The Stata Journal (2009)
9, Number 2, pp. 197–210

Updated tests for small-study effects in meta-analyses

Roger M. Harbord
Department of Social Medicine
University of Bristol
Bristol, UK
roger.harbord@bristol.ac.uk

Ross J. Harris
Centre for Infections
Health Protection Agency
London, UK
ross.harris@hpa.org.uk

Jonathan A. C. Sterne
Department of Social Medicine
University of Bristol
Bristol, UK

Abstract. This article describes an updated version of the `metabias` command, which provides statistical tests for funnel plot asymmetry. In addition to the previously implemented tests, `metabias` implements two new tests that are recommended in the recently updated *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2008). The first new test, proposed by Harbord, Egger, and Sterne (2006, *Statistics in Medicine* 25: 3443–3457), is a modified version of the commonly used test proposed by Egger et al. (1997, *British Medical Journal* 315: 629–634). It regresses $Z/\sqrt{V}$ against $\sqrt{V}$, where Z is the efficient score and V is Fisher's information (the variance of Z under the null hypothesis). The second new test is Peters' test, which is based on a weighted linear regression of the intervention effect estimate on the reciprocal of the sample size. Both of these tests maintain better control of the false-positive rate than the test proposed by Egger at al., while retaining similar power.

Keywords: sbe19_6, metabias, meta-analysis, publication bias, small-study effects, funnel plots

1 Introduction

Publication and related biases in meta-analysis are often examined by visually checking for asymmetry in funnel plots. However, such visual interpretation is inherently subjective. Tests for funnel plot asymmetry (small-study effects [Sterne, Gavaghan, and Egger 2000]) examine whether the association between estimated intervention effects and a measure of study size (such as the standard error of the intervention effect) is greater than might be expected to occur by chance.

This update to the `metabias` command (Steichen 1998; Steichen, Egger, and Sterne 1998) implements two new tests for funnel plot asymmetry that are recommended in the chapter addressing reporting biases (Sterne, Egger, and Moher 2008) in the recent update to the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2008). The modified version of Egger's test (Egger et al. 1997) proposed by Harbord, Egger, and Sterne (2006) still uses linear regression but is based on the efficient score and its variance, Fisher's information. The test proposed by Peters et al. (2006) is based on a weighted linear regression of the intervention effect estimate on the reciprocal of the sample size. These tests address mathematical problems that can occur with the commonly used Egger test and the rank correlation test proposed by Begg and Mazumdar (1994), which was also available in the original version of `metabias`. As with other recently updated meta-analysis commands, the syntax for `metabias` now corresponds to that for the main meta-analysis command, `metan`.

2 Syntax

`metabias` *varlist* [*if*] [*in*] `, egger harbord peters begg` [`graph nofit or rr` `level(`*#*`)` *graph_options*]

As in the `metan` command, *varlist* corresponds to either binary data—in this order: cases and noncases for the experimental group, then cases and noncases for the control group (d_1 h_1 d_0 h_0)—or the intervention effect and its standard error (*theta se_theta*).

The Harbord and Peters tests require binary data. Although the Egger test can be used with binary data, it is recommended only for studies with continuous (numerical) outcome variables and intervention effects measured as mean differences with the format *theta se_theta*.

`by` is allowed with `metabias`; see [D] **by**.

3 Options

`egger`, `harbord`, `peters`, and `begg` specify that the original Egger test, Harbord's modified test, Peters' test, or the rank correlation test proposed by Begg and Mazumdar (1994) be reported, respectively. There is no default; one test must be chosen.

`graph` displays a Galbraith plot (the standard normal deviate of intervention effect estimate against its precision) for the original Egger test or a modified Galbraith plot of $Z/\sqrt{V}$ versus $\sqrt{V}$ for Harbord's modified test. There is no corresponding plot for the Peters or Begg tests.

`nofit` suppresses the fitted regression line and confidence interval around the intercept in the Galbraith plot.

`or` (the default for binary data) uses odds ratios as the effect estimate of interest.

`rr` specifies that risk ratios rather than odds ratios be used. This option is not available for the Peters test.

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the width of confidence intervals**.

graph_options are any of the options documented in [G] **graph twoway scatter**. In particular, the options for specifying marker labels are useful.

4 Background

A funnel plot is a simple scatterplot of intervention effect estimates from individual studies against some measure of each study's size or precision (Light and Pillemer 1984; Begg and Berlin 1988; Sterne and Egger 2001). It is common to plot effect estimates on the horizontal axis and the measure of study size on the vertical axis. This is the opposite of the usual convention for twoway plots, in which the outcome (e.g., intervention effect) is plotted on the vertical axis and the covariate (e.g., study size) is plotted on the horizontal axis. The name "funnel plot" arises from the fact that precision of the estimated intervention effect increases as the size of the study increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias, the plot should approximately resemble a symmetrical (inverted) funnel. The `metafunnel` command (Sterne and Harbord 2004) can be used to display funnel plots, while the `confunnel` command (Palmer et al. 2008) can be used to display "contour-enhanced" funnel plots.

Funnel plots are commonly used to assess evidence that the studies included in a meta-analysis are affected by publication bias. If smaller studies without statistically significant effects remain unpublished, this can lead to an asymmetrical appearance of the funnel plot. However, the funnel plot is better seen as a generic means of displaying *small-study effects*—a tendency for the intervention effects estimated in smaller studies to differ from those estimated in larger studies (Sterne, Gavaghan, and Egger 2000). Small-study effects may be due to reporting biases, including publication bias and selective reporting of outcomes (Chan et al. 2004), poor methodological quality leading to spuriously inflated effects in smaller studies, or true heterogeneity (when the size of the intervention effect differs according to study size) (Egger et al. 1997; Sterne, Gavaghan, and Egger 2000). Apparent small-study effects can also be artifactual, because, in some circumstances, sampling variation can lead to an association between the intervention effect and its standard error (Irwig et al. 1998). Finally, small-study effects may be due to chance; this is addressed by statistical tests for funnel plot asymmetry.

For outcomes measured on a continuous (numerical) scale, tests for funnel plot asymmetry are reasonably straightforward. Using an approach proposed by Egger et al. (1997), we can perform a linear regression of the intervention effect estimates on their standard errors, weighting by 1/(variance of the intervention effect estimate). This looks for a straight-line relationship between the intervention effect and its standard error. Under the null hypothesis of no small-study effects, such a line would be vertical on a funnel plot. The greater the association between intervention effect and standard error, the more the slope would move away from vertical. The weighting is important to ensure that the regression estimates are not dominated by the smaller studies. It is mathematically equivalent, however, to a test of zero intercept in an unweighted regression on Galbraith's radial plot (Galbraith 1988) of the standard normal deviate, defined as the effect estimate divided by its standard error, against the precision, defined as the reciprocal of the standard error; and in fact, this method is used in `metabias`. If the regression line on a Galbraith plot is constrained to pass through the origin, its slope gives the summary estimate of fixed-effects meta-analysis as suggested by Galbraith. But if the intercept is estimated, a test of the null hypothesis of zero intercept tests for no association between the effect size and its standard error.

The Egger test has been by far the most widely used and cited approach to testing for funnel plot asymmetry. Unfortunately, there are statistical problems with this approach because the standard error of the log odds-ratio is correlated with the size of the odds ratio due to sampling variability alone, even in the absence of small-study effects (Irwig et al. 1998); see Deeks, Macaskill, and Irwig (2005) for an algebraic explanation of this phenomenon. This can cause funnel plots that were plotted using log odds-ratios (or odds ratios on a log scale) to appear asymmetric and can mean that p-values from the Egger test are too small, leading to false-positive test results. These problems are especially prone to occur when the intervention has a large effect, when there is substantial between-study heterogeneity, when there are few events per study, or when all studies are of similar sizes. Therefore, a number of authors have proposed alternative tests for funnel plot asymmetry. These are reviewed in a new chapter in the recently updated *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2008), which also gives guidance on testing for funnel plot asymmetry (Sterne, Egger, and Moher 2008).

4.1 Notation

We shall be primarily concerned with meta-analysis of 2×2 tables, where each study contains an intervention group and a control group, and the outcome is binary. We shall use the notation shown in table 1 for a single 2×2 table, using the letter d to denote those who experience the event of interest and h for those who do not, with subscripts 0 and 1 to indicate the control and intervention groups, respectively. We shall concentrate on the log odds-ratio, ϕ, as the measure of intervention effect, estimated by $\phi = \log(d_1 h_0 / d_0 h_1)$. The usual estimate of the variance of the log odds-ratio is the Woolf formula (Woolf 1955), $\text{Var}(\phi) = 1/d_0 + 1/h_0 + 1/d_1 + 1/h_1$, the square root of which gives the estimated standard error, $\text{SE}(\phi)$.

Table 1. Notation for a single 2×2 table

	Outcome		
	Experienced event d (disease)	Did not experience event h (healthy)	Total
Group 1 (intervention)	d_1	h_1	n_1
Group 2 (control)	d_0	h_0	n_0
Total	d	h	n

The Egger test is based on a two-sided t test of the null hypothesis of zero slope in a linear regression of ϕ against $\text{SE}(\phi)$, weighted by $1/\text{Var}(\phi)$ (Sterne, Gavaghan, and Egger 2000). This is equivalent to a two-sided t test of the null hypothesis of zero intercept in an unweighted linear regression of $\phi/\text{SE}(\phi)$ against $1/\text{SE}(\phi)$, which are the axes used in the Galbraith plot.

4.2 New tests for funnel plot asymmetry

Harbord's modification to Egger's test is based on the component statistics of the score test, namely, the efficient score, Z, and the score variance (Fisher's information), V. Z is the first derivative, and V is minus the second derivative of the log likelihood with respect to ϕ evaluated at $\phi = 0$ (Whitehead and Whitehead 1991; Whitehead 1997). The intercept in a regression of $Z/\sqrt{V}$ against $\sqrt{V}$ is used as a measure of the magnitude of small-study effects, with a two-sided t test of the null hypothesis of zero intercept giving a formal test for small-study effects. This is identical to a test of nonzero slope in a regression of Z/V against $1 = \sqrt{V}$ with weights V. If all marginal totals are considered fixed, V has no sampling error and hence no correlation with Z. If, as seems more realistic, n_0 and n_1 are considered fixed but d and h are not, the correlation remains lower than that between ϕ and its variance as calculated by the Woolf formula, leading to reduced false-positive rates (Harbord, Egger, and Sterne 2006).

Using standard likelihood theory (Whitehead 1997), it can also be shown that when ϕ is small and n is large, $\phi \approx Z/V$ and $\text{Var}(\phi) \approx 1/V$. It follows that the modified test becomes equivalent to the original Egger test when all trials are large and have small effect sizes. A plot of $Z = \sqrt{V}$ against $\sqrt{V}$ is therefore similar to Galbraith's radial plot of $\phi = \text{SE}(\phi)$ against $1/\text{SE}(\phi)$, as noted by Galbraith himself (Galbraith 1988).

When the parameter of interest is the log odds-ratio, ϕ, the efficient score is

$$Z = d_1 - dn_1/n$$

and the score variance evaluated at $\phi = 0$ is

$$V = n_0 n_1 dh/n^2(n-1)$$

The formula for V given above is obtained by using conditional likelihood, conditioning on the marginal totals d and h in table 1. When the parameter of interest is the log risk-ratio, it can be shown by using standard profile likelihood arguments that $Z = (d_1 n - d n_1)/h$ and $V = n_0 n_1 d/(nh)$.

The Peters test is based on a linear regression of ϕ on $1/n$, with weights dh/n. The slope of the regression line is used as a measure of the magnitude of small-study effects, with a two-sided t test of the null hypothesis of zero slope giving a formal test for small-study effects. This is a modification of Macaskill's test (Macaskill, Walter, and Irwig 2001), with the inverse of the total sample size as the independent variable rather than total sample size. The 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 (Peters et al. 2006). For balanced trials ($n_0 = n_1$), the weights dh/n are proportional to V.

When there is little or no between-trial heterogeneity, the Harbord and Peters tests have false-positive rates close to the nominal level while maintaining similar power to the original linear regression test proposed by Egger et al. (1997) (Harbord, Egger, and Sterne 2006; Peters et al. 2006; Rücker, Schwarzer, and Carpenter 2008).

5 Example

We shall use an example taken from a systematic review of randomized trials of nicotine replacement therapies in smoking cessation (Silagy et al. 2004), restricted to the 51 trials that used chewing gum as the method of delivery.

```
. use nicotinegum
(Nicotine gum for smoking cessation)
. describe
Contains data from nicotinegum.dta
  obs:            51                          Nicotine gum for smoking cessation
 vars:             5                          8 Jan 2009 12:02
 size:           663 (99.9% of memory free)   (_dta has notes)
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
trialid         byte   %9.0g
d1              int    %8.0g                  Intervention successes
h1              int    %9.0g                  Intervention failures
d0              int    %8.0g                  Control successes
h0              int    %9.0g                  Control failures
--------------------------------------------------------------------------------
Sorted by:  trialid
```

A standard fixed-effects meta-analysis, with intervention effects measured as odds ratios, suggests that there was a beneficial effect of the intervention (unusually for a medical meta-analysis, the event of interest here, smoking cessation, is good news rather than bad):

```
. metan d1 h1 d0 h0, or nograph
           Study     |     OR    [95% Conf. Interval]     % Weight
---------------------+---------------------------------------------------
1                    |  2.253       1.277     3.972          2.18
2                    |  1.850       0.989     3.460          1.98
3                    |  1.039       0.708     1.524          6.96
4                    |  1.416       0.599     3.350          1.21
5                    |  0.977       0.497     1.919          2.33
6                    |  4.773       1.910    11.932          0.70
7                    |  1.761       0.796     3.893          1.26
8                    |  3.159       1.138     8.768          0.69
9                    |  1.533       0.771     3.048          1.83
10                   |  1.385       0.888     2.160          4.55
11                   |  2.949       1.009     8.615          0.61
12                   |  2.293       1.239     4.245          1.92
13                   |  1.234       0.490     3.106          1.12
14                   |  2.624       1.026     6.708          0.87
15                   |  2.035       0.783     5.289          0.82
16                   |  2.822       1.329     5.994          1.13
17                   |  0.869       0.461     1.636          2.82
18                   |  0.887       0.326     2.408          1.10
19                   |  3.404       1.689     6.861          1.18
20                   |  2.170       1.101     4.279          1.59
21                   |  1.412       0.572     3.487          1.08
22                   |  2.029       0.800     5.148          0.97
23                   |  0.955       0.294     3.098          0.77
24                   |  1.250       0.472     3.311          1.00
25                   |  1.847       0.461     7.397          0.41
26                   |  3.327       1.371     8.077          0.76
27                   |  1.434       0.843     2.441          3.16
28                   |  1.333       0.428     4.155          0.72
29                   |  1.235       0.931     1.638         11.86
30                   |  3.142       1.776     5.558          1.84
31                   |  3.522       0.853    14.543          0.28
32                   |  1.168       0.704     1.937          3.81
33                   |  1.511       0.835     2.735          2.45
34                   |  3.824       1.150    12.713          0.39
35                   |  1.165       0.405     3.349          0.85
36                   |  1.345       0.349     5.188          0.50
37                   |  0.483       0.042     5.624          0.26
38                   |  1.713       1.212     2.421          6.33
39                   |  1.393       0.572     3.389          1.09
40                   |  1.844       1.204     2.822          4.30
41                   |  1.460       0.775     2.751          2.18
42                   |  1.269       0.776     2.075          3.84
43                   |  4.110       1.564    10.799          0.59
44                   |  2.082       1.504     2.881          6.57
45                   |  1.714       0.523     5.621          0.57
46                   |  1.294       0.749     2.236          2.98
47                   |  5.313       0.701    40.255          0.20
48                   |  2.703       0.509    14.357          0.25
49                   |  2.124       0.928     4.858          1.07
50                   |  1.760       0.549     5.643          0.58
51                   |  1.460       0.679     3.140          1.49
---------------------+---------------------------------------------------
M-H pooled OR        |  1.658       1.515     1.815        100.00
---------------------+---------------------------------------------------
```

```
Heterogeneity chi-squared =  62.04 (d.f. = 50) p = 0.118
I-squared (variation in OR attributable to heterogeneity) =  19.4%
Test of OR=1 : z=  10.99 p = 0.000
```

The `metan` command automatically creates the variables `_ES`, corresponding to the odds ratio, and `_selogES`, corresponding to the standard error of the log odds-ratio. We can use these to derive variables for input to the `metafunnel` command:

```
. generate logor = log(_ES)
. generate selogor = _selogES
```

We now use `metafunnel` to draw a funnel plot with the log odds-ratio, ϕ, on the horizontal axis and its standard error, $\mathrm{SE}(\phi)$, on the vertical axis. The `egger` option draws a line corresponding to the weighted regression of the log odds-ratio on its standard error that is the basis of Egger's regression test; see figure 1.

```
. metafunnel logor selogor, egger
```

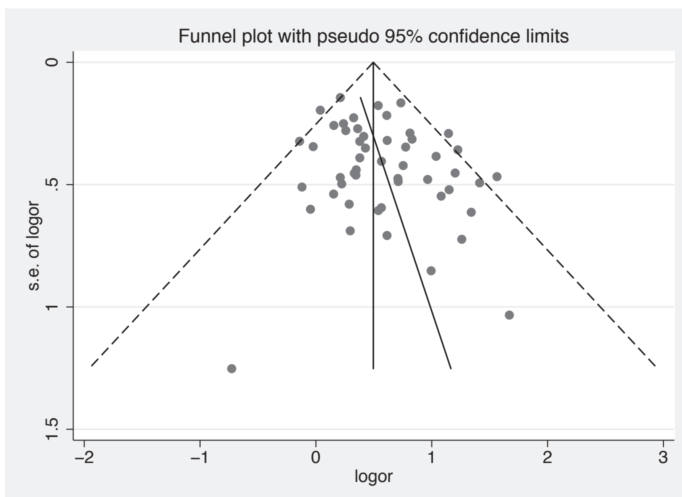

Figure 1. Funnel plot of the log odds-ratio, ϕ, against its standard error, $\mathrm{SE}(\phi)$, including the fitted regression line from the standard regression (Egger) test for small-study effects

The funnel plot appears asymmetric, with smaller studies (those with larger standard errors) tending to have larger (more beneficial) odds ratios. This may suggest publication bias.

We use the `metabias` command to perform a test of small-study effects employing the commonly used Egger test.

```
. metabias d1 h1 d0 h0, egger
Note: data input format tcases tnoncases ccases cnoncases assumed.
Note: odds ratios assumed as effect estimate of interest
Note: peters or harbord tests generally recommended for binary data
Egger´s test for small-study effects:
Regress standard normal deviate of intervention
effect estimate against its standard error
Number of studies =  51                             Root MSE      =   1.082
------------------------------------------------------------------------------
     Std_Eff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       slope |   .2832569   .1188368     2.38   0.021     .0444455    .5220683
        bias |   .7045941   .3566387     1.98   0.054    -.0120982    1.421286
------------------------------------------------------------------------------

Test of H0: no small-study effects          P = 0.054
```

The estimated bias coefficient is 0.705 with a standard error of 0.357, giving a p-value of 0.054. The test thus provides weak evidence for the presence of small-study effects.

The same results can be produced by using the derived variables `logor` and `selogor`:

```
. metabias logor selogor, egger
  (output omitted)
```

We now use Harbord's modified test:

```
. metabias d1 h1 d0 h0, harbord graph
Note: data input format tcases tnoncases ccases cnoncases assumed.
Note: odds ratios assumed as effect estimate of interest
Harbord´s modified test for small-study effects:
Regress Z/sqrt(V) on sqrt(V) where Z is efficient score and V is score variance
Number of studies =  51                             Root MSE      =   1.092
------------------------------------------------------------------------------
   Z/sqrt(V) |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sqrt(V) |   .3468707    .126528     2.74   0.009     .0926032    .6011382
        bias |   .5273137   .3866755     1.36   0.179    -.2497398    1.304367
------------------------------------------------------------------------------

Test of H0: no small-study effects          P = 0.179
```

The estimated intercept is 0.527 with a standard error of 0.387, giving a p-value of 0.179. The modified test thus suggests little evidence for small-study effects. The modified Galbraith plot of $Z/\sqrt{V}$ versus $\sqrt{V}$ is shown in figure 2.

(*Continued on next page*)

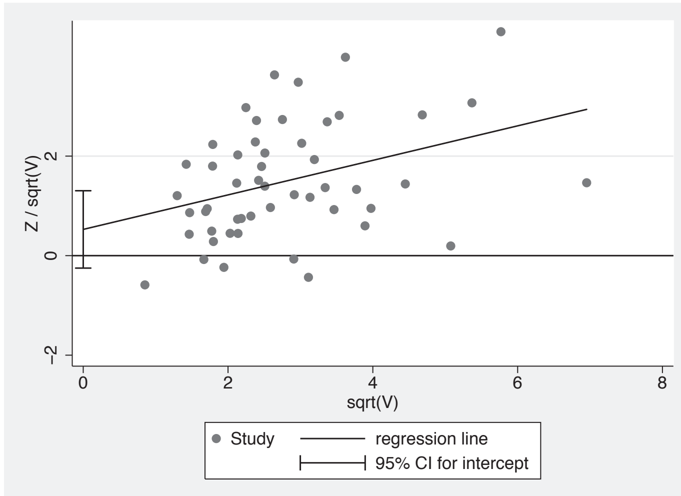

Figure 2. Modified Galbraith plot of $Z/\sqrt{V}$ versus $\sqrt{V}$

Finally, we will use Peters' test for small-study effects:

```
. metabias d1 h1 d0 h0, peters
Note: data input format tcases tnoncases ccases cnoncases assumed.
Note: odds ratios assumed as effect estimate of interest
Peter´s test for small-study effects:
Regress intervention effect estimate on 1/Ntot, with weights SF/Ntot
Number of studies =  51                                Root MSE      =   .3897
------------------------------------------------------------------------------
     Std_Eff |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bias |   26.20225   14.58572     1.80   0.079    -3.108842    55.51334
    constant |   .4197904   .0776552     5.41   0.000     .2637364    .5758443
------------------------------------------------------------------------------
Test of H0: no small-study effects          P = 0.079
```

In this example, the p-value from Peters' test is closer to that from Egger's test than it is to the p-value from Harbord's test. These differing results emphasize the importance of selecting a test in advance; picking a test result from among several is strongly discouraged.

6 Saved results

For all tests, the following scalars are returned:

`r(N)`	number of studies
`r(p_bias)`	p-value of the bias estimate

For the regression-based tests (Harbord, Peters, and Egger), the following scalars are returned:

`r(df_r)`	degrees of freedom
`r(bias)`	estimate of bias (the constant in the regression equation for the Egger and Harbord tests, and the slope for the Peters test)
`r(se_bias)`	standard error of bias estimate
`r(rmse)`	root mean squared error of fitted regression model

For Begg's test, the following scalars are returned:

`r(score)`	Kendall's score ($P-Q$)
`r(score_sd)`	standard deviation of Kendall's score
`r(p_bias_ncc)`	p-value for Begg's test (not continuity-corrected)

7 Discussion

We have described how to use the `metabias` command to perform two tests for funnel plot asymmetry. These tests are among those recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2008) because they reduce the inflation of the false-positive rate (type I error) that can occur for the Egger test, while retaining power compared with alternative tests. `metabias` allows only one test to be specified. Systematic reviewers should ideally specify their chosen test in advance of the analysis and should avoid choosing from among the results of several tests. Although simulation studies comparing the different tests have been reported (Harbord, Egger, and Sterne 2006; Peters et al. 2006; Rücker, Schwarzer, and Carpenter 2008), no test currently has been shown to be superior in all circumstances. A fuller discussion of these issues is available in chapter 10 (Sterne, Egger, and Moher 2008) of the *Cochrane Handbook*.

Tests for funnel plot asymmetry should not be seen as a foolproof method of detecting publication bias or other small-study effects. We recommend that tests for funnel plot asymmetry be used only when there are at least 10 studies included in the meta-analysis. Even then, power may be low. False-positive results may occur in the presence of substantial between-study heterogeneity, and no test performs well when all studies are of a similar size. Although funnel plots, and tests for funnel plot asymmetry, may alert us to a problem that needs considering when interpreting the results of a meta-analysis, they do not provide a solution to this problem.

(*Continued on next page*)

8 Acknowledgment

We are grateful to Thomas Steichen, who wrote the original version of the `metabias` command and gave us permission to update it.

Some of the guidance in this article is based on the chapter "Addressing reporting biases" (Sterne, Egger, and Moher 2008), published in the new *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins and Green 2008).

9 References

Begg, C. B., and J. A. Berlin. 1988. Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society, Series A* 151: 419–463.

Begg, C. B., and M. Mazumdar. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50: 1088–1101.

Chan, A.-W., A. Hróbjartsson, M. T. Haahr, P. C. Gøtzche, and D. G. Altman. 2004. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association* 291: 2457–2465.

Deeks, J. J., P. Macaskill, and L. Irwig. 2005. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 58: 882–893.

Egger, M., G. Davey Smith, M. Schneider, and C. Minder. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315: 629–634.

Galbraith, R. F. 1988. A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine* 7: 889–894.

Harbord, R. M., M. Egger, and J. A. C. Sterne. 2006. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 25: 3443–3457.

Higgins, J. P. T., and S. Green, ed. 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: Wiley.

Irwig, L., P. Macaskill, G. Berry, and P. Glasziou. 1998. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. *British Medical Journal* 316: 470.

Light, R. J., and D. B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press.

Macaskill, P., S. D. Walter, and L. Irwig. 2001. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 20: 641–654.

Palmer, T. M., J. L. Peters, A. J. Sutton, and S. G. Moreno. 2008. Contour-enhanced funnel plots for meta-analysis. *Stata Journal* 8: 242–254.

Peters, J. L., A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton. 2006. Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association* 295: 676–680.

Rücker, G., G. Schwarzer, and J. Carpenter. 2008. Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* 27: 746–763.

Silagy, C., T. Lancaster, L. Stead, D. Mant, and G. Fowler. 2004. Nicotine replacement therapy for smoking cessation. *Cochrane Database of Systematic Reviews* 3: CD000146.

Steichen, T. J. 1998. sbe19: Tests for publication bias in meta-analysis. *Stata Technical Bulletin* 41: 9–15. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 125–133. College Station, TX: Stata Press.

Steichen, T. J., M. Egger, and J. A. C. Sterne. 1998. sbe19.1: Tests for publication bias in meta-analysis. *Stata Technical Bulletin* 44: 3–4. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 84–85. College Station, TX: Stata Press.

Sterne, J. A. C., and M. Egger. 2001. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 54: 1046–1055.

Sterne, J. A. C., M. Egger, and D. Moher. 2008. Addressing reporting biases. In *Cochrane Handbook for Systematic Reviews of Interventions*, ed. J. P. T. Higgins and S. Green, 297–334. Chichester, UK: Wiley.

Sterne, J. A. C., D. Gavaghan, and M. Egger. 2000. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 53: 1119–1129.

Sterne, J. A. C., and R. M. Harbord. 2004. Funnel plots in meta-analysis. *Stata Journal* 4: 127–141.

Whitehead, A., and J. Whitehead. 1991. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 10: 1665–1677.

Whitehead, J. 1997. *The Design and Analysis of Sequential Clinical Trials*. Rev. 2nd ed. Chichester, UK: Wiley.

Woolf, B. 1955. On estimating the relation between blood group and disease. *Annals of Human Genetics* 19: 251–253.

About the authors

Roger Harbord is a research associate in medical statistics in the Department of Social Medicine at the University of Bristol, UK. He is a co-convenor of the Cochrane Collaboration's Screening and Diagnostic Tests Methods Group.

Ross Harris is a statistician for the Health Protection Agency. His research interests include evidence synthesis and the epidemiology of sexually transmitted diseases.

Jonathan Sterne is Professor of Medical Statistics and Epidemiology in the Department of Social Medicine at the University of Bristol, UK. His research interests include statistical methods for epidemiology and health services research, causal models, meta-analysis and systematic reviews, the epidemiology of sexually transmitted infections, and the epidemiology of asthma and allergic diseases.