Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis

Jonathan A. C. Sterne*, Matthias Egger

MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR UK

Received 03 August 2000; received in revised form 11 April 2001; accepted 12 April 2001

Abstract

Asymmetry in funnel plots may indicate publication bias in meta-analysis, but the shape of the plot in the absence of bias depends on the choice of axes. We evaluated standard error, precision (inverse of standard error), variance, inverse of variance, sample size and log sample size (vertical axis) and log odds ratio, log risk ratio and risk difference (horizontal axis). Standard error is likely to be the best choice for the vertical axis: the expected shape in the absence of bias corresponds to a symmetrical funnel, straight lines to indicate 95% confidence intervals can be included and the plot emphasises smaller studies which are more prone to bias. Precision or inverse of variance is useful when comparing meta-analyses of small trials with subsequent large trials. The use of sample size or log sample size is problematic because the expected shape of the plot in the absence of bias is unpredictable. We found similar evidence for asymmetry and between trial variation in a sample of 78 published meta-analyses whether odds ratios or risk ratios were used on the horizontal axis. Different conclusions were reached for risk differences and this was related to increased between-trial variation. We conclude that funnel plots of meta-analyses should generally use standard error as the measure of study size and ratio measures of treatment effect.

© 2001 Elsevier Science Inc. All rights reserved.

Keywords: Funnel plot; Meta-analysis; Randomized controlled trials; Selection bias; Publication bias; Statistical method; Systematic reviews

1. Introduction

Funnel plots—scatter plots in which the treatment effects estimated from individual studies on the horizontal axis are plotted against a measure of study precision on the vertical axis—have been proposed as a means of detecting publication bias in meta-analysis [1]. In the absence of bias the graph resembles a symmetrical inverted funnel because the treatment effect estimates from smaller studies scatter more widely at the bottom of the graph, with the spread narrowing with increasing precision among larger studies. If there is publication bias because smaller studies which show no statistically significant effects remain unpublished [2,3], then the funnel plot will appear asymmetrical [4,5]. Funnel plot asymmetry cannot, however, be interpreted as proof of publication bias in meta-analysis [6]. Asymmetry could also result from the overestimation of treatment effects in smaller studies of inadequate methodological quality [7]. Furthermore, heterogeneity of treatment effects will lead to funnel plot asymmetry if the true treatment effect is larger in the smaller trials [6,8]. For example, if a combined outcome is considered then substantial benefit may be seen only in patients at high risk for the component of the combined outcome which is affected by the intervention [9]. Trials conducted in high-risk patients will also tend to be smaller, because of the difficulty in recruiting such patients.

Funnel plots were first used in educational research and psychology [1], mainly for meta-analyses of continuous outcome variables in which standardized mean difference was plotted against sample size. In medical research vertical axes based on the standard error or variance of the treatment effect estimate have been increasingly used. A majority of trials in medicine have binary outcomes, and treatment effects are usually expressed as risk or odds ratios, although risk differences may also be used to measure treatment effects.

Meta-analysts thus face a wide array of choices for both vertical and horizontal axes in funnel plots. This leads to the danger that the funnel plot chosen for a particular meta-analysis may be that which best conveys the message desired by the investigator, or may not be appropriate for detecting bias [10]. The purpose of this article is to provide guidelines for the choice of axes in funnel plots of meta-analyses with binary outcomes.

* Corresponding author. Tel.: +44 117 928 7396; Fax: +44 117 928 7325. E-mail address: jonathan.sterne@bristol.ac.uk (J.A.C. Sterne).

0895-4356/01/$ – see front matter © 2001 Elsevier Science Inc. All rights reserved.
PII: S0895-4356(01)00377-8
2. Choice of vertical axis in funnel plots: case study

The randomized controlled trials of magnesium treatment in the prevention of death following myocardial infarction (Table 1) are a well known example where publication bias, demonstrated by an asymmetrical funnel plot [5,6], has been suggested as an explanation for the discrepancy between meta-analyses which showed a clear beneficial effect of magnesium therapy on mortality [11,12] and a subsequent large trial which showed no effect [13]. Fig. 1 shows funnel plots for these 16 trials, using six different choices of vertical axis. In each case, the horizontal axis is the log odds ratio, and the vertical line shows the summary log odds ratio calculated using fixed effects meta-analysis. 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 [14].

Fig. 1A uses standard error (SE) as the vertical axis. The largest studies have the smallest standard errors, so to place the largest trials at the top of the graph, the axis has to be inverted (standard error 0 at the top). The diagonal lines show the expected 95% confidence intervals around the summary estimate, i.e. [summary effect estimate − (1.96 × SE)] and [summary effect estimate + (1.96 × SE)] for each SE on the vertical axis. They indicate the extent of between-trial heterogeneity: in the absence of heterogeneity 95% of the trials should lie within the funnel defined by these straight lines.

Fig. 1B uses precision (defined as 1/SE) as the vertical axis. The study to investigate whether the prevalence of funnel plot asymmetry in published meta-analyses depends on the choice of treatment effect.

As described in detail elsewhere [8], we hand searched volumes 1993–97 of four general medicine and four specialist journals for meta-analyses based on at least five trials with binary endpoints which randomized individuals and reported the number of patients and events in each trial. We

Table 1
Mortality results from 16 trials of intravenous magnesium in acute myocardial infarction

<table>
<thead>
<tr>
<th>Study (first author)</th>
<th>Year</th>
<th>Deaths</th>
<th>Total no. of patients</th>
<th>Deaths</th>
<th>Total no. of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morton</td>
<td>1984</td>
<td>1</td>
<td>40</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>Rasmussen</td>
<td>1986</td>
<td>9</td>
<td>135</td>
<td>23</td>
<td>135</td>
</tr>
<tr>
<td>Smith</td>
<td>1986</td>
<td>2</td>
<td>200</td>
<td>7</td>
<td>200</td>
</tr>
<tr>
<td>Abrahorn</td>
<td>1987</td>
<td>1</td>
<td>48</td>
<td>1</td>
<td>46</td>
</tr>
<tr>
<td>Feldstedt</td>
<td>1988</td>
<td>10</td>
<td>150</td>
<td>8</td>
<td>148</td>
</tr>
<tr>
<td>Schechter</td>
<td>1989</td>
<td>1</td>
<td>59</td>
<td>9</td>
<td>56</td>
</tr>
<tr>
<td>Ceremuzynski</td>
<td>1989</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Bertschat</td>
<td>1989</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Singh</td>
<td>1990</td>
<td>6</td>
<td>76</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>Pereira</td>
<td>1990</td>
<td>1</td>
<td>27</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>Schechter &amp; Hod 1</td>
<td>1991</td>
<td>2</td>
<td>89</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>Golf</td>
<td>1991</td>
<td>5</td>
<td>23</td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Thögersen</td>
<td>1991</td>
<td>4</td>
<td>130</td>
<td>8</td>
<td>122</td>
</tr>
<tr>
<td>LIMIT-2</td>
<td>1992</td>
<td>90</td>
<td>1159</td>
<td>118</td>
<td>1157</td>
</tr>
<tr>
<td>Schechter &amp; Hod 2</td>
<td>1995</td>
<td>4</td>
<td>107</td>
<td>17</td>
<td>108</td>
</tr>
<tr>
<td>ISIS-4</td>
<td>1995</td>
<td>2216</td>
<td>29011</td>
<td>2103</td>
<td>29039</td>
</tr>
</tbody>
</table>

0.5 was added to all cells in the 2 × 2 table to allow calculation of the odds ratio for the Bertschart trial.
Bibliographic references are available from the authors on request.

3. Choice of horizontal axis in funnel plots: empirical study

The choice of treatment effect measure may affect the interpretation of randomized trials and meta-analyses [16]. In practice, most meta-analyses use ratio measures of treatment effect (odds ratio or relative risk), although risk differences are sometimes also used. We conducted an empirical study to investigate whether the prevalence of funnel plot asymmetry in published meta-analyses depends on the choice of treatment effect.

As described in detail elsewhere [8], we hand searched volumes 1993–97 of four general medicine and four specialist journals for meta-analyses based on at least five trials with binary endpoints which randomized individuals and reported the number of patients and events in each trial. We
found 78 meta-analyses, containing a total of 1145 trials, whose characteristics are shown in Table 2. When we performed standard chi-square tests of heterogeneity using the odds ratio, risk ratio and risk difference as outcome measures, the proportions of the 78 meta-analyses with heterogeneity P-value less than 0.05 were 0.29, 0.27 and 0.35, respectively.

For each meta-analysis, we investigated the evidence for funnel plot asymmetry by doing the analysis proposed by Egger et al. [6]. As explained in more detail elsewhere [8], this is algebraically identical to a regression (weighted by the inverse variance of the treatment effect) of treatment effect (log odds ratio, log risk ratio or risk difference) on its standard error:

\[
treatment\ effect = b_0 + b_1 \times SE\ of\ treatment\ effect
\]

Evidence for funnel plot asymmetry is measured by the estimated “asymmetry coefficient” \( b_1 \); if \( b_1 = 0 \) then the funnel plot is symmetric. Because of the low power of tests for funnel plot asymmetry [8], and consistent with previous work [6], we used \( P < 0.1 \) in testing for asymmetry. We have shown elsewhere [8] that this regression test for funnel plot asymmetry is more powerful than the rank correlation method proposed by Begg and Mazumdar [17]. Analyses were done using the statistical package Stata (Stata Corporation, College Station, TX). Both the regression and rank correlation tests may be done using the Stata command “metabias” [18], which also draws funnel plots.

There was evidence for funnel plot asymmetry for one or more of the treatment effect measures in 32 of the 78 meta-analyses. The number of meta-analyses with evidence of asymmetry was 21 (27%), 22 (28%) and 17 (22%) for log odds ratio, log risk ratio and risk difference, respectively.

Fig. 4 is a Venn diagram showing the extent of agreement on the evidence for asymmetry between the three treatment effects. While there was substantial concordance between log odds ratio and log risk ratio, use of risk difference iden-
tified asymmetry in eight trials not identified using the other treatment effect measures.

4. Discussion

The potential for bias in the location, selection or conduct of the component studies included in meta-analyses is increasingly recognized. Funnel plots are a useful graphic means of checking whether “small study effects”—a tendency for treatment effect estimates in small studies to differ from those in larger studies—may have distorted the results of a meta-analysis [19]. This could be due to publication bias, other reporting biases, low methodological quality of smaller studies or true heterogeneity in treatment effects [20]. However, there is confusion over how funnel plots should be constructed, and the choice of method may affect the shape of plots [10]. We addressed this issue in a case study and an empirical investigation of published meta-analyses.

4.1. Choice of vertical axis

Clearly, the choice of variables used in funnel plots should be consistent with the fundamental assumption underlying the method: that in the absence of bias a plot from trials which estimate the same treatment effect will be symmetrical and bear some resemblance to a funnel. This is the case for standard error and variance (and functions thereof) but not necessarily for sample size or functions of sample size. Further, the major factor affecting the probability of publication of a study is its P-value [2,3,21–23], which is a function of standard error or variance rather than of sample size.

The power of a study, and the precision of the estimate of a treatment’s effect, is determined both by the total sample size and by the number of participants developing the event of interest. A study with 100,000 patients and 10 events is less powerful and will produce less precise estimates than a study with 1000 patients and 100 events. In other words, studies with very different sample sizes may have the same standard error and precision, and vice versa. In the absence
of bias the shape of plots using sample size on the vertical axis will therefore not necessarily correspond to a funnel. A predictable shape in the absence of bias is required in order to use any plot to assess evidence for bias. For the same reason it is not possible to calculate expected 95% confidence limits for funnel plots of sample size. This is illustrated in Fig. 5, which shows funnel plots of the log odds ratio of end-stage renal failure against standard error and sample size for a meta-analysis of 10 trials of ACE inhibitors in non-diabetic renal disease [24]. The event rate was low in the largest trial [25] so that the estimate of the treatment effect was less precise than the estimates from all except one of the smaller trials. Sample size and functions of sample size are therefore not a good choice for the vertical axis in funnel plots.

Standard error, precision (the inverse of standard error), variance and the inverse of variance are four valid choices for the vertical axis. However, there are important differences in the shape of plots. Our case study indicates that the use of standard error is likely to be preferable in many situations. Interpretation of funnel plots is facilitated by inclusion of lines representing the 95% confidence limits around the summary treatment effect which show the expected distribution of studies in the absence of heterogeneity or of selection biases. Such lines can be drawn for all four variables; however, lines will be curved and, in the case of precision and inverse of variance, smaller studies will be compressed at the bottom of the graph. This makes the visual assessment of asymmetry more difficult and gives emphasis to the larger studies which on average are less prone to bias. Precision or the inverse of variance might be preferable in studies which focus on a comparison of meta-analyses of small trials with subsequent large trials [26]. Finally, funnel plots of standard error correspond to the regression test to detect asymmetry [6] and therefore allow the identification of trials which may be influential in the regression analysis.

Tang and Liu [10] examined the evidence for funnel plot asymmetry in a sample of 198 meta-analyses published in the Cochrane Library using either precision or sample size on the vertical axis. They found that the shape of a funnel plot depends on the choice of the method to construct the plot but did not note that plots against sample size will not necessarily have a funnel shape. Tang and Liu used both the regression test proposed by Egger et al. [6] and a regression method of their own in which the covariate is the square root of the total sample size (rather than the standard error) and often obtained discrepant results depending on which
statistical method was used. This is not surprising: the regression test proposed by Egger et al. [6] is based on the assumption that the process of selection for publication is driven by the statistical significance of the trials’ results, which will lead to an association between treatment effect and standard error. A recent study compared the power of different methods to detect publication bias, using simulations in which a study’s probability of inclusion in a meta-analysis was based on its P-value [27]. This study concluded that the regression test proposed by Egger et al. [6] had greater power than a regression against total sample size, although concerns were raised about the Type I error rate (see also the discussion in Sterne et al. [8]).

Copas [28] performed sensitivity analyses in which the value of the estimated treatment effect is computed under a range of assumptions about the severity of the selection bias and showed that these can lead to an approximately linear association between treatment effect and its standard error [29]. There is also a considerable body of empirical evidence that statistically significant results are more likely to
be published than non-significant ones: this includes studies comparing published trials with trials included in trial registries [30,31], investigations into the full publication of conference abstracts [32,33] and follow-up studies of proposals submitted to ethics committees [2,3,21,23] and of research funded by the National Institutes of Health [34].

The Tang and Liu approach [10] appears to postulate that it is the square root of the total sample size of a trial which is most closely related to the probability of publication; an assumption which is supported by little empirical evidence. Indeed, a recent synthesis of the available data found that the sample size of a study was only weakly associated with the probability of publication whereas a consistent association was found with statistically significant results [22]. We
acknowledge, however, that in some situations other factors, such as the sample size, may be more strongly related to the probability of publication than the level of statistical significance.

4.2. Choice of horizontal axis

We found that tests for funnel plot asymmetry reached different conclusions when treatment effects were expressed as risk differences than when ratio measures were used. Among the considerations when choosing the effect measure in meta-analyses of trials with binary outcomes are the measure’s consistency across trials and its ease of interpretation [35]. The less heterogeneous the statistic, the greater the justification for statistically combining the trials.

In our sample of published meta-analyses we found that treatment effects measured using risk differences showed more between-trial heterogeneity than log odds ratio or log risk ratio. This is consistent with the findings of Deeks and Altman [35,36] in a study of almost 2000 meta-analyses of binary outcomes published on the Cochrane Library. Engels et al. [37] also found that more meta-analyses had heterogeneous risk differences than odds ratios, in a sample of 125 meta-analyses published between 1990 and 1996. They pointed out that use of risk differences can lead to trials with very few events in either treatment or control groups being given more weight than trials with many events [37]. This, and the greater heterogeneity of the risk difference between trials, will have contributed to the discrepant results of tests for funnel plot asymmetry in our study of 78 meta-analyses.

Fewer funnel plots were classified as asymmetric using risk differences (22%) than using odds ratios (27%) or risk ratios (28%): increased between-trial heterogeneity will reduce power to detect funnel plot asymmetry.

Causes of between-trial heterogeneity may also lead to funnel plot asymmetry, if they are associated with study size. For example, if risk ratios were identical across trials then risk differences would by definition be larger in trials in which event rates were larger (i.e., trials in high-risk groups). In some meta-analyses in high-risk patients are smaller because high-risk patients are less common. This would lead to funnel plot asymmetry, not caused by bias, in funnel plots using risk differences. Given that there tends to be more heterogeneity using risk differences, these will not usually be the best choice of horizontal axis in funnel plots. Independent of the reason for funnel plot asymmetry, its presence calls into question the wisdom of statistically combining studies in meta-analysis.

Conclusions from tests of funnel plot asymmetry were similar whether log odds ratio or log risk ratio was chosen. This could be expected as odds ratio and risk ratio give similar results unless event rates are high. Indeed, discrepancies between these measures arose only when the event rate was > 30% in one or more trials. In this situation the odds ratio scale is preferable because it is not constrained (for example, a risk of 50% cannot increase by more than a factor of 2). Furthermore, funnel plots based on the odds ratio have the same shape whether the outcome is defined as occurrence or non-occurrence of the event of interest (for example, smoking cessation or continued smoking). In contrast,

| Possible choices of axis in funnel plots: advantages, disadvantages and recommendations |
|---|---|---|
| **Axis / measure** | **Advantages and disadvantages** | **Recommendations** |
| **Vertical axis** |  |  |
| Standard error | Funnel shape with straight 95% confidence lines. Emphasis of the plot is on smaller studies where bias is more likely. Axis has to be inverted to place the largest trials at the top of the graph. | The best choice in most cases |
| Precision (1/SE) | Plot is not funnel shaped; 95% confidence lines are curved. Emphasis of the plot is on larger studies; smaller studies are compressed at the bottom. | An option in studies which focus on a comparison of meta-analyses of small trials with subsequent large trials |
| Variance (SE²) | Plot is not funnel shaped; 95% confidence lines are curved. Emphasis of the plot is on smaller studies where bias is more likely. | Not recommended |
| Inverse variance (1/variance) | Plot is not funnel shaped; 95% confidence lines are curved. Emphasis of plot is on larger studies, smaller studies are compressed at the bottom. | An option in studies which focus on a comparison of meta-analyses of small trials with subsequent large trials, but precision would usually be better. |
| Sample size or log sample size | Expected shape of plot in absence of bias is unpredictable. | Invalid choice |
| **Horizontal axis** |  |  |
| Log odds ratio | Scale is not constrained and plots have the same shape whether the outcome is defined as occurrence or non-occurrence of the disease. Odds ratios may be misinterpreted as risk ratios. | The best choice in most cases |
| Log risk ratio | Readily understood measure. Scale is naturally constrained so that heterogeneity may be introduced if the event rate is high. | Valid choice in many cases but not recommended if the event rate is high. Can give different conclusions depending on outcome definition. |
| Risk difference | Readily understood measure. Often associated with increased heterogeneity which may result in funnel plot asymmetry which is not apparent when ratio measures are used. | Not recommended in most cases |
the shape of funnel plots based on risk ratios will differ depending on the definition of the outcome as occurrence or non-occurrence. This is illustrated in Fig. 6 for a meta-analysis of trials of the effect of ACE inhibitors on mortality in heart failure [38]. Most trials had control group mortality rates below 10% but there were also trials with higher rates, including the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS) [39] with a control group mortality rate above 50%. When the outcome is mortality and treatment effects are measured using risk ratios (upper plot), the funnel plot seems moderately asymmetrical and the P-value from the regression test is 0.04. When the outcome is survival and treatment effects are again measured using risk ratios (middle plot), the plot is more clearly asymmetrical, as is the evidence from the regression test (P = 0.006). A funnel plot using odds ratios (lower plot) looks similar to the plot using risk ratios and with mortality as the outcome, but the shape of this plot would be unchanged if the outcome were survival (the plot would simply be the mirror image, transposed around 0 on the y-axis). In this meta-analysis the most powerful studies were those which had the highest mortality rates, and so the extent to which the risk ratio is constrained is related to the position of the study in the funnel plot.

Our recommendations for the choice of axis in funnel plots are summarized in Table 3. Although our understanding of the selection process leading to publication or non-publication of research is imperfect, standard error is likely to be the best choice of the vertical axis in funnel plots, unless the aim is to emphasize the difference between the largest trials and all others, when precision is more appropriate. A log ratio measure, preferably the log odds ratio, should generally be used for the horizontal axis. Funnel plots are a useful graphic method but their interpretation is subjective. Standardizing the choice of axes should contribute to making them a more reliable tool in meta-analytic research.

References

[23] Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281–6.
[31] Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 1993;341:418–22.
[34] Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clini Trials 1993;Doc No 50.


