



Online article and related content
current as of July 28, 2009.

Comparison of Two Methods to Detect Publication Bias in Meta-analysis

Jaime L. Peters; Alex J. Sutton; David R. Jones; et al.

JAMA. 2006;295(6):676-680 (doi:10.1001/jama.295.6.676)

<http://jama.ama-assn.org/cgi/content/full/295/6/676>

Correction

[Contact me if this article is corrected.](#)

Citations

[This article has been cited 49 times.](#)
[Contact me when this article is cited.](#)

Topic collections

Journalology/ Peer Review/ Authorship; Quality of Care; Evidence-Based Medicine
[Contact me when new articles are published in these topic areas.](#)

Subscribe

<http://jama.com/subscribe>

Permissions

permissions@ama-assn.org
<http://pubs.ama-assn.org/misc/permissions.dtl>

Email Alerts

<http://jamaarchives.com/alerts>

Reprints/E-prints

reprints@ama-assn.org

Comparison of Two Methods to Detect Publication Bias in Meta-analysis

Jaime L. Peters, MSc

Alex J. Sutton, PhD

David R. Jones, PhD

Keith R. Abrams, PhD

Lesley Rushton, PhD

SYSTEMATIC REVIEWS AND META-analyses are commonly used to identify and evaluate evidence about interventions or exposures in human health. Even when conducted thoroughly, systematic reviews and meta-analyses can be subject to publication bias—studies being less likely to be published, hence less likely to be included in a systematic review or meta-analysis because of the size and/or statistical significance of their estimate of effect.¹ If publication bias occurs, the subsequent systematic review or meta-analysis of published literature may be misleading.

Of the methods available to researchers for the detection of publication bias, one of the simplest is the funnel plot.² This is a scatterplot of the estimate of effect from each study in the meta-analysis against a measure of its precision, usually $1/SE$ (FIGURE 1A). In the absence of bias, the plot should resemble a “funnel shape,” as smaller, less precise studies are more subject to random variation than larger studies when estimating an effect. In the presence of publication bias, some smaller studies reporting negative results will be missing, leading to an asymmetrical funnel plot. Of course, publication bias is not the only possible explanation for observed (or tested) funnel plot asymmetry.³ Between-study heterogeneity and small-study effects (the tendency for

Context Egger’s regression test is often used to help detect publication bias in meta-analyses. However, the performance of this test and the usual funnel plot have been challenged particularly when the summary estimate is the natural log of the odds ratio (lnOR).

Objective To compare the performance of Egger’s regression test with a regression test based on sample size (a modification of Macaskill’s test) with lnOR as the summary estimate.

Design Simulation of meta-analyses under a number of scenarios in the presence and absence of publication bias and between-study heterogeneity.

Main Outcome Measures Type I error rates (the proportion of false-positive results) for each regression test and their power to detect publication bias when it is present (the proportion of true-positive results).

Results Type I error rates for Egger’s regression test are higher than those for the alternative regression test. The alternative regression test has the appropriate type I error rates regardless of the size of the underlying OR, the number of primary studies in the meta-analysis, and the level of between-study heterogeneity. The alternative regression test has comparable power to Egger’s regression test to detect publication bias under conditions of low between-study heterogeneity.

Conclusion Because of appropriate type I error rates and reduction in the correlation between the lnOR and its variance, the alternative regression test can be used in place of Egger’s regression test when the summary estimates are lnORs.

JAMA. 2006;295:676-680

www.jama.com

smaller studies to show greater effects than larger studies) are also possible explanations.³ However, when the study summary estimates are odds ratios (ORs), there is a correlation between the natural log of OR (lnOR) and its SE, since the variance is a function of lnOR.⁴ This correlation is stronger the further the estimated OR is from unity.

Thus, some asymmetry observed in a funnel plot may be due to this correlation rather than publication bias. The effect of this correlation can be avoided by plotting effect size estimates against sample size, rather than precision. The meta-analysis plotted in Figure 1A uses data simulated from a model with no publication bias. However, it appears that some small negative studies could be missing from the bottom left-hand corner, which could

be interpreted as indicating publication bias. When these data are plotted against sample size (Figure 1B), the funnel plot looks more symmetrical. Although Figures 1A and 1B are not remarkably different, since only the y-axis has changed, the impact on Egger’s regression test can be quite striking, especially if the underlying OR is far from null.

Author Affiliations: Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, Leicester, England (Drs Sutton, Jones, and Abrams, and Ms Peters); MRC Institute for Environment and Health, Leicester, England (Dr Rushton). Dr Rushton is now with the Department of Epidemiology and Public Health, Imperial College London, London, England.

Corresponding Author: Jaime Peters, MSc, Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 22-28 Princess Rd W, Leicester, LE1 6TP, England (jlp9@leicester.ac.uk).

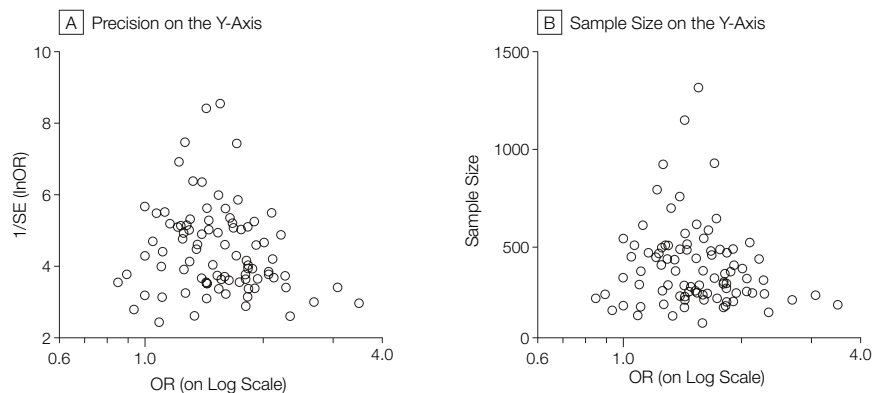
Statistical tests have been developed to provide more formal assessments for publication bias than the inspection of funnel plots. Egger's regression test⁵ is widely used (eg, as of January 11, 2006, the Web of Knowledge⁶ included 819 articles citing this article), is implemented in a number of software packages,⁷⁻¹⁰ and has become a standard procedure (eg, of 43 meta-analyses published in *JAMA* since 1997 in which an assessment of publication bias was made, 13 reported using Egger's regression test). Since it is based directly on the funnel plot, where the standardized effect estimate (*effect/SE*) is regressed on a measure of precision ($1/SE$), Egger's regression test is also subject to the effects of the correlation when using ORs.

In fact, Egger's regression test has been challenged because of its high type I error rates (the proportion of false-positive results) when ORs are used,^{3,11,12} a probable symptom of this correlation. As almost one third of the *JAMA* articles reviewed above used Egger's regression test when the summary estimates were ORs, this needs investigation. Using simulation analyses, we confirm that Egger's regression test is indeed inappropriate for ORs, particularly when the ORs are large and there is considerable between-study heterogeneity.^{3,4,11} We describe a simple alternative (a modified version of Macaskill's test,⁴ which is little used in practice) for detecting funnel plot asymmetry that avoids this correlation.

METHODS

We assessed the performance of 8 regression tests for funnel plot asymmetry, including Egger's regression test, using simulation methods. The tests¹³ differ in terms of the independent variable used, the weighting used, and whether random effects were included. In this article we compare the performance of Egger's regression test and the test found to have the most desirable properties compared with the remaining regression tests (results for all tests can be found in Peters et al¹³). Other modified tests are also being developed.¹⁴

Figure 1. Funnel Plots of a Meta-analysis Simulated With No Publication Bias



lnOR indicates natural log of the odds ratio; SE, standard error.

Characteristics of the simulated meta-analyses were based on a systematic review of meta-analyses of animal experiments,¹⁵ but the findings can be applied generally. Meta-analyses of 6, 16, 30, and 90 primary studies with underlying ORs of 1, 1.2, 1.5, 3, and 5 were simulated. The control group event rate was allowed to vary for each primary study. It was sampled from a uniform distribution with lower and upper limits of 0.3 and 0.7, respectively, representing a fairly common event in the control group. The treatment group event rate was calculated from this and the assumed underlying OR. The number of subjects in the control group in each study was based on the exponential of the normal distribution with a mean of 5 and variance of 0.3. The ratio of control to treated/exposed subjects was 1. The median sample size was around 300 in each simulated meta-analysis. Fixed- and random-effects models were used to simulate the meta-analyses. Since between-study heterogeneity is often found in meta-analyses,^{16,17} an understanding of the performance of tests for publication bias in such situations is essential in practice. The between-study heterogeneity parameter was calculated as a percentage of the average within-study variance estimate. From the fixed-effects model, the average within-study variance was calculated and between-study heterogeneity was then

defined to be 20%, 150%, and 500% of the within-study variance estimate. This reflects scenarios ranging from modest to considerable between-study heterogeneity. These levels of between-study heterogeneity corresponds to values of I^2 , the percentage of total variation across studies that is due to heterogeneity rather than chance,¹⁸ of 16.7%, 60%, and 83.3%, respectively.

Performance of the regression tests was assessed in the absence and presence of induced funnel plot asymmetry. Asymmetry was induced in 2 ways. First, it was induced on the basis of the P value associated with a study's effect size^{4,19,20} (the larger the P value the more likely that study was excluded from the meta-analysis). Since a study estimate is more likely to be statistically significant when the underlying OR is large, little publication bias is actually induced for the larger underlying ORs. Therefore, publication bias was also induced on the basis of study effect size.²¹ Studies with the most extreme negative effect sizes were excluded from the meta-analysis. Results are based on 1000 replications. The maximum SE of estimates for the type I error rates and power in the simulations is 1.7%. All simulations and analyses were carried out in Stata 8.2.⁷ For ease of presentation, only results for the underlying ORs of 1, 1.5, and 5 are given in the Figures (findings for underlying ORs of 1.2 and 3 follow the same general trend¹³).

RESULTS

An ideal test has the desired type I error rate (eg, 10% when statistical significance is specified from a 2-tailed test at $P < .10$, as is advocated for these tests⁵) and good power to detect asymmetry when it exists. In FIGURE 2, Egger's regression test exceeds the appropriate type I error rate of 10% for large underlying ORs. As the amount of between-study heterogeneity and number of primary studies increases, the type I error rates also increase, even for moderate ORs (ie, OR=1.5). We also

observed an imbalance in the tail probability areas for Egger's 2-tailed test,¹³ as previously demonstrated.^{4,11}

In the presence of funnel plot asymmetry, Egger's regression test appears reasonably powerful to detect this asymmetry (FIGURE 3), especially as the underlying OR and number of studies in the meta-analysis increase.

However, in assessing practical use of the test, power must be considered in light of the type I error rates (so that false-positive results are not mistaken for true-positive results). This

trade-off between power and type I error rates is similar to that between the sensitivity and specificity of a diagnostic test. Our findings and those of others^{3,4,11} lead us to have serious concerns over the practical use of Egger's test to identify funnel plot asymmetry for lnORs.

Of the 7 further regression models assessed, one model stands out in that its performance is superior to all the others, including Egger's regression test.¹³ This model and the simulated results from it are now discussed.

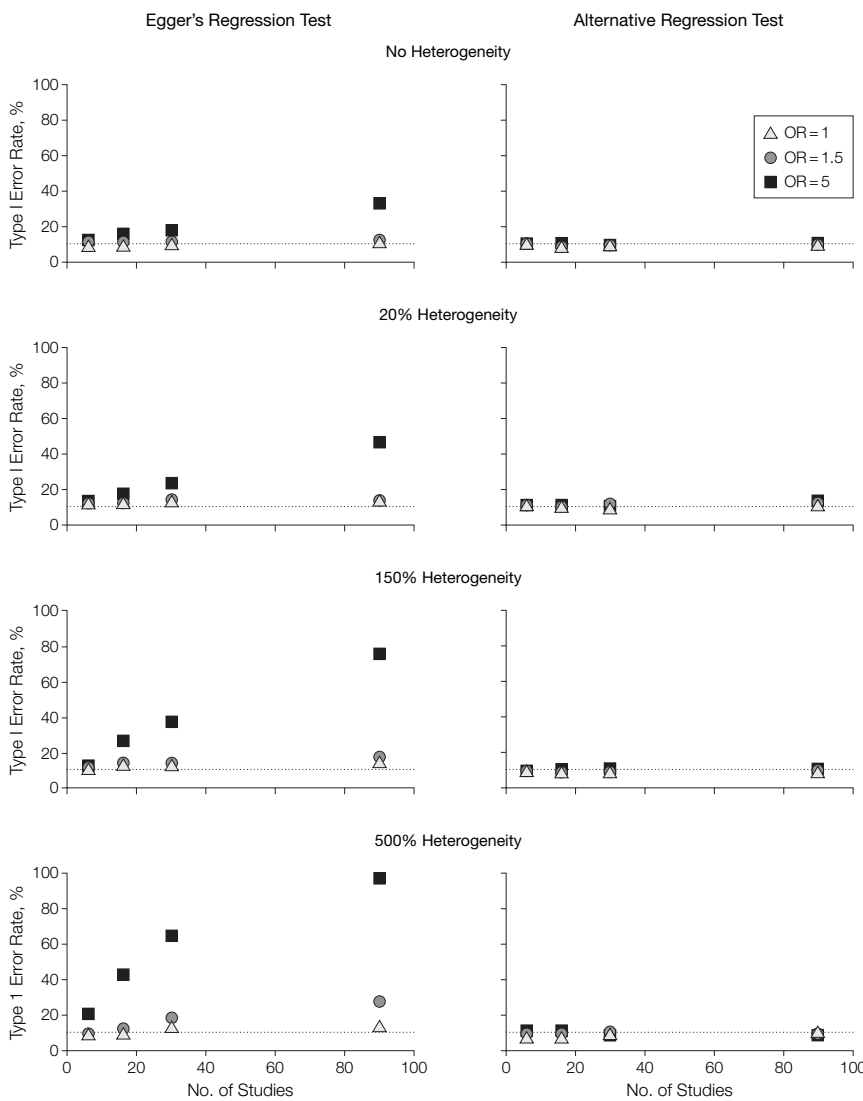
An Alternative to Egger's Regression Test

In preference to Egger's regression test, we recommend a simple weighted linear regression with lnOR as the dependent variable and the inverse of the total sample size as the independent variable. This is a minor modification of Macaskill's test,⁴ with the inverse of the total sample size as the independent variable rather than total sample size. Our results indicate that use of the inverse of the total sample size gives more balanced type I error rates in the tail probability areas than where there is no transformation of sample size.¹³ Use of sample size reduces the correlation between the lnOR and its SE.^{4,13} It also avoids violating an assumption of regression models that Egger's regression test does not avoid, as the independent variable, SE, is subject to random error (so that Egger's regression test is affected by regression dilution bias²²).

The weighting given to each study by the alternative regression test is based on the assumption that the null hypothesis is true, ie, the underlying OR=1. Choice of this weighting helps to reduce the correlation between the lnOR and the weight given to each study when the standard inverse variance weighting is used. Thus, appropriate type I error rates and balance in the tail probabilities are achieved.

Further explanation of the implications of this choice of weighting can be found in Macaskill et al⁴ and details of the weighting given to each study are given in Peters et al.¹³ Figure 2 shows

Figure 2. Type I Error Rates for Egger's Regression Test and the Alternative Regression Test



OR indicates odds ratio. Dotted line indicates the expected type I error rate (ie, 10%).

that the type I error rates for this alternative regression test are approximately 10%, as expected, regardless of the size of the underlying OR, the number of studies in the meta-analysis, and the amount of between-study heterogeneity, unlike those for Egger's regression test (Figure 2).

When there is little between-study heterogeneity, the alternative regression test and Egger's regression test appear to have moderate power to detect asymmetry when it is induced on the basis of P value (Figure 3) and high power when asymmetry is induced on the magnitude of the effect (data not shown).

When there is considerable heterogeneity (Figure 3), Egger's regression test is more powerful than the alternative regression test, however as discussed it is difficult to disentangle the high type I error rates of Egger's regression test from power.

COMMENT

Neither Egger's regression test nor the alternative regression test are particularly powerful in all scenarios. However, a test that may not be optimal, but performs well in all situations, is needed. Thus, although the alternative regression test is no more powerful than Egger's regression test, we recommend that the alternative be routinely used rather than Egger's regression test because it reduces the correlation between $\ln\text{OR}$ and its SE^4 through the choice of weighting and has appropriate type I error rates. The alternative regression test can easily be run in any software package allowing weighted linear regression. (Details on implementing this test in Stata⁷ are available from the authors.) In fact, applying this test to the meta-analysis illustrated in Figure 1 gives a nonsignificant result ($P = .18$), as one would expect since the data were simulated with no publication bias; Egger's regression test yields $P = .07$.

The alternative regression test is analogous to a funnel plot based on sample size. Thus, although contrary to the recommendations of Sterne and Egger²³

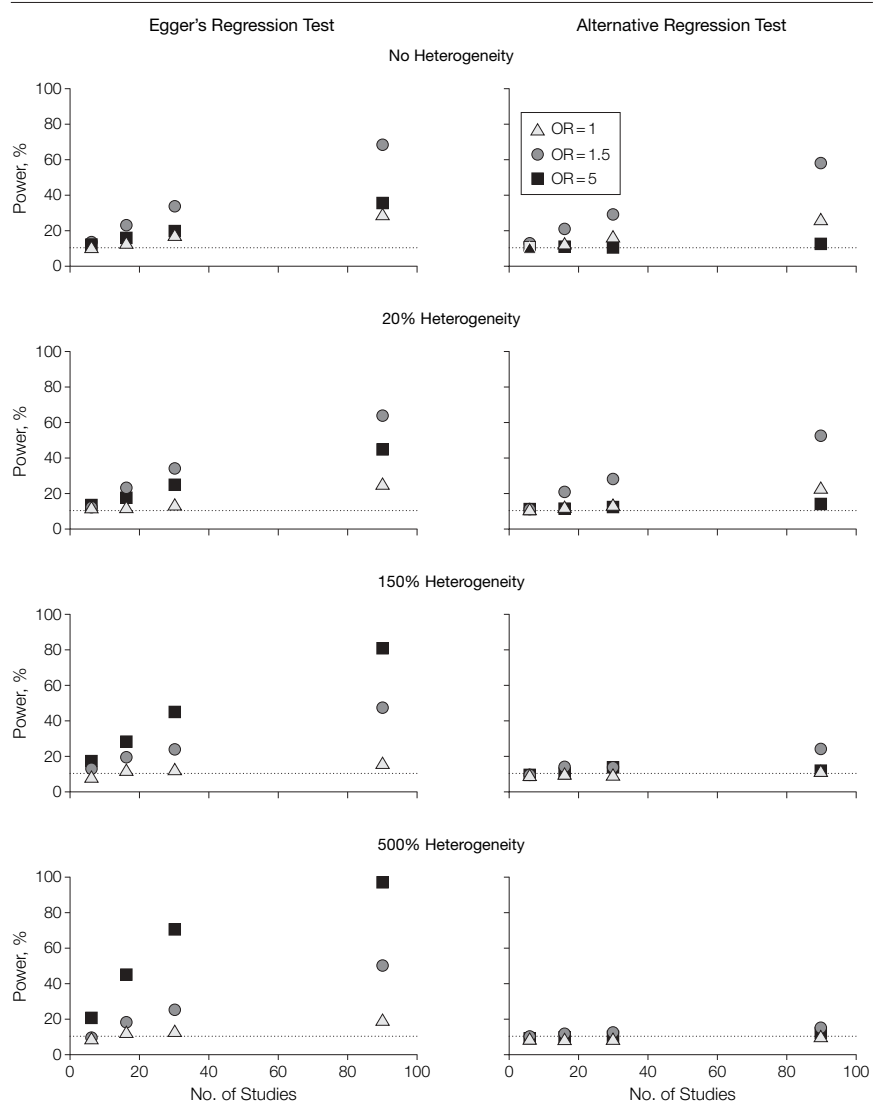
for choice of funnel plot axis, we advocate use of sample size¹³ for $\ln\text{OR}$ s.

We have also assessed use of the permutation test to obtain the P value for each test. The permutation test has been advocated for use in meta-regression to deal with inflated type I error rates.²⁴ Preliminary findings do not necessarily suggest better performance of tests based on P values from the permutation test compared with the usual t test.¹³ Extensions to, and performance of, these regression tests when some of the between-study heterogeneity can be

explained by a study-level covariate is ongoing work.

Our results, like those of some others,^{3,4} only concern synthesis of ORs. Findings of an investigation of Egger's regression test using relative risks (RRs) suggests a similar result: excessive type I error rates.¹¹ Although more work is needed on the performance of both tests when the summary estimate is not the OR, it is likely that other relative summary estimates (eg, RRs and risk differences) will be subject to effects similar to the correlation described above for the

Figure 3. Power of Egger's Regression Test and the Alternative Regression Test to Detect Publication Bias Induced by P Value



OR indicates odds ratio. Dotted line indicates the expected type I error rate (ie, 10%).

OR, thus suggesting Egger's regression test may not be appropriate. We did not consider meta-analyses of rare events. Evidence suggests the type I error rates for Egger's regression test are particularly high in these situations,^{3,11} but performance of the alternative regression test needs exploring.

Simply testing for the presence of asymmetry does not help obtain an unbiased estimate from the meta-analysis, particularly as there is overreliance on these tests (eg, a nonsignificant *P* value being taken as evidence that publication bias is not an issue). Our review of 43 meta-analyses published in *JAMA* since 1997 found that a number of approaches are taken when publication bias is suspected (as in 11 of the 43 meta-analyses). These include

acknowledging possible publication bias, but giving no detail on the extent or impact of such bias; discussing the possible implications of suspected publication bias and advising caution in the interpretation of the pooled estimate; and attributing inconsistent findings to the possible existence of publication bias. Other possible approaches include the trim and fill method²⁵ and best evidence synthesis approach.^{26,27} None of these approaches is adequate; while better methods of detecting and dealing with publication bias are being developed, we recommend that authors draw their conclusions cautiously, keeping the possibility of sensitivity to publication and related biases in mind.

Author Contributions: Ms Peters had full access to all of the data in the study and takes responsibility for

the integrity of the data and the accuracy of the data analysis.

Study concept and design: Peters, Sutton, Jones, Abrams, Rushton.

Analysis and interpretation of data: Peters, Sutton, Jones, Abrams, Rushton.

Drafting of the manuscript: Peters, Sutton, Jones, Abrams, Rushton.

Critical revision of the manuscript for important intellectual content: Peters, Sutton, Jones, Abrams, Rushton.

Statistical analysis: Peters, Sutton, Jones, Abrams, Rushton.

Obtained funding: Peters, Sutton, Jones, Abrams, Rushton.

Administrative, technical, or material support: Peters, Sutton, Jones, Abrams, Rushton.

Study supervision: Sutton.

Financial Disclosures: None reported.

Funding/Support: Ms Peters is funded through a UK Department of Health Evidence Synthesis Award.

Role of the Sponsor: The funding source had no role in any aspect of the study.

Acknowledgment: We are pleased to thank Petra Macaskill, PhD (School of Public Health, Sydney, Australia), for her comments on an earlier draft and suggestions for its improvement. Dr Macaskill did not receive any compensation.

REFERENCES

- Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess.* 2000;4:1-115.
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research.* Chichester, England: Wiley; 2000.
- Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000; 53:1119-1129.
- Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med.* 2001;20:641-654.
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315:629-634.
- Web of Knowledge. Available at: <http://wok.mimas.ac.uk>. Accessed September 20, 2005.
- Stata Statistical Software, Release 8.2.* College Station, Tex: Stata Corp; 2004.
- Borenstein M, Rothstein H. Comprehensive meta-analysis: a computer program for research synthesis. 1999. Available at: <http://www.meta-analysis.com>. Accessibility verified January 6, 2006.
- Rosenberg MS, Adams DC, Gurevitch J. *Metawin: Statistical Software for Meta-analysis: Version 2.0.* Sunderland, Mass: Sinauer Association; 1999.
- StatsDirect Statistical Software. Available at: <http://www.statsdirect.com>. Accessibility verified January 6, 2006.
- Schwarzer G, Antes G, Schumacher M. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Stat Med.* 2002;21:2465-2477.
- Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test: graphical test is itself biased. *BMJ.* 1998;316:470.
- Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. *Performance of Tests and Adjustments for Publication Bias in the Presence of Heterogeneity: Technical Report 05-01.* Leicester, England: Dept of Health Sciences, University of Leicester; 2005.
- Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints [published online ahead of print December 12, 2005]. *Stat Med.* doi: 10.1002/sim.2380. Accessed January 18, 2006.
- Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KA. *A Review of the Use of Systematic Review and Meta-analysis Methods to Evaluate Animal Toxicology Studies: Technical Report 04-02.* Leicester, England: Dept of Health Sciences, University of Leicester; 2004.
- Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med.* 2000;19:1707-1728.
- Villar J, Mackey ME, Carroli G, Donner A. Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: comparison of fixed and random effects models. *Stat Med.* 2001;20:3635-3647.
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539-1558.
- Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. *J Educ Behav Stat.* 1996;21:299-332.
- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* 1994;50:1088-1101.
- Duval S, Tweedie RL. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J Am Stat Soc.* 2000;95:89-98.
- Irwig L, Glasziou P, Wilson A, Macaskill P. Estimating an individual's true cholesterol level and response to intervention. *JAMA.* 1991;266:1678-1685.
- Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol.* 2001;54:1046-1055.
- Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med.* 2004;23:1663-1682.
- 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;56:455-463.
- Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educ Res.* 1986;15:5-11.
- Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol.* 1995; 48:9-18.