A MODIFIED TEST FOR BIAS IN META-ANALYSIS

Roger M Harbord, Jonathan A C Sterne, Matthias Egger. Department of Social Medicine, University of Bristol, UK.
email: roger.harbord@bristol.ac.uk

University of BRISTOL

ABSTRACT

Objective

To develop a test for funnel-plot asymmetry in meta-analysis that retains the reasonable power of the regression test (Egger *et al.* 1997)[1] while avoiding its tendency to give false-positive results too often in certain unusual situations.

Background

- Publication bias and related bias in meta-analysis is often examined by visually checking for asymmetry in funnel plots of standard error against treatment effect.
- Formal tests of funnel plot asymmetry have been proposed by Begg and Mazumdar[2] based on rank correlation, and Egger *et al.*[1] based on linear regression.
- When applied to binary outcome data, these tests can give false-positive rates that are higher than the nominal level in some situations (large treatment effects, or few events per trial, or all trials of similar sizes), although these situations are rare in practice. [3;4]
- This occurs chiefly because the estimates of treatment effect and its standard error are correlated, particularly in small trials. [3;4]

Methods

We suggest a modified test for funnel plot asymmetry in meta-analysis, still based on simple linear regression, but on a graph of:

$$Z/\sqrt{V} \text{ vs. } \sqrt{V}$$

– where Z is the efficient score and V is Fisher's information (both evaluated under the null hypothesis).

Results

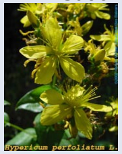
In simulations based on 78 published meta-analyses the false-positive rate of the modified test was within 2% of the nominal level of 10% in 75 (96%) of the studies, compared with 58 (74%) for the unmodified linear regression method and 29 (37%) for the rank correlation method.

Conclusions

This modified regression test is less likely to give false-positive results than the original version of the regression test. The power of the test requires further investigation but is likely to be similar to that of the unmodified test (i.e. higher than the rank correlation test though still limited, particularly when the number of studies is small). Issues remain about the best way of handling between-study heterogeneity that is still present after allowing for funnel-plot asymmetry.

EXAMPLE

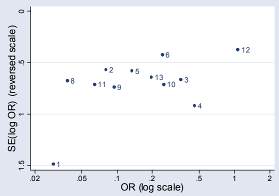
St John's Wort for depression[5]



	Treatment group		Control group	
study	non-responders	responders	non-responders	responders
1	15	10	25	0
2	7	27	29	9
3	6	14	11	9
4	21	4	23	2
5	12	20	27	6
6	22	28	42	13
7	23	0	24	0
8	15	29	41	3
9	10	15	21	3
10	10	10	16	4
11	11	19	27	3
12	26	29	26	31
13	5	20	14	11

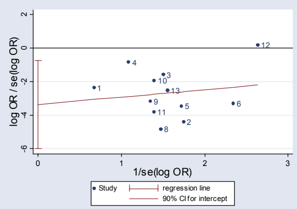
This meta-analysis is unusual in having a large effect size, low numbers of events, *and* all trials of similar sizes. Hence it may well cause problems for the unmodified regression test. We analyse odds ratios for non-response:

Funnel plot



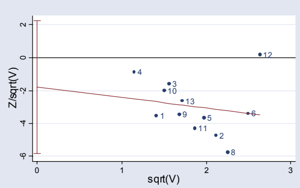
It is hard to judge whether or not there is funnel plot asymmetry visually.

Galbraith's radial plot showing unmodified regression test



Intercept is significant (P=0.043), suggesting funnel plot asymmetry and the possibility of bias.

A 'modified radial plot' showing modified regression test



Intercept is no longer significant (P=0.45), suggesting that the unmodified test gave a false-positive result.

The fact that a plot of $Z/\sqrt{V}$ *vs.* $\sqrt{V}$ is a close approximation to Galbraith's radial plot was pointed out by Galbraith himself [6].

EXPLANATION

Notation for 2 x 2 table for a single trial:

	Events	Non-events	Total
Treated group	s_T	f_T	n_T
Control group	s_C	f_C	n_C
Total	S	F	N

We will use odds ratios (*OR*) to measure treatment effect, although the method can be adapted for risk ratios.

The unmodified regression test is based on assessing the significance of the intercept in a simple linear regression of log *OR* / SE(log *OR*) *vs.* 1 / SE(log *OR*), (or equivalently, of the *slope* in a regression of log *OR* on SE(log *OR*) with weights 1 / Var(log *OR*)), where

$$\text{Estimated } \log OR = \log\left(\frac{s_T/f_T}{s_C/f_C}\right)$$

$$\text{Estimated } \quad \text{SE}(\log \; OR\,) = \sqrt{1/s_T + 1/f_T + 1/s_C + 1/f_C}$$

These estimates are correlated, especially in trials with low small numbers of events and a large treatment effect. This is the main cause of the high false-positive rate of the original regression test.

An alternative approach:

The efficient score for log *OR*
(evaluated under the null hypothesis, i.e. log *OR* = 0) is:

$$Z = s_T - n_T S / N$$

– i.e. simply the "Observed – Expected" (*O* – *E*) count in the top left cell.

Fisher's Information (the score variance) is:

$$V = \frac{n_T n_C S F}{N^3}$$

The usual Pearson's chi-squared test statistic is $\chi^2 = Z^2/V$

Notice that *V* depends only on the table *margins*, not the internal entries. *V is therefore (virtually) uncorrelated with the estimate of treatment effect.* However,

$$\text{SE}(\log OR) \approx 1\big/\sqrt{V}$$

One possibility would be to simply replace SE(log *OR*) on the horizontal axis by $1/\sqrt{V}$. We can do slightly better if we also replace log OR / se(log *OR*) on the vertical axis by $Z/\sqrt{V}$. This measure of the standardised effect size is the square-root of the Pearson chi-squared test statistic and has better properties in small studies.

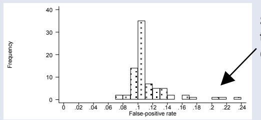
Hence we arrive at a modified regression test, based on regression of $Z/\sqrt{V}$ on $1/\sqrt{V}$.

FALSE-POSITIVE RATE

(Type I error rate) in simulations based on the characteristics of 78 published meta-analyses[4]

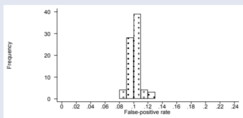
Using control group event rate, fixed-effect estimate and group sizes from each study, assuming no funnel plot asymmetry *and* no heterogeneity, so the true odds ratio is the same for all studies.

Unmodified regression test[4]



Some are much higher than the nominal level (set at P=0.1)

Modified regression test



False-positive rate always within 3% of nominal level.

HETEROGENEITY

A large amount of between-trial heterogeneity can cause problems for both versions of the test, as using simple linear regression amounts to using a *multiplicative* variation inflation factor to model heterogeneity rather than the more usual additive between-study variance component[7]. Work is ongoing on assessing whether more complex methods[7] then become necessary, or whether this simpler method is a satisfactory approximation.

Acknowledgement

RMH would like to acknowledge the influence of the teaching of John and Anne Whitehead, and to thank Anne Whitehead for useful email correspondence.

References

1. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629-34.
2. Begg CB,.Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;**50**:1088-101.
3. Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased (Letter). *BMJ* 1998;**316**:470.
4. Sterne JAC, Gavaghan DJ, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature . *Journal of Clinical Epidemiology* 2000;**53** :1119-29.
5. Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D. St John's wort for depression--an overview and meta-analysis of randomised clinical trials. *BMJ* 1996;**313**:253-8.
6. Galbraith R. A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med* 1988;**7**:889-94.
7. Thompson SG,.Sharp SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat.Med.* 1999;**18**:2693-708