The Stata Journal (2004)
4, Number 2, pp. 127–141

Funnel plots in meta-analysis

Jonathan A. C. Sterne and Roger M. Harbord
Department of Social Medicine, University of Bristol
Canynge Hall, Whiteladies Road, Bristol BS8 2PR UK

Abstract. Funnel plots are a visual tool for investigating publication and other bias in meta-analysis. They are simple scatterplots of the treatment effects estimated from individual studies (horizontal axis) against a measure of study size (vertical axis). The name "funnel plot" is based on the precision in the estimation of the underlying treatment effect increasing as the sample size of component studies increases. Therefore, in the absence of bias, results from small studies will scatter widely at the bottom of the graph, with the spread narrowing among larger studies. Publication bias (the association of publication probability with the statistical significance of study results) may lead to asymmetrical funnel plots. It is, however, important to realize that publication bias is only one of a number of possible causes of funnel-plot asymmetry—funnel plots should be seen as a generic means of examining small study effects (the tendency for the smaller studies in a meta-analysis to show larger treatment effects) rather than a tool to diagnose specific types of bias. This article introduces the `metafunnel` command, which produces funnel plots in Stata. In accordance with published recommendations, standard error is used as the measure of study size. Treatment effects expressed as ratio measures (for example risk ratios or odds ratios) may be plotted on a log scale.

Keywords: st0061, metafunnel, funnel plots, meta-analysis, publication bias, small-study effects

1 Introduction

The substantial recent interest in meta-analysis (the statistical methods that are used to combine results from a number of different studies) is reflected in a number of user-written commands that do meta-analysis in Stata. Meta-analyses should be based on *systematic reviews* of relevant literature. A systematic review is a systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. The main feature that distinguishes systematic from narrative reviews is a methods section that clearly states the question to be addressed and the methods and criteria to be employed for identifying and selecting relevant studies and extracting and analyzing information (Egger, Davey Smith, and Altman 2001).

While systematic reviews and meta-analyses have the potential to produce precise estimates of treatment effects that reflect all of the relevant literature, they are not immune to bias. *Publication bias*—the association of publication probability with the statistical significance of study results—is well documented as a problem in the medical research literature (Stern and Simes 1997). Further, it has been demonstrated that randomized controlled trials for which concealment of treatment allocation is not adequate,

© 2004 StataCorp LP

or which are not double blind, produce estimated treatment effects that appear more beneficial (Schulz et al. 1995).

2 Funnel plots

Funnel plots are simple scatterplots of the treatment effects estimated from individual studies against a measure of study size. The name "funnel plot" is based on the precision in the estimation of the underlying treatment effect increasing as the sample size of component studies increases. Results from small studies will therefore scatter widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias, the plot will resemble a symmetrical, inverted funnel, as shown in the top graph of figure 1.

If there is bias, for example, because smaller studies showing no statistically significant effects (open circles in figure 1) remain unpublished, then such publication bias will lead to an asymmetrical appearance of the funnel plot with a gap in the right bottom side of the graph (middle graph of figure 1). In this situation, the combined effect from meta-analysis will overestimate the treatment's effect. The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

It is important to realize that publication bias is only one of a number of possible explanations for funnel-plot asymmetry; these are discussed in more detail in section 2.3. For example, trials of lower quality yield exaggerated estimates of treatment effects (Schulz et al. 1995). Smaller studies are, on average, conducted and analyzed with less methodological rigor than larger studies (Egger et al. 2003), so asymmetry may also result from the overestimation of treatment effects in smaller studies of lower methodological quality (bottom graph of figure 1). Unfortunately, funnel-plot asymmetry has often been equated with publication bias without consideration of its other possible explanations; for example, the help file for the `metabias` command in Stata (written in 1998) refers only to publication bias.

(*Continued on next page*)

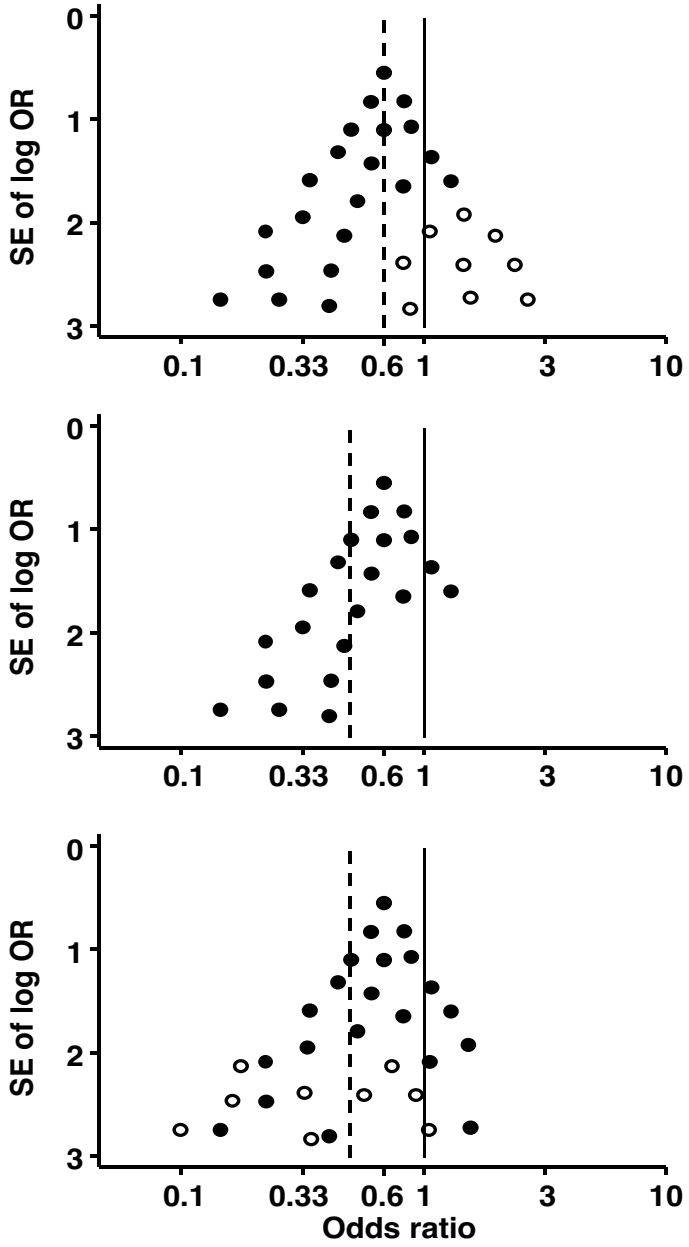

Figure 1: Hypothetical funnel plots: (top) symmetrical plot in the absence of bias (open circles indicate smaller studies showing no beneficial effects); (middle) asymmetrical plot in the presence of publication bias (smaller studies showing no beneficial effects are missing); (bottom) asymmetrical plot in the presence of bias due to low methodological quality of smaller studies (open circles indicate small studies of inadequate quality whose results are biased towards larger beneficial effects).

Although it is conventional to plot treatment effects on the horizontal axis and the measure of study size on the vertical axis, it is certainly not an error to plot the axes the other way around. Indeed, such a choice is arguably more consistent with standard statistical practice in that the variable on the vertical axis is usually hypothesized to depend on the variable on the horizontal axis. Such funnel plots can be plotted in Stata using the `metabias` command (Steichen 1998; Steichen, Egger, and Sterne 1998).

2.1 Choice of axis in funnel plots

The majority of endpoints in randomized trials of medical treatments are binary, with treatment effects most commonly expressed as ratio measures (odds ratio, risk ratio, or hazard ratio). (This may not be true of trials in other disciplines, such as psychology or social research.) The use of ratio measures is justified by empirical evidence that there is less between-trial heterogeneity in treatment effects based on ratio measures than difference measures (Deeks and Altman 2001; Engels et al. 2000). As is generally the case in meta-analysis, the *log* of the ratio measure and its standard error are used in funnel plots.

Sterne and Egger (2001) consider choice of axis in funnel plots of meta-analyses with binary outcomes. Although sample size or functions of sample size have often been used as the vertical axis, this is problematic because the precision of a treatment effect estimate is determined by both the sample size and by the number of events. Thus, studies with very different sample sizes may have the same standard error and precision and vice versa. Therefore, the shape of plots using sample size on the vertical axis is not predictable except that, in the absence of bias, it should be symmetric. After considering various possible choices of vertical axis, Sterne and Egger conclude that standard error of the treatment effect estimate is likely to be preferable in many situations. Funnel plots may also be drawn using precision (= 1/(standard error)) on the vertical axis using the `funnel2` command distributed as part of the `metaggr` package (Bradburn, Deeks, and Altman 1998). Such plots tend to emphasize differences between the largest study and the others.

2.2 Example

The trials of magnesium therapy following myocardial infarction (heart attack) are a well-known example in which the results of a meta-analysis, which appeared to provide clear evidence that magnesium therapy reduced mortality, were contradicted by subsequent larger trials that found no evidence that magnesium influenced mortality. Figure 2 is a funnel plot based on the results of 15 trials of the effect of magnesium on mortality following myocardial infarction. Because the smaller trials produced smaller odds ratios (more substantial reductions in mortality associated with magnesium therapy), the funnel plot is clearly asymmetric.

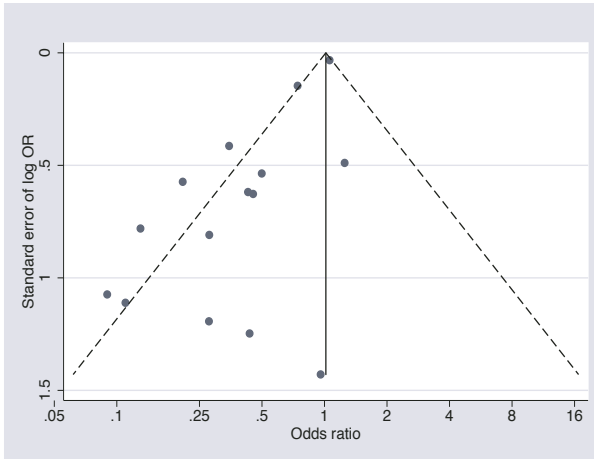

Figure 2: Funnel plot, using data from 15 trials of magnesium therapy following myocardial infarction.

The horizontal axis of figure 2 (treatment odds ratio) is drawn on a log scale, so that (for example) odds ratios of 2 and 0.5 are the same distance from the null value of 1 (no treatment effect). This is equivalent to plotting the log-odds ratio on the horizontal axis. The standard error of the log OR is plotted on the vertical axis. Note that the largest studies have the smallest standard errors, so to place the largest studies at the top of the graph, the vertical axis must be reversed (standard error 0 at the top).

The solid vertical line represents the summary estimate of the treatment effect, derived using fixed-effect meta-analysis. This is close to 1 because the estimated treatment odds ratios in the largest studies were close to 1. For the purposes of displaying the center of the plot in the absence of bias, calculation of the summary log-odds ratio using fixed rather than random-effects meta-analysis is preferable because the random-effects estimate gives greater relative weight to smaller studies and will, therefore, be more affected if publication bias is present (Poole and Greenland 1999).

Interpretation of funnel plots is facilitated by inclusion of diagonal lines representing the 95% confidence limits around the summary treatment effect, i.e., [summary effect estimate − (1.96 × standard error)] and [summary effect estimate + (1.96 × standard error)] for each standard error on the vertical axis. These show the expected distribution of studies in the absence of heterogeneity or of selection biases: in the absence of heterogeneity, 95% of the studies should lie within the funnel defined by these straight lines. Because these lines are not strict 95% limits, they are referred to as "pseudo 95% confidence limits".

2.3 Sources of funnel-plot asymmetry

Funnel plots were first proposed as a means of detecting a specific form of bias—publication bias. However as explained earlier (see the bottom graph of figure 1),

the exaggeration of treatment effects in small studies of low quality provides a plausible alternative mechanism for funnel-plot asymmetry. Egger et al. (1997) list different possible reasons for funnel-plot asymmetry, which are summarized in table 1.

Table 1: Potential sources of asymmetry in funnel plots

1. Selection biases
Publication bias
Location biases
Language bias
Citation bias
Multiple publication bias
2. True heterogeneity
Size of effect differs according to study size:
Intensity of intervention
Differences in underlying risk
3. Data irregularities
Poor methodological design of small studies
Inadequate analysis
Fraud
4. Artifact
Heterogeneity due to poor choice of effect measure
5. Chance

In addition to selective publication of studies according to their results, other possible biases affecting the selection of studies for inclusion in meta-analyses include the propensity for the results to affect the language of publication (Jüni et al. 2002); the possibility that results affect the frequency with which a study is cited and, hence, its probability of inclusion in a meta-analysis, and the multiple publication of studies with demonstrating an effect of the intervention (Tramer et al. 1997).

It is important to realize that funnel-plot asymmetry need not result from bias. The studies displayed in a funnel plot may not always estimate the same underlying effect of the same intervention, and such heterogeneity in results may lead to asymmetry in funnel plots if the true treatment effect is larger in the smaller studies. For example, if a combined outcome is considered, then substantial benefit may be seen only in subjects at high risk for the component of the combined outcome which is affected by the intervention (Davey Smith and Egger 1994; Glasziou and Irwig 1995). Some interventions may have been implemented less thoroughly in larger studies, thus explaining the more

positive results in smaller studies. For example, an asymmetrical funnel plot was found in a meta-analysis of trials examining the effect of inpatient comprehensive geriatric assessment programs on mortality. An experienced consultant geriatrician was more likely to be actively involved in the smaller trials and this may explain the larger treatment effects observed in these trials (Egger et al. 1997; Stuck et al. 1993).

The way in which data irregularities such as low methodological quality of smaller studies may result in funnel-plot asymmetry was described earlier. Poor choice of effect measure may also result in funnel-plot asymmetry; for example, it has been shown that meta-analyses in which intervention effects are measured as risk differences are more heterogeneous than those in which intervention effects are measured as risk ratios or odds ratios (Deeks and Altman 2001; Engels et al. 2000). The inappropriate use of risk differences may also result in funnel-plot asymmetry—if the effect of intervention is homogeneous on the risk ratio scale, then the risk difference will be smaller in studies that have low event rates.

2.4 Tests for funnel-plot asymmetry

It is, of course, possible that an asymmetrical funnel plot arises merely by the play of chance. Statistical tests for funnel-plot asymmetry have been proposed by Begg and Mazumdar (1994) and by Egger et al. (1997). These are available in the Stata command `metabias` (Steichen 1998; Steichen, Egger, and Sterne 1998). The test proposed by Egger et al. (1997) is algebraically identical to a test that there is no linear association between the treatment effect and its standard error and, hence, that there is no straight-line association in the funnel plot of treatment effect against its standard error (see Sterne, Gavaghan, and Egger [2000] for details). The corresponding fitted line may be added to the funnel plot using the `egger` option of the `metafunnel` command—see section 5 below.

2.5 Small-study effects

Funnel-plot asymmetry thus raises the possibility of bias, but it is not proof of bias. It is important to note, however, that asymmetry (unless produced by chance alone) will always lead us to question the interpretation of the overall estimate of effect when studies are combined in a meta-analysis; for example, if the study size predicts the treatment effect, what treatment effect will apply if the treatment is adopted in routine practice? Sterne, Egger, and Davey Smith (2001) and Sterne, Gavaghan, and Egger (2000) have suggested that the funnel plot should be seen as a generic means of examining "small-study effects" (the tendency for the smaller studies in a meta-analysis to show larger treatment effects) rather than as a tool to diagnose specific types of bias.

When funnel-plot asymmetry is found, its possible causes should be carefully considered. For example, how comprehensive was the literature search that located the trials included in the meta-analysis? Does reported trial quality differ between larger and smaller studies? Is there a plausible reason for the effect of intervention to be greater

in smaller trials? It is possible that differences between smaller and larger trials are accounted for by a trial characteristic; this may be investigated using the `by()` option of the `metafunnel` command, as described in section 6 below. Explanations for heterogeneity may be investigated more formally using meta-regression (Thompson and Sharp 1999) to investigate associations between study characteristics and intervention effect estimates. For example, we might investigate evidence that studies in which reported allocation concealment is unclear or inadequate tend to result in more beneficial treatment effect estimates. Meta-regression analyses may be done using the Stata command `metareg` (Sharp 1998); however, it will not necessarily be possible to provide a definitive explanation for funnel-plot asymmetry. In medical research, meta-analyses typically contain 10 or fewer trials (Sterne, Gavaghan, and Egger 2000). Power to detect associations between study characteristics and intervention effect estimates will therefore often be low, in which case it may not be possible to identify a particular study characteristic as the cause of the heterogeneity.

3 Syntax

`metafunnel` { *theta* { *se* | *var* } | *exp*(*theta*) { *ll ul* [*cl*] } } [`if` *exp*] [`in` *range*]
[, `by`(*by_var*) [`var` | `ci`] `nolines` `forcenull` `reverse` `eform` `egger`
graph_options]

4 Description

`metafunnel` plots funnel plots. The syntax for `metafunnel` is based on the same framework as for the `meta`, `metabias`, `metacum`, and `metatrim` commands. The user provides the effect estimate as *theta* (e.g., the log-odds ratio) and a measure of theta's variability (i.e., its standard error or its variance). Alternatively, the user provides *exp*(*theta*) (e.g., an odds ratio), its confidence interval, and, optionally, the confidence level.

5 Options

`by`(*by_var*) displays subgroups according to the value of *by_var*. The legend displays the value labels for the levels of *by_var* if these are present; otherwise, it displays the value of each level of *by_var*.

`var` and `ci` indicate the meaning of the input variables in the same way as for the other meta-analysis commands listed above. The help file for `meta` gives a full explanation.

`nolines` specifies that pseudo 95% confidence interval lines not be included in the plot. The default is to include them.

`forcenull` forces the vertical line at the center of the funnel to be plotted at the null treatment effect of zero (1 when the treatment effect is exponentiated). The default is for the line to be plotted at the value of the fixed-effect summary estimate.

`reverse` inverts the funnel plot so that larger studies are displayed at the bottom of the plot with smaller studies at the top. This may also be achieved by specifying `noreverse` as part of the `yscale(`*axis_description*`)` graphics option.

`eform` exponentiates the treatment effect *theta* and displays the horizontal axis (treatment effect) on a log scale. As discussed in section 2.2, this is useful for displaying ratio measures, such as odds ratios and risk ratios.

`egger` adds the fitted line corresponding to the regression test for funnel-plot asymmetry proposed by Egger et al. (1997) and implemented in `metabias` (see section 2.4). This option may not be combined with the `by()` option.

graph_options are any options allowed by the `twoway scatter` command that can be used to change the appearance of the points and add labels. If option `egger` is specified, the look of the fitted line can be changed using the options `clstyle`, `clpattern`, `clwidth`, and `clcolor` explained under *connect_options* in Stata's built-in help system and the graphics manual.

6 Examples

Listing the data for the 15 magnesium trials produces the following output:

```
. list trial trialnam year dead1 alive1 dead0 alive0, noobs
```

trial	trialnam	year	dead1	alive1	dead0	alive0
1	Morton	1984	1	39	2	34
2	Rasmussen	1986	9	126	23	112
3	Smith	1986	2	198	7	193
4	Abraham	1987	1	47	1	45
5	Feldstedt	1988	10	140	8	140
6	Schechter	1989	1	58	9	47
7	Ceremuzynski	1989	1	24	3	20
8	Singh	1990	6	70	11	64
9	Pereira	1990	1	26	7	20
10	Schechter 1	1991	2	87	12	68
11	Golf	1991	5	18	13	20
12	Thogersen	1991	4	126	8	114
13	LIMIT-2	1992	90	1069	118	1039
14	Schechter 2	1995	4	103	17	91
15	ISIS-4	1995	2216	26795	2103	26936

To use the `metafunnel` command, we first need to derive the treatment effect and its standard error for each trial. Here, we will express the treatment effects as log-odds ratios.

```
. generate or = (dead1/alive1)/(dead0/alive0)
. generate logor = log(or)
. generate selogor = sqrt((1/dead1)+(1/alive1)+(1/dead0)+(1/alive0))
```

A funnel plot can then be drawn using the following syntax, which includes the regression line corresponding to the regression test for funnel-plot asymmetry proposed by Egger et al. (1997):

```
. metafunnel logor selogor, xtitle(Log odds ratio) ytitle(Standard error of log OR)
> egger
```

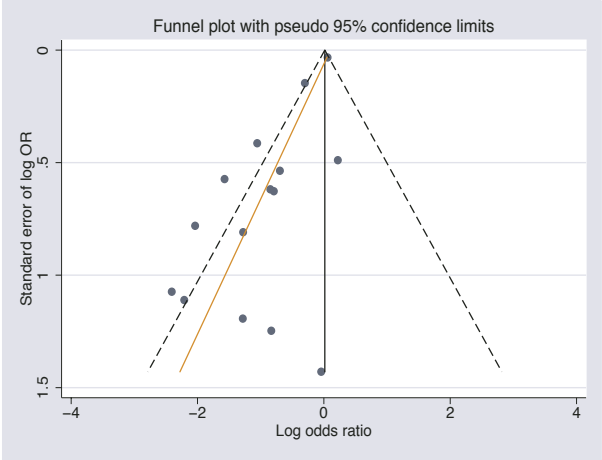

Figure 3: Funnel plot, using data from 15 trials of magnesium therapy following myocardial infarction, with log-odds ratios displayed on the horizontal axis.

By default, the subtitle "Funnel plot with pseudo 95% confidence limits" is displayed. ("Funnel plot" is displayed if the `nolines` options is specified.) This may be changed using the graphics option `subtitle(`*tinfo*`)`.

Note that the log-odds ratio and its standard error may be derived automatically using the `metan` command. (The latest version of this command may be installed by typing `ssc install metaaggr.pkg, replace` in the Stata Command window.) Typing

```
. metan dead1 alive1 dead0 alive0, or
```

produces a meta-analysis of the effect of magnesium and creates variables `_ES`, containing the odds ratio in each study, and `_selogES`, containing the standard error of the log-odds ratio. Thus, we may derive the log-odds ratio by typing

```
. genenerate log_ES = log(_ES)
```

The list output below shows that variables `log_ES` `_selogES` are identical to variables `logor` and `selogor` derived earlier.

```
. list trial trialnam year logor selogor _ES log_ES _selogES, noobs
```

trial	trialnam	year	logor	selogor	_ES	log_ES	_selogES
1	Morton	1984	-.8303483	1.247018	.4358974	-.8303483	1.247018
2	Rasmussen	1986	-1.056053	.4140706	.3478261	-1.056053	.4140706
3	Smith	1986	-1.27834	.8081392	.2784993	-1.27834	.8081392
4	Abraham	1987	-.0434851	1.42951	.9574468	-.0434851	1.42951
5	Feldstedt	1988	.2231435	.4891684	1.25	.2231435	.4891684
6	Schechter	1989	-2.40752	1.072208	.0900383	-2.40752	1.072208
7	Ceremuzynski	1989	-1.280934	1.193734	.2777778	-1.280934	1.193734
8	Singh	1990	-.695748	.5361776	.4987013	-.695748	.5361776
9	Pereira	1990	-2.208274	1.109648	.1098901	-2.208274	1.109648
10	Schechter 1	1991	-2.03816	.7807263	.1302682	-2.03816	.7807263
11	Golf	1991	-.8501509	.6184486	.4273504	-.8501509	.6184486
12	Thogersen	1991	-.7932307	.6258662	.452381	-.7932307	.6258662
13	LIMIT-2	1992	-.2993398	.1465729	.7413074	-.2993398	.1465729
14	Schechter 2	1995	-1.570789	.5740395	.2078812	-1.570789	.5740395
15	ISIS-4	1995	.0575872	.0316421	1.059278	.0575872	.0316421

The following command was used to produce figure 2 (see section 2.2), in which the horizontal axis is the treatment odds ratio, displayed on a log scale:

```
. metafunnel logor selogor, xlab(.05 .1 .25 .5 1 2 4 8 16)
> xscale(log) xtitle(Odds ratio) eform subtitle( )
> ytitle(Standard error of log OR)
```

When the `eform` option is used, the label of the horizontal axis (treatment effect, *theta*) is changed accordingly, unless there is a variable label for *theta* or the `xtitle(`*axis_title*`)` graphics option is used.

Finally, we will illustrate the use of the `by()` option by grouping the studies according to whether they were published during the 1980s or the 1990s:

```
. generate period = year
. recode period 1980/1989=1 1990/1999=2
(period: 15 changes made)
. label define periodlab 1 "1980s" 2 "1990s"
. label values period periodlab
. tab period
```

period	Freq.	Percent	Cum.
1980s	7	46.67	46.67
1990s	8	53.33	100.00
Total	15	100.00	

Using the latest version of the `metan` command (Bradburn, Deeks, and Altman 1998), we can examine the effect of magnesium separately, according to time period.

```
. metan dead1 alive1 dead0 alive0, or by(period) label(namevar=trialnam)
            Study        |    OR    [95% Conf. Interval]     % Weight
-----------------------+-----------------------------------------------
     1980s             |
Morton                 |   0.436     0.038     5.022          0.09
Rasmussen              |   0.348     0.154     0.783          0.99
Smith                  |   0.278     0.057     1.357          0.32
Abraham                |   0.957     0.058    15.773          0.05
Feldstedt              |   1.250     0.479     3.261          0.35
Schechter              |   0.090     0.011     0.736          0.42
Ceremuzynski           |   0.278     0.027     2.883          0.14
 Sub-total             |
  M-H pooled OR        |   0.437     0.267     0.714          2.36
-----------------------+-----------------------------------------------
     1990s             |
Singh                  |   0.499     0.174     1.426          0.47
Pereira                |   0.110     0.012     0.967          0.31
Schechter 1            |   0.130     0.028     0.602          0.57
Golf                   |   0.427     0.127     1.436          0.39
Thogersen              |   0.452     0.133     1.543          0.37
LIMIT-2                |   0.741     0.556     0.988          5.04
Schechter 2            |   0.208     0.067     0.640          0.75
ISIS-4                 |   1.059     0.996     1.127         89.74
 Sub-total             |
  M-H pooled OR        |   1.020     0.961     1.083         97.64
-----------------------+-----------------------------------------------
Overall                |
  M-H pooled OR        |   1.007     0.948     1.068        100.00
-----------------------+-----------------------------------------------
Test(s) of heterogeneity:
                Heterogeneity  degrees of
                  statistic     freedom      P    I-squared**
1980s                  7.85          6      0.250     23.5%
1990s                 30.27          7      0.000     76.9%
Overall               46.61         14      0.000     70.0%
Overall Test for heterogeneity between sub-groups :
                       8.50          1      0.004
** I-squared: the variation in OR attributable to heterogeneity

Significance test(s) of OR=1
1980s                   z=  3.31     p = 0.001
1990s                   z=  0.66     p = 0.511
Overall                 z=  0.22     p = 0.829
-----------------------------------------------------------------------
```

The `by()` option of the `metafunnel` command is used to display separate symbols for the two time periods; the resulting funnel plot is displayed in figure 4.

```
. metafunnel logor selogor, xlab(.05 .1 .25 .5 1 2 4 8 16)
> xscale(log) xtitle(Odds ratio) eform subtitle( )
> ytitle(Standard error of log OR) by(period)
```

As demonstrated by the analysis according to time period, the larger studies were published later. Perhaps more surprisingly, the asymmetry appears to result more from the studies published during the 1990s than from those published during the 1980s.

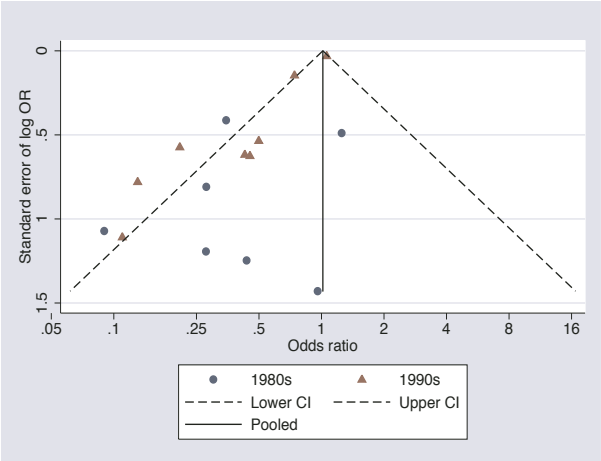

Figure 4: Funnel plot, using data from 15 trials of magnesium therapy following myocardial infarction, grouped according to date of publication.

7 Acknowledgments

Portions of the code for `metafunnel` were originally written by Thomas Steichen, who also gave helpful comments on an early version of the command. We are grateful to Nicholas J. Cox, who provided extensive programming advice.

8 References

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

Bradburn, M. J., J. J. Deeks, and D. G. Altman. 1998. sbe24: metan – an alternative meta-analysis command. *Stata Technical Bulletin* 44: 4–15. In *Stata Technical Bulletin Reprints*, vol. 8, 86–100. College Station, TX: Stata Press.

Davey Smith, G. and M. Egger. 1994. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. *British Medical Journal* 308(6921): 72–74.

Deeks, J. J. and D. G. Altman. 2001. Effect measures for meta-analysis of trials with binary outcomes. In *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2d ed., ed. M. Egger, G. Davey Smith, and D. G. Altman, 313–335. London: BMJ Publishing Group.

Egger, M., G. Davey Smith, and D. G. Altman. 2001. *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2d ed. London: BMJ Publishing Group.

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(7109): 629–634.

Egger, M., P. Jüni, C. Bartlett, F. Holenstein, and J. Sterne. 2003. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment* 7: 1–68.

Engels, E. A., C. H. Schmid, N. T. Terrin, I. Olkin, and J. Lau. 2000. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in Medicine* 19: 1707–1728.

Glasziou, P. P. and L. M. Irwig. 1995. An evidence based approach to individualizing treatment. *British Medical Journal* 311(7016): 1356–1359.

Jüni, P., F. Holenstein, J. A. C. Sterne, C. Bartlett, and M. Egger. 2002. Direction and impact of language bias in meta-analysis of controlled trials: empirical study. *International Journal of Epidemiology* 31: 115–123.

Poole, C. and S. Greenland. 1999. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 150(5): 469–475.

Schulz, K. F., I. Chalmers, R. J. Hayes, and D. G. Altman. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* 273(5): 408–412.

Sharp, S. 1998. sbe23: Meta-analysis regression. *Stata Technical Bulletin* 42: 16–22. In *Stata Technical Bulletin Reprints*, vol. 7, 148–155. College Station, TX: Stata Press.

Steichen, T. J. 1998. sbe19: Tests for publication bias in meta-analysis. *Stata Technical Bulletin* 41: 9–15. In *Stata Technical Bulletin Reprints*, vol. 7, 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. In *Stata Technical Bulletin Reprints*, vol. 8, 84–85. College Station, TX: Stata Press.

Stern, J. M. and R. J. Simes. 1997. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* 315(7109): 640–645.

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(10): 1046–1055.

Sterne, J. A. C., M. Egger, and G. Davey Smith. 2001. Investigating and dealing with publication and other bias. In *Systematic Reviews in Health Care: Meta-Analysis in Context*. 2d ed., ed. M. Egger, D. G. Altman, and G. Davey Smith, 189–208. London: BMJ Publishing Group.

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(11): 1119–1129.

Stuck, A. E., A. L. Siu, G. D. Wieland, J. Adams, and L. Z. Rubenstein. 1993. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 342: 1032–1036.

Thompson, S. G. and S. J. Sharp. 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 18: 2693–2708.

Tramer, M. R., D. J. Reynolds, R. A. Moore, and H. J. McQuay. 1997. Impact of covert duplicate publication on meta-analysis: a case study. *British Medical Journal* 315(7109): 635–640.

About the Authors

Jonathan Sterne is a Reader in Medical Statistics and Epidemiology in the Department of Social Medicine, 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.

Roger Harbord is a Research Associate in Medical Statistics in the Department of Social Medicine, University of Bristol, UK. His research interests include statistical methods for epidemiology and health services research, meta-analysis and systematic reviews, and genetic epidemiology.