Publication bias diagnosis		Final bias conclusion*		Total	Difference*	Kappa	AR [§] (%)	Association [#]
		No bias	Bias					
PET	$P>0.05$	94(a)	2(b)	96	0.013	0.310, $P<0.001$	87.50	$P=0.004$
	$P\leq 0.05$	12(c)	4(d)	16				
	Total	106	6	112 ^Δ				
PBT	$P>0.05$	62(a)	3(b)	65	1.000	0.354, $P=0.003$	91.43	$P=0.038$
	$P\leq 0.05$	3(c)	2(d)	5				
	Total	65	5	70 ^Δ				

*Source from authors' opinion of the original meta-analyses papers (mainly judged by the forest plot);

^Δ Number less than 117 due to missing (some papers only used one of Egger's test, Begg's test and funnel plots);

*Difference between two methods was based on McNemar test;

[§] $AR = (a + d)/(a + b + c + d)$; [#] Association between two methods was based on Fisher's exact test.

Table3. The association, AR, kappa and difference between the results of Egger's test and Begg's test

Publication bias diagnosis		PBT		Total	Difference*	Kappa	AR [§] (%)	Association [#]
		$P>0.05$	$P\leq 0.05$					
PET	$P>0.05$	58(a)	1(b)	59	$P=0.219$	0.457, $P=0.001$	91.04	$P=0.004$
	$P\leq 0.05$	5(c)	3(d)	8				
Total		63	4	67 ^Δ				

^Δ Number less than 117 due to missing (some papers only used one of Egger's test, Begg's test and funnel plots);

*Difference between two methods was based on McNemar test;

[§] $AR = (a + d)/(a + b + c + d)$; [#] Association between two methods was based on Fisher's exact test.

4. DISCUSSIONS

It is clear that the potential presence of PB is often ignored by researchers performing meta-analyses [6, 21]. Sometimes though the methods of PB diagnosis were used, they were only based on the statistical significance. As is well known, whether the PB existed is the key aspect of the meta-analysis results' reliability. In order to lower the effect of the PB, precisely assessing PB is necessary. Assessment of PB in systematic reviews and meta-analyses in many fields of medical science are believed to require improvement [22]. Though there are several PB methods such as funnel plot, ET, BT, trim and filling method, fail-safe number, and it was still lack of a standard procedure to assess PB. General speaking, among these methods, ET and BT are mostly used methods.

In our study, we found that for the same data, the PET is significantly less than the PBT. Moreover, 14.29% published meta-analysis papers have a less than 0.05 P-value by ET, while 7.14% papers have a less than 0.05 P-value by ET, which suggested PB. These results suggest that the detective power of ET is possible a little greater than BT, and the findings are also consistent to another research by Shi et al, [10].

Though the ARs are relatively high (nearly 90%), and the Kappa value between ET and BT is also statistically significant. We find the difference between ET and BT is objective existence, especially sometimes the PBT is greater than the PET instead [12], even though ET seems more sensitive to biases than BT, but their diagnosis effect may affect by many factors, such as heterogeneity, I-square

(E.g. In our study, there was 59.81% meta-analysis paper had a less than 0.05 P-value in the heterogeneity test, and 53.21% of heterogeneity I^2 were greater than 50% (which suggested significant heterogeneity)), literatures' quality, the included numbers of original studies, the normality of the extracted data. Just as Shi et al had reported the power difference between these two tests was affected by the results of the normality tests [10].

We also found that only in 5.98% meta-analysis papers, authors declared the PB were existed. These results indicate that using different methods may not lead to an inconsistent result. However, from the results, it is obviously suggested that the PB problem is underestimated. In fact, a few funnel plots in the published meta-analysis papers look a little asymmetrical in some extent, though the asymmetrical degree is not serious.

For the ET, nowadays, it is widely used to an assessment of PB. But estimating the standard error of independent variable in Egger's linear regression only comes from included studies, which will affirmatively produce sampling errors. Accordingly, the estimations of the intercept and slope were biased in the regression model. BT is considered as a test of the direct analogue model of funnel plots. General speaking, both ET and BT are considered to have stronger statistical and discriminatory powers than other methods in detecting PB with the same type I error level [23], as they both can fully use the information rather than qualitative descriptions and graphs. From our results, we also supported that BT was consistent to ET in PB diagnosis, but unfortunately we drew a low kappa value (only 0.457).

However, we should know that both ET and BT have some limitations. Such as (i) ET could not explain the reason of the asymmetry of the funnel plot. In addition, the high type I error rates (the proportion of false-positive results) of ET should be treated with more attention [24]; (ii) BT has an obvious variation of its power. The major affecting factors of the power are the number of included studies, effect size and its variance. One study of points out one of the limitations of a large part of PB methods based on the asymmetry of the funnel plot, namely that they do not take between-study heterogeneity into account [5]. One difficulty with the use of them is that they require researchers to make certain assumptions about the nature of PB. The performance of a certain PB method depends on whether or not the method's assumptions are met [25]. Through comparing their advantages and defects of ET and BT, we aimed to offer an approach to select the appropriate PB detect methods in the future meta-analyses.

5. CONCLUSION

Though ET is sensitive (high statistical power) than BT, other methods such as funnel plot, trim and filling method, fail-safe number were still recommended to a combined use (as assistant detective methods to the final conclusion of whether the PB was existed), after synthetically balancing all the results, the correctness and the robustness will significantly enhance to detect PB, so as to increase the reliability and validity of evidences came from the meta-analysis papers.

Of course, the best solution to the problem of PB is to complete reporting of study results, moreover, the key way is to establish a pre-established system of trials registers. In addition, logistic regression analysis can be used to control the confounders, other statistical methods, such as Bayesian hierarchical models and the multilevel models; cumulative meta-analysis should also be considered in future analyses, to lower the effects of other biases.

ACKNOWLEDGEMENTS

The authors wish to thank Zhishun Lu (who had previously worked in Department of Pathology, Stanford University, USA, now she works in People's hospital of Guizhou Province, P.R. China) for her critical review and scientific editing of this manuscript, and this study was supported by a supported grant of "Discipline construction funds of doctoral degree" of Zunyi Medical University (No. [2015]-996036).

REFERENCES

- [1] Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. *Stat Med* 2015; 34: 343-360.
- [2] Sedgwick P. What is publication bias in a meta-analysis? *BMJ* 2015; 351:h4419.
- [3] Mavridis D, Efthimiou O, Leucht S, Salanti G. Publication bias and small-study effects magnified effectiveness of antipsychotics but their relative ranking remained invariant. *J Clin Epidemiol* 2016; 69:161-169.

- [4] van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol* 2014; 14:70.
- [5] Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ* 2007; 176: 1091-1096.
- [6] Onishi A, Furukawa TA. Publication bias is underreported in systematic reviews published in high-impact-factor journals: metaepidemiologic study. *J Clin Epidemiol* 2014; 6: 1320-1326.
- [7] Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Stat Med* 2015; 34: 2781-2793.
- [8] Sutton AJ, Song F, Gilbody SM, Abrams KR. Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res* 2000; 9: 421-445.
- [9] McShane BB, Böckenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. *Perspect. Psychol Sci* 2016; 11(5):730-749.
- [10] Shi XQ, Wang ZZ. Comparison of the power difference of Egger's test and Begg's test and the reason analysis. *Acta Med Univ Sci Technol Huazhong* 2009; 38: 91-93,102. [In Chinese, English abstract]
- [11] Bi GL, Chen FL, Huang WM. The association between hypertensive disorders in pregnancy and bronchopulmonary dysplasia: a systematic review. *World J Pediat* 2013; 9: 300-306.
- [12] Hosseini SM, Poorolajal J, Karami M, Ameri P. Prevalence of Nasopharyngeal Carriage of *Streptococcus pneumoniae* in Iran: A Meta-Analysis. *J Res Health Sci* 2015; 15: 141-146.
- [13] Yang H, Wei D, Yang K, Tang W, Luo Y, Zhang J. The prognosis of MGMT promoter methylation in glioblastoma patients of different race: a meta-analysis. *Neurochem. Res* 2014; 39: 2277-2287.
- [14] Zhang B, Liang L, Chen W, Liang C, Zhang S. An updated study to determine association between Gadolinium-based contrast agents and Nephrogenic Systemic Fibrosis. *Plos One* 2015; 10: e0129720.
- [15] Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-634.
- [16] Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088-1101.
- [17] Atakpo P, Vassar M. Cumulative meta-analysis by precision as a method to evaluate publication bias. *J Dermatol Sci* 2016; 83(3):251-253.
- [18] Poorolajal J, Haghdoost AA, Mahmoodi M, Majdzadeh R, Nasser-Moghaddam S, Fotouhi A. Capture-recapture method for assessing publication bias. *J Res Med Sci* 2010; 15: 107-115.
- [19] Guan M, Vandekerckhove J. A Bayesian approach to mitigation of publication bias. *Psychon. Bull Rev* 2016; 23(1):74-86.
- [20] Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Stat Med* 2010; 29: 2969-2983.
- [21] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess* 2010; 14: iii, ix-xi, 1-193.
- [22] Koletsi D, Valla K, Fleming PS, Chaimani A, Pandis N. Assessment of publication bias required improvement in oral health systematic reviews. *J Clin Epidemiol* 2016; 76:118-124.
- [23] Hayashino Y, Noguchi Y, Fukui T. Systematic evaluation and comparison of statistical tests for publication bias. *J Epidemiol* 2005; 15: 235-243.
- [24] 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.
- [25] Kicinski M. How does under-reporting of negative and inconclusive results affect the false-positive rate in meta-analysis? A simulation study. *BMJ Open* 2014; 4: e004831.