

Comparison of Different Methods of Detecting Publication Bias

Janhavi Kale, Cytel, Pune, India

Anwaya Nirpharake, Cytel, Pune, India

ABSTRACT:

Health professionals and policy makers want to make healthcare decisions based on the relevant research evidence. It can be achieved by doing systematic review. Systematic review and meta-analysis are conducted to evaluate the evidence about an intervention by combining results of multiple studies. Most of the time, even when conducted thoroughly, systematic review and meta-analysis are subject to bias. A considerable number of research findings are not published, especially in case of negative or non-confirmatory results. This phenomenon is widely recognized as publication bias. Various methods have been developed to detect publication bias in meta-analysis, such as graphical approaches and formal statistical tests. This paper compares the performance of some of these methods for various combinations of levels of heterogeneity and number of studies in meta-analysis. The results suggest that Fail Safe N and Trim and Fill methods perform better when number of component studies is small. On the other hand when number of component studies is large, Begg-Mazumdar and Egger's regression methods give better results. Our findings can serve as guideline for choosing appropriate method of detecting publication bias.

INTRODUCTION:

Important questions in clinical research are typically studied more than once by different research teams. Many times results from such different studies are diverse and conflicting, which makes the health care decision difficult. Hence a common practice is to make a comprehensive assessment of such results available in literature. This can be accomplished by systematic review and meta-analysis.

Publication bias is one of the major issues that researchers face while conducting systematic review and meta-analysis. Since published studies are more likely to find their way into a meta-analysis, any bias in the literature is likely to be reflected in the meta-analysis as well.

We will see more details about systematic review, meta-analysis and publication bias in the background section. After that we will briefly describe 4 methods of detecting publication bias. These four methods are compared using simulated data. The last section summarises the finding from simulation exercise.

BACKGROUND:

SYSTEMATIC REVIEW AND META ANALYSIS

Systematic review and meta-analysis are commonly used to identify and evaluate evidence about the intervention or exposures in human health. Systematic review shows what we know and what we don't know about the topic of interest. It is a method of summarizing research evidence. However, there is a need to assess the methodological quality of the design and execution of each study.

Systematic review is performed in below steps:

- Define a research problem
- Search and identify the relevant work
- Define study selection criteria i.e. inclusion and exclusion criteria
- Meta-analysis: Data synthesis and overall effect computations
- Interpret the findings

Searching of relevant literature has become easier because of online libraries such as Cochrane, PubMed, Medline, Embase etc. Hence accessibility to most of the publications has increased and this helps to include more studies in systematic review and meta-analysis. Meta-analysis is the heart of any systematic review. When

PhUSE 2017

results in identified publications are statistically combined then it is called as meta-analysis. Pharmaceutical companies use meta-analysis to gain approval for a new drug. It plays an important role in the planning of new studies. Results of Meta-analysis can improve the precision of the estimate of effect size and it resolves controversies arising from conflicting results and generates a new hypothesis.

Steps involved in Meta-analysis:

- Calculate outcome measure for each study and compute combined study effect
- Test for heterogeneity
- Test for publication bias
- Sensitivity analysis

Although the intent of a meta-analysis is to find and assess all studies meeting the inclusion criteria, still due to the selective publication of scientific research or because of neglecting (or not publishing) non-significant results, publication bias gets introduced. This may give misleading meta-analysis results.

PROBLEM OF PUBLICATION BIAS

Publication bias occurs when the publication of study results depends on statistical significance and direction of the intervention effect. So that published studies may be systematically different from those of unpublished studies. Since it is easier to include published studies in meta-analysis, any bias in the literature is likely to be reflected in the meta-analysis as well. This issue is generally known as publication bias. There are many reasons for presence of publication bias; below is the list for some of them:

1. Many studies remain unpublished because researchers do not published their work in the publication
2. Journals may be biased toward positive results because negative results are less likely to be published
3. Study sponsors or funding sources may be biased towards results that favor their interests.
4. Some studies get started but do not complete. Such studies are available sometimes in the abstracts of conferences but not in the clinical trial study databases. Hence they are missed
5. While searching for studies for systematic review, people rely on publications in only one language (generally only English). This introduces language bias.

The risk of publication bias can be reduced by identifying and including unpublished studies, redefining the inclusion and exclusion criteria. So first, one has to assess the extent of publication bias in a systematic review. There are several methods developed to detect publication bias. In this paper we will discuss some of those methods.

METHODS TO DETECT PUBLICATION BIAS

Researchers have developed many methods to assess the publication bias on meta-analysis. Five of them are discussed below.

GENERAL FUNNEL PLOT

The Simple method to detect publication bias is through subjective inspection of funnel plot. A funnel plot is a scatter plot of estimates of the treatment effects of each study against the measure of its precision ($1/\text{Standard Error}$). In the absence of publication bias, plot will look like symmetric inverted funnel. In the presence of publication bias, it will be an asymmetric funnel plot. This just a graphical representation, there are other statistical tests such as Fail-Safe N, Egger's Regression, Begg-Mazumdar rank correlation and Trim and Fill which provide objective examination of publication bias.

Figure (1a): Funnel Plot: No Publication Bias

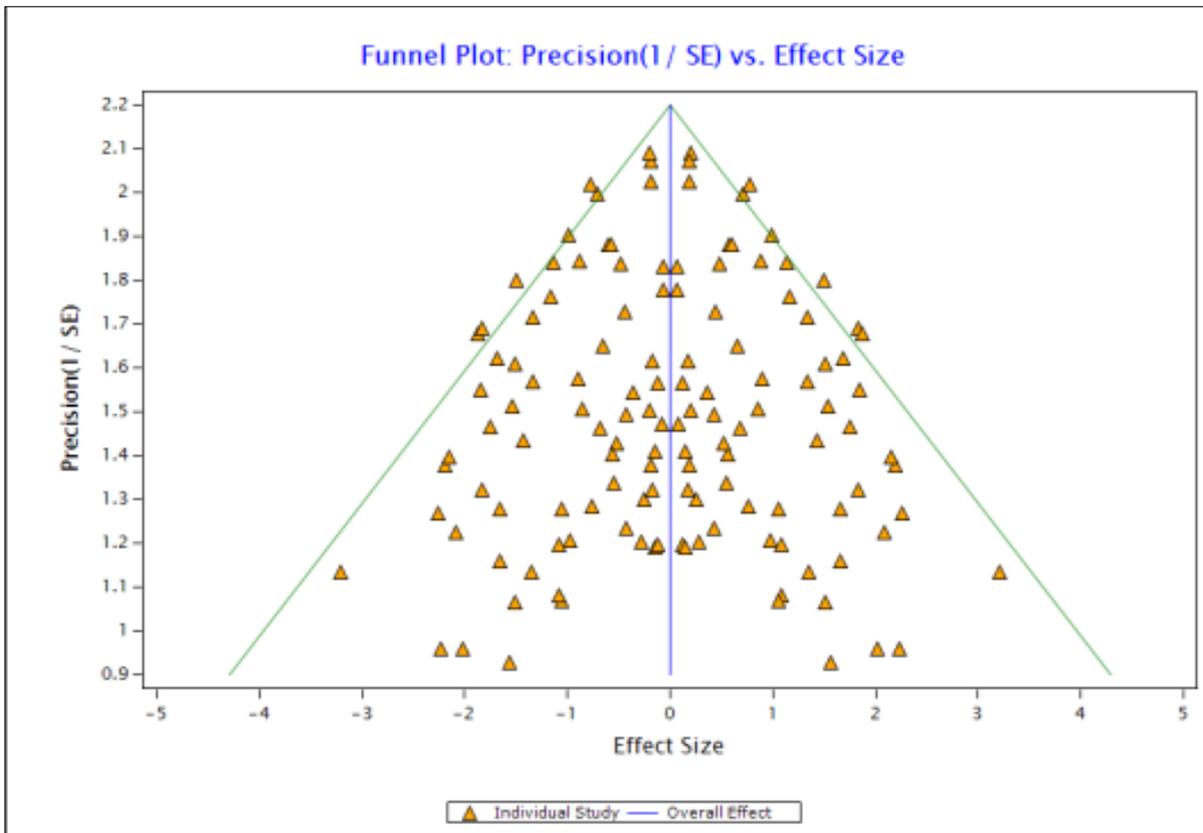
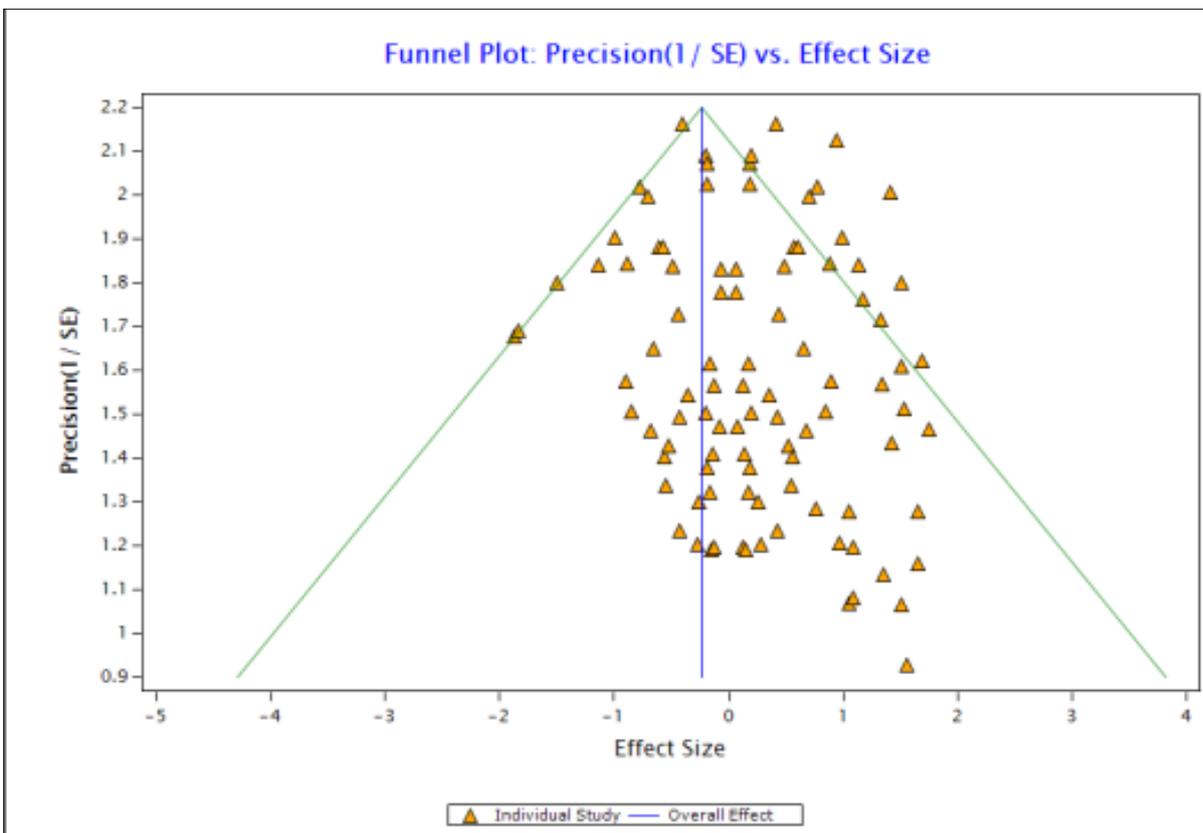


Figure (1b): Funnel Plot: Publication Bias



FAIL-SAFE N

Fail-Safe N stands for number of studies which are non-significant and hence are missing from meta-analysis. In this method number of non-significant studies is estimated. If these studies are added to Meta analysis it may change the overall p-value significant to non-significant. Hence, this method suggests how many missing studies we would need to retrieve and incorporate in the analysis before the p-value becomes non-significant. Small value of fail-safe N indicates presence of publication bias.

BEGG-MAZUMDAR

An adjusted rank correlation test is suggested by Begg-Mazumdar for identifying publication bias in a meta-analysis. A simple, formal test is constructed using correlation between standardized effect estimates and their variances. This tests the interdependence of variance and effect size using Kendall's method. The method uses Kendall's rank correlation rather than the ordinary product moment correlation, because Normal distributions of effect sizes are unlikely here. This test is fairly powerful for meta-analysis containing 75 or more component studies but has low power when meta-analysis has less than 25 studies. Hence the results of this test must be interpreted cautiously in small meta-analyses.

EGGERS REGRESSION

Egger et. al. (1997) suggested use of linear regression instead of correlation. This is the test for asymmetry of the funnel plot. The test checks for Y intercept = 0 from a linear regression of normalized effect estimate (effect size/standard error) against precision (1/Standard error). The regression line is expected to pass through the origin. According to Egger et. al., when publication bias is absent, all studies will be around the regression line of Y (effect size/standard error) on X (1/Standard error), with small sample studies near origin and large sample studies away from origin. In other words if the regression line has intercept value close to zero, there is lesser publication bias.

Figure (2a): Regression Plot: Publication Bias

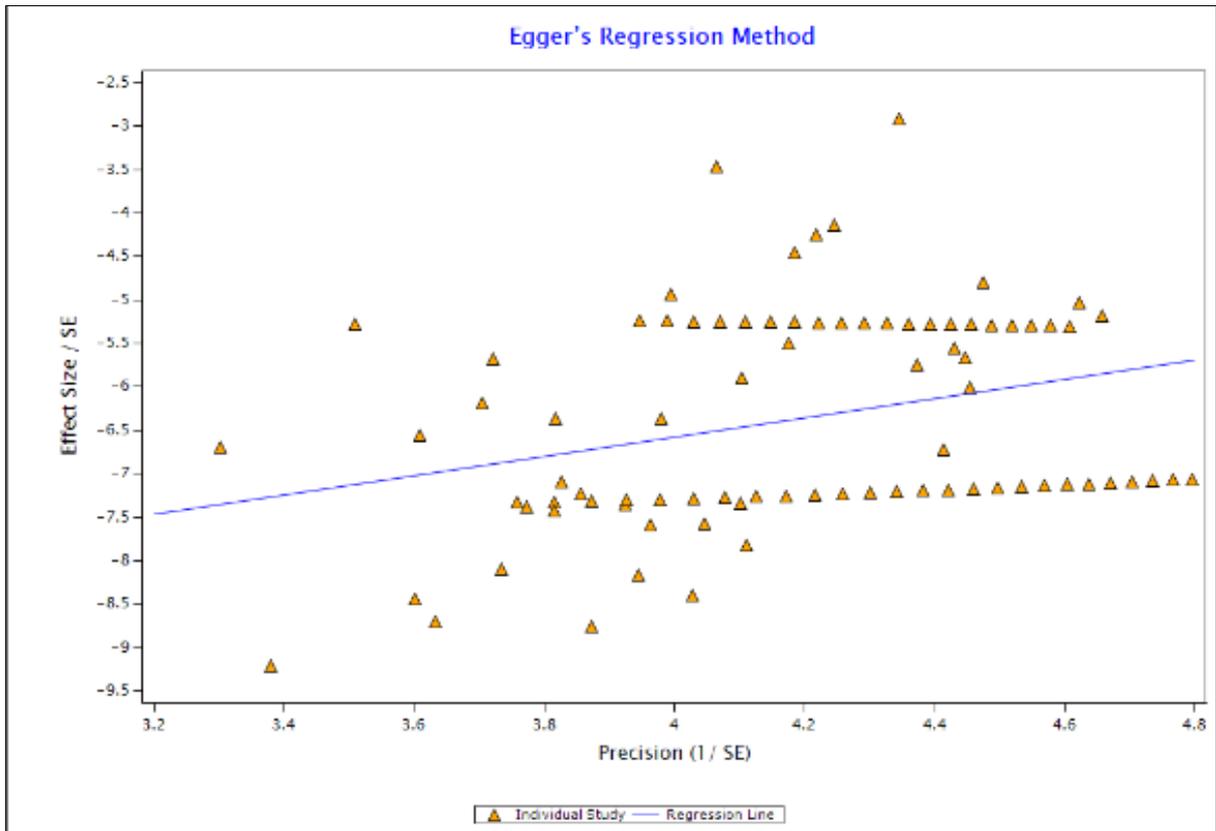
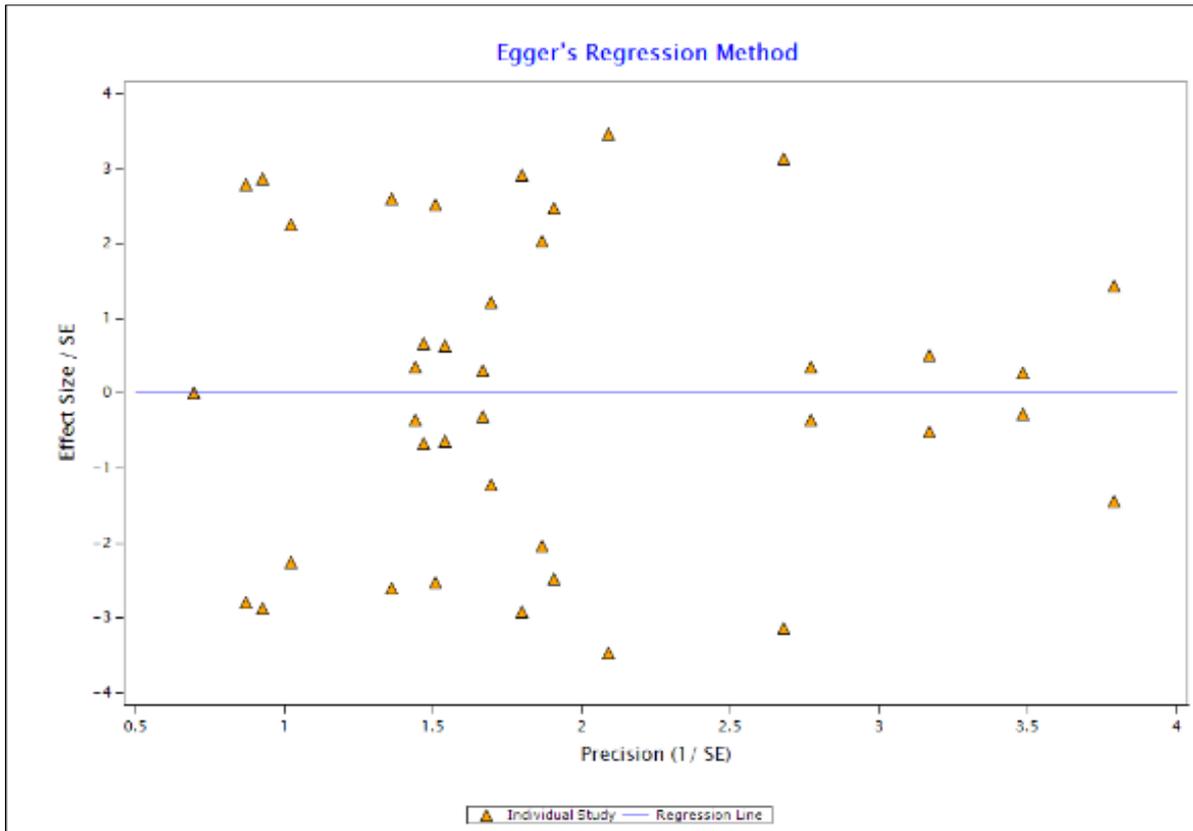


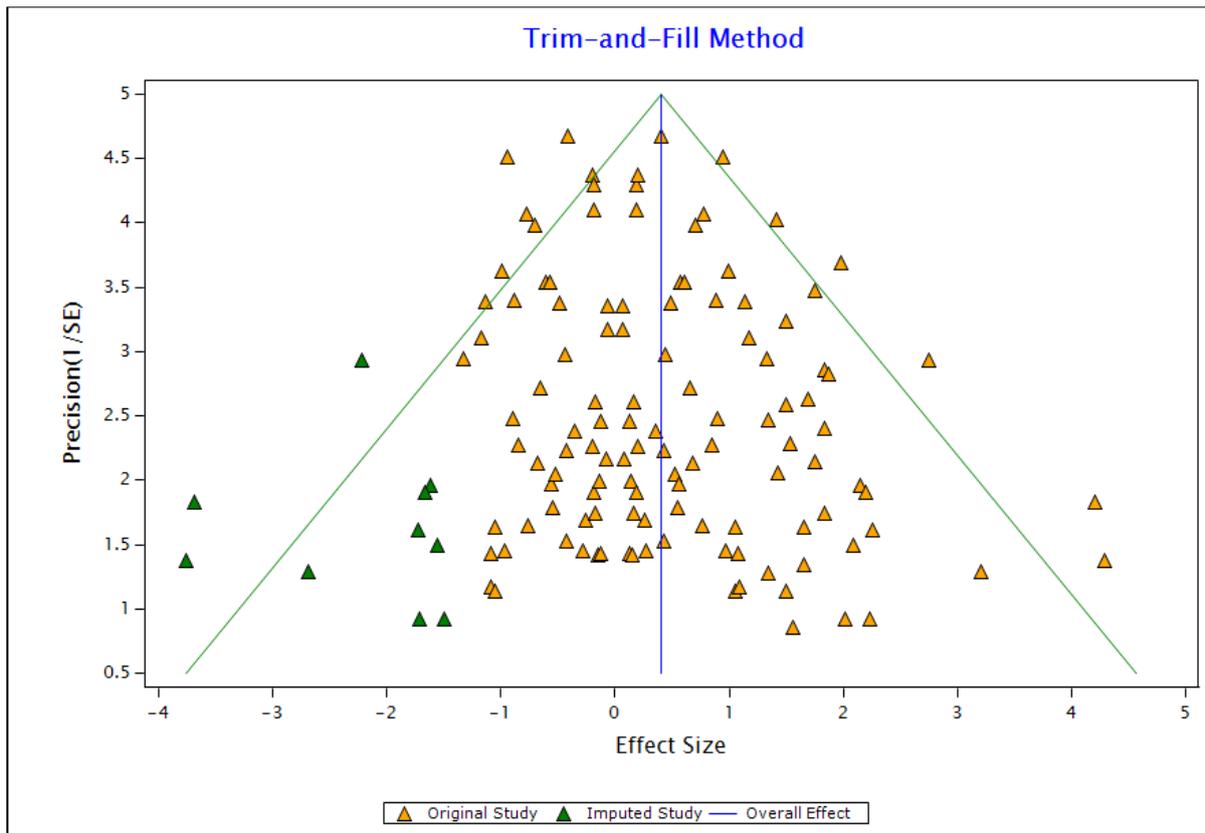
Figure (2b): Regression Plot: No Publication Bias



TRIM AND FILL

The trim and fill method gives us approximate number of studies to be imputed to make the Funnel plot symmetric. The trim and fill method has two stages. In the first stage data points (studies) are trimmed sequentially, starting with the least powerful (i.e. with smallest weight). The trimming process continues until symmetry is achieved in the funnel plot. Then from the remaining studies a new pooled estimate is calculated. Then we fill the studies removed in stage one by changing the sign of their effect sizes. This is second stage. Hence the name is trim and fill. Of course, we do not actually know that the filled studies match reality in any close way. But we can see the potential for incorrect conclusions if there was a publication bias. The trim and fill method does seem to provide a robust diagnostic method to enable us to detect the bias.

Figure (3): Trim-and-Fill Funnel Plot



NEED OF COMPARISON OF THESE METHODS

The methods discussed above have different approaches to detect the presence of publication bias. The problem is which method to refer for checking presence or absence of publication bias. In this paper above four statistical methods are compared for different scenarios.

ANALYSIS

For comparison purpose different datasets were simulated with following specifications-

- Four levels of number of studies, small (N=20), medium (N=50), large (N=100) and very large (N=150) and two levels of heterogeneity, low ($I^2 \leq 30$) and high ($I^2 > 30$) were considered. The I^2 statistic describes the percentage of variation among studies that is due to heterogeneity rather than chance.
- For each combination given above 500 datasets were generated. Hence, in total 4000 datasets were simulated.
- These datasets were symmetric or in other words have no publication bias.
- To induce publication bias, 20% of the studies from negative side of the effect size were removed from each dataset. This leads to generation of another 4000 datasets.

Meta-analysis was performed on each of the datasets using odds ratio of proportions test. Publication bias in the datasets was checked using following four methods-

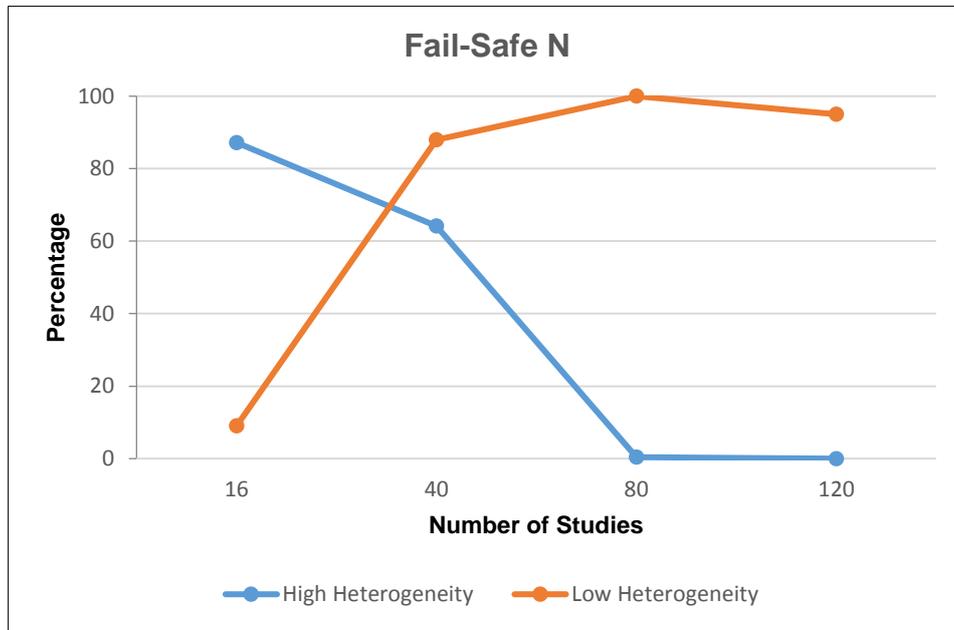
1. Fail-Safe N
2. Begg-Mazumdar
3. Egger's regression
4. Trim and Fill

PhUSE 2017

For datasets with no publication bias all the four methods showed correct results for all number of studies and both levels of heterogeneity. The next aim was to check the performance of these methods for datasets in presence of publication bias and varying levels of number of studies and heterogeneity. Performance of all these methods was compared through graphs as shown below-

The number of studies shown as X axis labels are 80% of N, here N = (20, 50, 100 and 150) as mentioned in 'Analysis' section bullet1. Here 20% studies with from negative side of effect size are were removed.

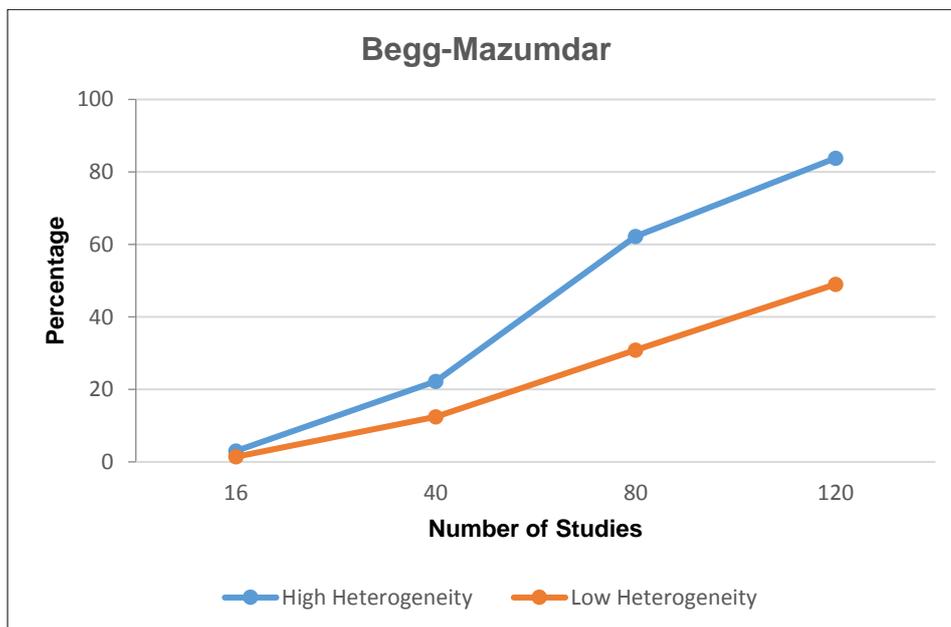
Figure (4a): Percentage of correctly detecting publication bias in Fail- Safe N method (Data with 20% publication bias)



From the above graph we can say that, as number of studies increases in a high heterogeneous data the ability of detecting publication bias of Fail-Safe N method goes on decreasing. When number of studies is above 80, this method clearly fails to detect publication bias.

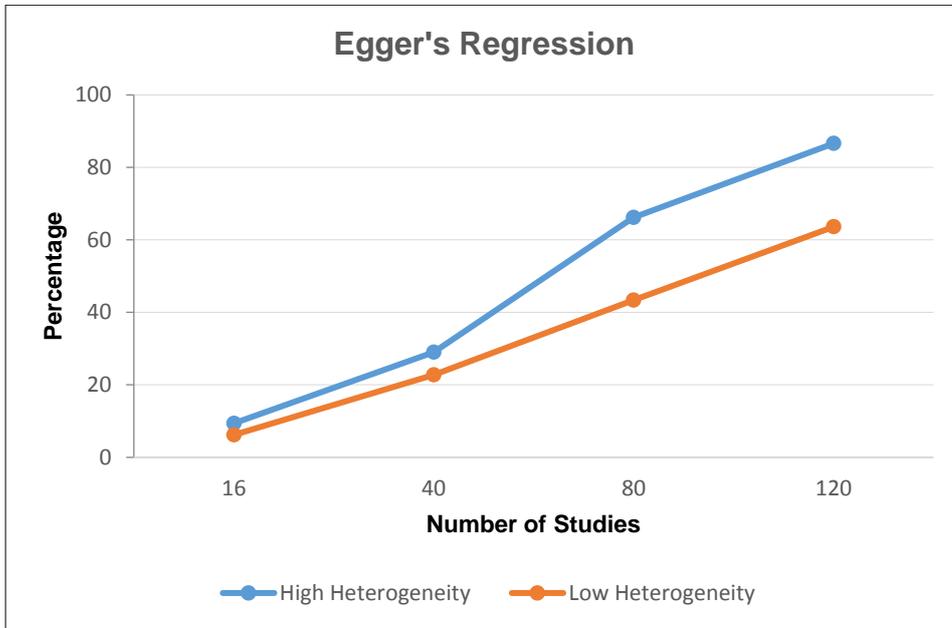
In case of low heterogeneous data, as number of studies increases ability of detecting publication bias of this method increases.

Figure (4b): Percentage of correctly detecting publication bias in Begg-Mazumdar method (Data with 20% publication bias)



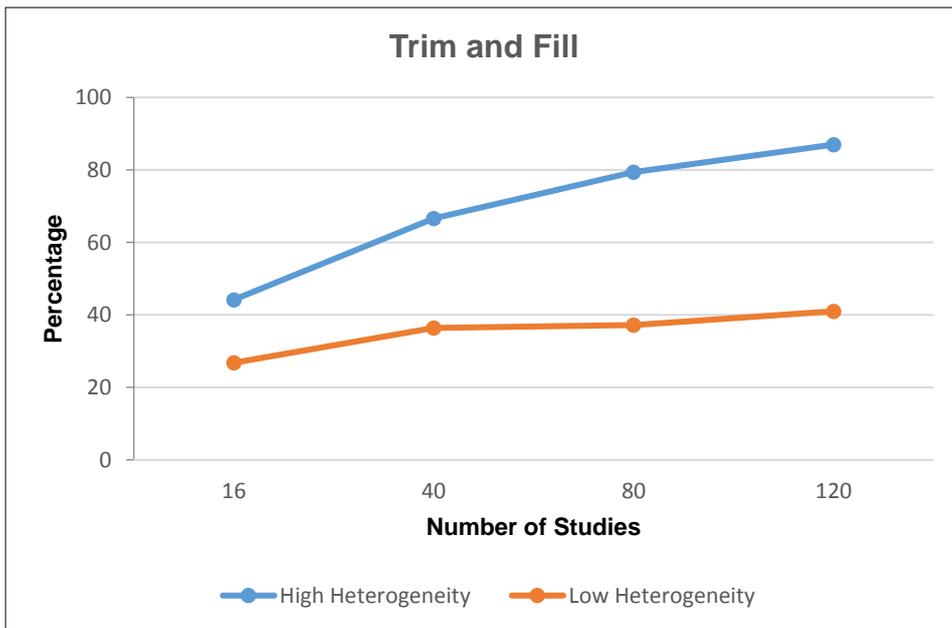
The performance of Begg-Mazumdar method becomes better with increasing number of studies in both low and high heterogeneous data. At all number of studies, percentage of correct detection of bias is higher for data with high heterogeneity than corresponding percentage of data with low heterogeneity.

Figure (4c): Percentage of correctly detecting publication bias in Egger's Regression method (Data with 20% publication bias)



The performance of Egger's regression method is similar to that of Begg-Mazumdar method.

Figure (4d): Percentage of correctly detecting publication bias in Trim and Fill method (Data with 20% publication bias)



For low heterogeneity the capability of detecting publication bias of Trim and Fill method doesn't change much with respect to number of studies. In high heterogeneous data this capability increase as number of studies

PhUSE 2017

increases. In case of high heterogeneity with large number of studies this methods detects publication bias in about 87% of datasets whereas in low heterogeneity this percentage is about 41.

CONCLUSION

1. Performance of Begg-Mazumdar, Eggers regression and Trim-and-Fill is better in highly heterogeneous data while Fail-Safe N works better in case of low heterogeneous data.
2. Begg-Mazumdar and Egger's regression methods perform quite similar. For high heterogeneous data Fail-Safe N method gives opposite results compared to other three methods.
3. In case of small number of studies it is advisable to use Fail Safe N and Trim and Fill method to detect publication bias, while in large dataset Begg-Mazumdar and Egger's regression methods gives reliable results.

FUTURE SCOPE

In this paper, only one level of publication bias (20% of missing studies) is considered. We can further explore performance of these four methods for different levels of publication bias. The comparison of methods can be extended for larger datasets. This work can also be extended for different types of effect size (Mean Difference, Ratio of Means, Log Hazard Ratio etc.)

REFERENCES

1. Colin B. Begg and Madhuchhanda Mazumdar (1994), Operating characteristics of a rank correlation test for publication bias, *Biometrics*, Vol. 50, No. 4 (Dec., 1994), pp. 1088-1101
2. Julian P T Higgins, Jonathan J Deeks (2003), Measuring inconsistency in meta-analyses, *BMJ* 2003;327:557
3. Matthias Egger, George Davey Smith, Martin Schneider and Christoph Minder (1997), Bias in Meta-Analysis Detected by a Simple, Graphical Test, *BMJ: British Medical Journal*, Vol. 315, No. 7109, *Meta-Analyses: Separating the Good from the Bad* (Sep. 13, 1997), pp. 629-634
4. Michael S. Rosenberg(2005), "The file drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis", *Evolution*, 59(2), 2005, pp. 464-468
5. Robert G. Orwin (1983), A Fail-Safe N for Effect Size in Meta-Analysis, *Journal of Educational Statistics*, Vol. 8, No. 2 (Summer, 1983), pp. 157-159
6. Sue Duval & Richard Tweedie (2000), A Nonparametric Trim and Fill Method of Accounting for Publication Bias in Meta Analysis, *Journal of the American Statistical Association*, 95:449, 89-98, DOI: 10.1080/01621459.2000.10473905
7. Sue Duval & Richard Tweedie (2000), Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis, *Biometrics*, Vol. 56, No. 2 (Jun., 2000), pp. 455-463

ACKNOWLEDGMENTS

The authors would like to thank all the colleagues at Cytel who gave their valuable inputs as and when required.

CONTACT INFORMATION

Janhavi Kale janhavi.kale@cytel.com +91-20-67090213
Anwaya Nirpharake anwaya.nirpharake@cytel.com +91-20-67090212

Cytel Statistical Software & Services Pvt. Ltd.
S.No. 150, Lohia Jain IT Park, "A" wing, 6th Floor, Paud Road, Kothrud
Pune 411038, India
Web: www.cytel.com