

# Chapter 15

## Meta-analysis II

### Interpretation and Use of Outputs

Adedayo A. Onitilo, Suhail A.R. Doi, and Jan J. Barendregt

**Abstract** Outputs in meta-analysis give us measures of evidence dissemination bias or graphical representation of the pooled results and their underlying heterogeneity. This chapter discusses the various outputs with a focus on their utility and interpretation. Examples focus on the use of MetaXL, which is our own software developed for meta-analysis and is freely available from [www.epigear.com](http://www.epigear.com). This is the only software currently available that can perform a bias-adjusted meta-analysis.

#### Introduction

The main output of a meta-analysis is the pooled estimate and its confidence interval. In addition, there are also a number of graphical and numerical outputs that aid with interpretation of results by presenting information such as study heterogeneity, detection of publication bias, and other important aspects of the meta-analysis. Graphical and statistical representation should not replace, but should be used in addition to narrative description of study design, setting, methods, follow-up analysis methods as well as the strengths and limitations of the individual studies pooled together in the meta-analysis. MetaXL is our

---

A.A. Onitilo  
Marshfield Clinic - Weston Center, Weston, USA

University of Queensland, School of Population Health, Brisbane, QLD, Australia  
University of Wisconsin, School of Medicine and Public Health, Madison, USA

S.A.R. Doi (✉)  
University of Queensland, School of Population Health, Brisbane, QLD, Australia  
Princess Alexandra Hospital, Brisbane, Australia  
e-mail: [sardoi@gmx.net](mailto:sardoi@gmx.net)

J.J. Barendregt  
University of Queensland, School of Population Health, Brisbane, QLD, Australia

preferred software for meta-analysis (downloadable freely from [www.epigear.com](http://www.epigear.com)) and all outputs we discuss use MetaXL as far as possible. In addition, bias-adjusted meta-analyses can only be run using MetaXL.

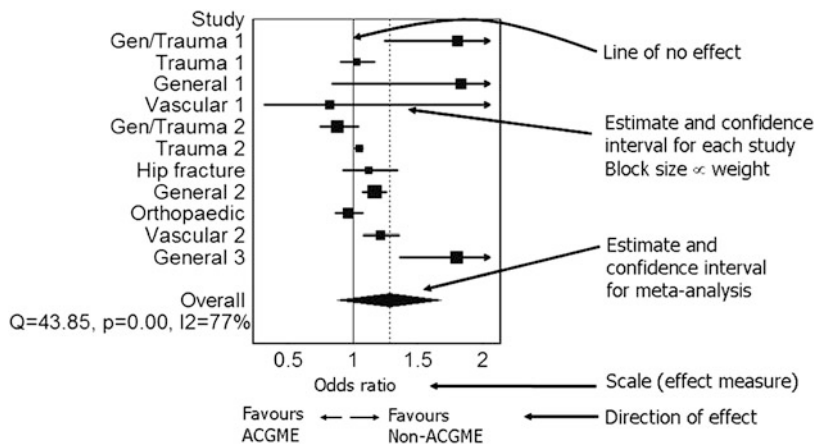
## Individual and Pooled Results

### *Forest Plot*

The forest plot, a graphical presentation of meta-analysis results, first used in 1982 by Lewis and Ellis, has now become standard practice and is arguably the most important output from a meta-analysis. A forest plot presents individual study estimates, the pooled estimate, confidence intervals, as well as the weight of each study in the analysis and heterogeneity statistics. Individual studies used in the meta-analysis are represented in the plot by horizontal lines; the length of each line represents the confidence interval around each study estimate. Shorter lines represent a narrower confidence interval thus higher precision of study effect size (ES), usually found in larger studies. Conversely, longer lines represent a wider confidence interval and less precision around effect side, usually found in smaller studies. The point estimate from each study is represented by a shape on the line such as a dot or box, and the size of this shape represents the weight of the study in the meta-analysis. A summary estimate of the point estimates is also represented by a shape at the end of the graph. Most forest plots also have two vertical lines. A dotted vertical line represents the pooled estimate and a solid vertical line represents the null estimate. For example, for the odds ratio this is 1 and for the mean difference the null has a value of 0. Horizontal lines that cross the null vertical line represent non-significant studies. Most forest plots in meta-analysis will arrange studies in chronologic order or by subgroups. This allows for further subgroup analysis or stratification. It can also be a way to represent heterogeneity in the meta-analysis. The plot can either be on a normal scale or logarithmic scale. The normal scale is usually used for mean difference and rates, while logarithm scales are used for ratios.

Figure 15.1 presents the forest plot from a quality effects model analysis of patient mortality before and after changes to the working hour regulations for surgeons. The square on the plot for individual studies is proportional to the weight it has in the meta-analysis; the horizontal lines represent the study's confidence interval. The dotted vertical line on the right gives the pooled estimate, the solid vertical line on the left is the null result, in this case  $OR = 0$ . Inspection of the forest plot can give a good indication of the amount of heterogeneity. In MetaXL, the forest plot is obtained by choosing Results from the MetaXL menu.

Forest plots are easy to read and interpret, although one drawback is that attention is often drawn to the least precise study which has the longest horizontal line and actually carries less weight in the meta-analysis. While the forest plots



**Fig. 15.1** The forest plot from a quality effects model analysis of mortality after (compared with before) the ACGME (Accreditation Council for Graduate Medical Education) regulations that reduced working hours for surgeons. The *dotted vertical line* indicates the pooled effect size.  $Q$  is the Cochran  $Q$  statistic for heterogeneity and  $I^2$  is the  $I^2$  statistic and both are discussed below

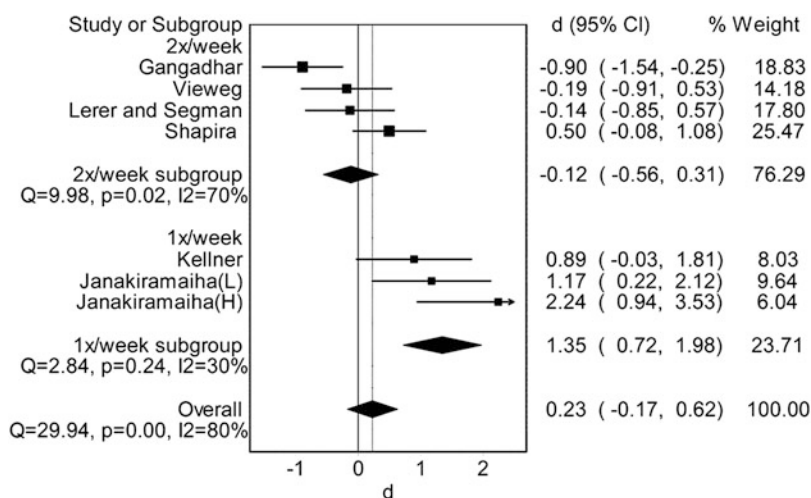
should be considered whenever feasible and appropriate in reporting a meta-analysis it may not be the most appropriate representation of meta-analysis that involves too many studies, in this case a summary forest plot or other plots such as the Galbraith plot should be considered. In the summary forest plot, individual study results are replaced with pooled results from either different outcomes or different subgroups. Thus individual points represent meta-analyses rather than studies.

### Sensitivity Analysis

Sensitivity analysis explores the ways in which the main findings are changed by varying the selection criteria for studies that are combined. The sensitivity analysis is executed by running the meta-analysis across categories of selected studies; for example, published versus unpublished studies or other selection criteria based on patient group, type of intervention or setting. A meta-analysis can also be performed by leaving out one study at a time to see if any single study has a large influence on the pooled results. Sensitivity analysis can also be done by running different meta-analysis models to examine the robustness of the method utilized in the meta-analysis. Usually if there is no significant heterogeneity in the studies used, most methods should yield comparable summary estimates. When dose–response or open-ended variables are examined in the meta-analysis, a sensitivity analysis can limit the range in the dose–response or open-ended variable that produce most of the effect. In meta-analyses without sensitivity analyses, the likely

**Table 15.1** Leave-one-out sensitivity analysis results (Charlson et al. 2012)

Study excluded	Meta-analysis results with study excluded			Weight (%) of the study in the complete analysis
	ES	Lower 95 % confidence limit	Higher 95 % confidence limit	
Gangadhar	-0.897	-1.540	-0.255	18.8
Vieweg	-0.190	-0.907	0.527	14.2
Lerer and Segman	-0.138	-0.846	0.570	17.8
Shapira	0.497	-0.084	1.078	25.5
Kellner	0.892	-0.027	1.811	8.0
Janakiramaiha (L)	1.169	0.220	2.117	9.6
Janakiramaiha (H)	2.236	0.944	3.528	6.0



**Fig. 15.2** Forest plot comparing 3x/week ECT with 2x/week (*upper* subgroup) or 1x/week (*lower* subgroup). It is evident that the 1x/week frequency results in a greater difference from 3x/week compared with 2x/week (Data from Charlson et al. 2012). Heterogeneity is diminished within subgroups suggesting that ECT frequency contributes to overall heterogeneity

impact of these important factors on the key finding is ignored and thus the results are less robust. An example of a leave-one-out sensitivity analysis is given in Table 15.1 and also depicted in the forest plot for 2x/week versus 1x/week comparisons after electroconvulsive therapy (ECT) for depression (Fig. 15.2).

## Heterogeneity

One of the most important aspects of meta-analysis is to determine whether heterogeneity exists in the studies combined in the analysis and investigate the source of such heterogeneity. It underscores the use of meta-analysis as a means of

generating a summary estimate, lends conclusiveness to otherwise inconclusive clinical situations and extends meta-analysis to explain differences between the combined studies. Heterogeneity can be clinical or statistical. Clinical heterogeneity is based on the characteristics of studies combined (e.g., study design, follow-up length, duration of therapy) and characteristics of subjects in the studies. Thus clinical heterogeneity can be within or between studies. Statistical heterogeneity refers to situations when the estimates from different studies deviate considerably from each other. Below we describe some of the formal statistical tests and plots for assessing heterogeneity.

### *Cochran's Q*

Cochran's  $Q$  is a heterogeneity statistic. It is the classical measure of heterogeneity and is given by

$$Q = \sum_i w_i (\theta_i - \theta_p)^2$$

where  $i$  is an index for the study,  $w_i$  is the fixed effect weight of study  $i$ ,  $\theta_i$  is the estimate from study  $i$ , and  $\theta_p$  is the fixed effect pooled estimate. The  $Q$  statistic follows a chi-squared distribution with  $k - 1$  degrees of freedom under the null hypothesis of homogeneity, where  $k$  is the number of studies in the meta-analysis. If the probability of the value of  $Q$  occurring by chance is low ( $p < 0.05$ ), the null hypothesis is rejected and heterogeneity is assumed. Unfortunately, the  $Q$  statistic is not very sensitive when the number of studies is not large. In that case, some authors prefer a critical value for  $p$  of 0.1 instead of 0.05. When the number of studies is large, the  $Q$  statistic becomes too sensitive. In MetaXL the MACochranQ function returns the  $Q$  statistic in the spreadsheet. The test can then be performed using Excel's CHIDIST function. The  $Q$  statistic and its test result are also presented in the forest plot and tabular output.

The magnitude of the computed  $Q$  is dependent on the weight and the number of studies in the meta-analysis. If there are limited number of small studies (<20 studies), it has been shown that the asymptotic  $Q$  statistic gives the correct type I error under the null hypothesis but has low power (Takkouche et al. 1999) and null for heterogeneity is not likely to be rejected. Whereas if there are large number of studies or large sample size studies in the meta-analysis, irrespective of true clinical heterogeneity  $Q$  has too much power and null for heterogeneity is likely to be rejected. For this reason, it is always important to examine the studies in the meta-analysis for clinical heterogeneity.

## $I^2$

The  $I^2$  statistic is another means to detect heterogeneity, and is derived from the  $Q$  statistic. The  $I^2$  examines the percentages of variation across studies due to heterogeneity rather than by chance and it is given by

$$I^2 = \begin{cases} 100 \frac{Q-df}{Q} & \text{if } > 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $df = k - 1$  is degrees of freedom. Confidence intervals for  $I^2$  can be derived as follows:

Define  $H = \sqrt{Q/(k - 1)}$ . Then,

$$SE[\ln(H)] = \begin{cases} \frac{1}{2} \frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q - \sqrt{2k-3}}} & \text{if } Q > k \\ \sqrt{\left\{ \frac{1}{2(k-2)} \left( 1 - \frac{1}{3(k-2)^2} \right) \right\}} & \text{otherwise} \end{cases}$$

95 % confidence intervals for  $H$  are then derived by

$$\exp(\ln H \pm 1.96SE[\ln(H)])$$

Since

$$I^2 = \frac{H^2 - 1}{H^2}$$

the confidence intervals for  $I^2$  are derived from those of  $H$ . The  $I^2$  statistic is thus a number between 0 and 100. A rule of thumb is that heterogeneity is low for an  $I^2$  of 25, moderate for an  $I^2$  of 50, and high for an  $I^2$  of 75. In MetaXL the MAISquare function returns the  $I^2$  statistic in the spreadsheet. It is also presented in the forest plot and tabular output; the latter includes the confidence interval. Effectively,  $I^2$  is  $(Q - (k - 1))$  divided by  $Q$  where  $k$  denotes the number of studies.  $I^2$  has the same problem of low statistical power with small numbers of studies. Specifically, the confidence intervals around  $I^2$  behave very similarly to tests of  $Q$  in terms of type I error and statistical power. Also,  $I^2$  increases with the number of subjects included in the studies in a meta-analysis. It thus seems counterintuitive to criticize  $Q$  as having low power on the one hand and to define a measure (and an assessment rule) that would require the heterogeneity test to be even more significant. From the point of view of validity, power and computational ease, the  $Q$  statistic is probably a better choice compared with  $I^2$ . Unlike the  $Q$  statistic, the  $I^2$  statistic does not vary based on the number of studies included in the meta-analysis, it is possible to compare the statistical heterogeneity of meta-analyses with different numbers of studies. However,  $I^2$  will tend to increase artificially as evidence accumulates since

it increases with number of subjects included in the meta-analysis. Additionally, as  $I^2$  is the percentage of variability that is due to between-study heterogeneity,  $1 - I^2$  is the percentage of variability that is due to sampling error. When the studies become very large, the sampling error tends to 0 and  $I^2$  tends to 1 (Rucker et al. 2008). Such heterogeneity may not be clinically relevant and studies with relatively large  $I^2$  in this situation may still be usefully pooled if other measures such as  $\tau^2$  remain relatively small and clinically relevant heterogeneity is unlikely to be present.

## $\tau^2$

Yet another statistic is  $\tau^2$ , which is the random effects variance component calculated as part of a random effects meta-analysis. The  $\tau^2$  statistic examines the between-study variance and is given by

$$\tau^2 = \frac{Q - (k - 1)}{\sum w_i - \left( \frac{\sum w_i^2}{\sum w_i} \right)}$$

which is set to zero if  $Q < k - 1$ , and  $w_i$  is the inverse variance weight. The  $\tau^2$  statistic is the variance of the presumed normally distributed individual study estimates under the assumptions of the random effects model.

In MetaXL the MATauSquare function returns the  $\tau^2$  statistic in the spreadsheet, and it is also presented in the tabular output of random effects analyses. It may also be used as a marker of heterogeneity if its value is greater than zero. Similar to the  $Q$  and  $I^2$  statistic the  $\tau^2$  statistic has its limitation; it is not very powerful if the number of studies is small or if the conditional variances between the studies are large. The advantage, however, is that it does not depend on the number or size of studies in the meta-analysis, i.e., it can be kept fixed with increasing subjects in the meta-analysis. Furthermore, since  $\tau^2$  is measured on the same scale as the outcome, it can therefore be directly used to quantify variability. Note that assessment of  $\tau^2$  does not give us a  $p$  value but rather a yes/no answer only, and certainly there will be little heterogeneity if  $\tau^2 = 0$  regardless of the value of  $I^2$ . We must keep in mind however that  $\tau^2$  assumes normality of the random effects and the error terms.

## *Q Index*

The  $Q$  index is applicable to the quality effects model only. It expresses the percentage of study weight that is re-distributed in the quality effects analysis. It is given by

$$Q_{\text{index}} = 100 \left( \frac{\sum_i w_i (1 - q_i)}{\sum_j w_j} \right)$$

where  $q_i$  is the quality score of study  $i$  and  $w_i$  is the inverse variance weight.

In MetaXL the MAQIndex function returns the Q index statistic in the spreadsheet, and it is also presented in the tabular output of quality effects analyses. The Q index is the only measure that inputs study quality as a source of heterogeneity. It therefore has the advantage of imputing clinical heterogeneity in statistical terms, a strength not seen in any other statistical test for heterogeneity.

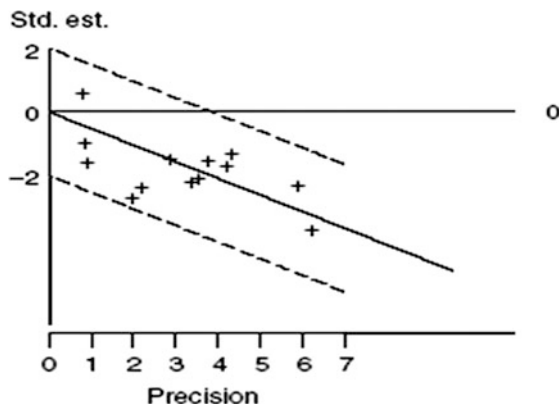
### ***Galbraith Plot***

The Galbraith plot (Fig. 15.3) presents standardized effect estimate on the vertical axis plotted against the inverse of the standard error on the horizontal axis. It is a linear regression constrained through the origin of the standardized treatment effects (treatment effect divided by its standard error) on their inverse standard errors which yields a regression line. Typically a dotted line is used at  $\pm 2SD$  confidence interval above and below this line. The slope of the regression line provides details of the unstandardized effect estimates. Galbraith plots facilitate examination of heterogeneity, including detection of outliers. With a fixed effect model, 95 % of studies in a meta-analysis should be found on this plot to be within the two confidence interval lines and the more precise studies are farthest from the origin of the linear regression line. Different symbols can be used in the plots to represent sub-sets or stratification thus making identification of the source of heterogeneity easier. Also the graph can be labelled to show the direction the effect the estimate favors. Compared to the forest plot, the Galbraith plot is able to display more studies that cannot be easily done by the forest plot and it also has the additional advantage that it gives a better representation of heterogeneity.

### ***L'Abbé Plot***

The L'Abbé plot is used to present the results of multiple clinical trials with dichotomous outcomes showing for each study; the observed event rate in the experimental group plotted against the observed event rates in the control group. It is used to view the range of event rates among the trials and highlight excessive heterogeneity. The L'Abbé plot is also ideally suited to diagnostic meta-analyses (Fig. 15.4) where diseased (group 1) and healthy (group 2) rates of test positivity can be compared across studies. The shape representing each study is usually proportional to the size of each study (or study weights) since unlike the forest





**Fig. 15.3** A Galbraith Plot with standardized effect estimate on the vertical axis plotted against the inverse of the standard error as a measure of precision on the horizontal axis. The intercept is constrained to zero. *Solid lines* represent the unweighted regression line constrained at 0 with a slope equal to the overall effect size of a fixed effects meta-analysis, and its 95 % confidence intervals (*dashed line*). The position of the studies on the y-axis indicates their contribution to the Q statistic for heterogeneity. The position of the studies on the x-axis indicates the weight of each study in the meta-analysis

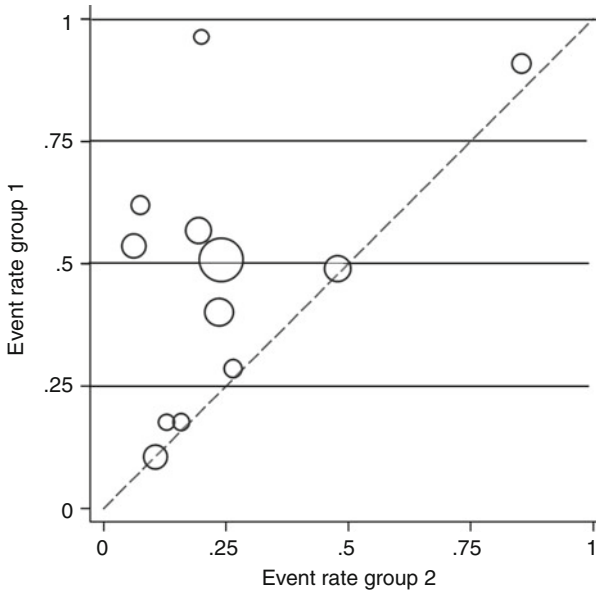
plot or Galbraith plot there is no information about the precision of the studies on the plotted axes. The L'Abbé plot should be considered when outcomes are dichotomous across studies (treatment vs. control) or for diagnostic studies (sensitivity vs. false-positive rates).

## Publication Bias

Publication bias refers to the phenomenon whereby studies with significant outcomes are more likely to be submitted for publication compared to null result or non-significant studies. This is usually assessed by several statistical and graphical (quasi-statistical) means.

### *Funnel Plot*

Funnel plots assess publication bias or heterogeneity by plotting the trials' effect estimates against a measure of precision. Asymmetrical plots are interpreted to suggest that selection biases are present. The use of such a plot is based on the fact that precision in estimating the underlying treatment effect will increase as the precision of the study increases and thus results from small studies will scatter widely at the bottom of the plot, with the spread narrowing with increasing

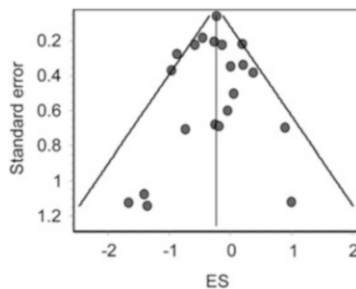


**Fig. 15.4** L'Abbé plot demonstrating true-positive (group 1) and false-positive (group 2) rates in diseased and healthy subjects, respectively, in a diagnostic meta-analysis (Data used from Whiting et al. 2005)

precision. In the absence of selection bias, the plot is expected to resemble a symmetrical inverted funnel. It is usually recommended that ratio measures of intervention effect should be plotted on a logarithmic scale, so that effects of the same magnitude but opposite directions (e.g., odds ratios of 0.5 and 2) are equidistant from 1.0.

Figure 15.5 shows the funnel plot from the meta-analysis of fibrinolysis in myocardial infarction study. The vertical line represents the pooled estimate from the inverse variance model; the funnel sides represent the 95 % confidence intervals around the pooled estimate, given the standard error on the y-axis; and the dots represent the individual study results. The aim of the funnel plot is to examine publication bias. When the study dots are largely symmetrical around the pooled estimate, there is no evidence for publication bias. In the present case, there is a large degree of asymmetry, which suggests publication bias is present. Funnel plots can look quite different, depending on the choice of y-axis. MetaXL offers three options: inverted standard error (default, and used in Fig. 15.5), precision, and inverse variance. For the log of risk or odds ratio, the inverted standard error is recommended. In MetaXL the funnel plot is obtained by choosing Results from the MetaXL menu.

While there has been much focus on selection biases in relation to the association between size and effect in a meta-analysis, it must be kept in mind that asymmetry can also occur for reasons other than selection biases due to selective



**Fig. 15.5** The funnel plot from the fibrinolysis in myocardial infarction meta-analysis (Yusuf et al. 1985). The ES is the LnOR. The *central line* depicts the fixed effects pooled estimate and the limbs of the funnel are made up by the limits of the confidence interval around the pooled estimate being computed successively based on the standard error of each study

publication or selective outcome reporting. These other factors related to study size include study quality (smaller usually thought to be worse), presence of true heterogeneity (e.g. different baseline risks in small and large studies), an association between the intervention effect and its standard error (artefactual) or even chance. Despite the initial expectations, assessment of publication bias using the classic funnel plot continues to misrepresent bias because the appearance of the standard funnel plot has been shown to be misleading. Furthermore, it has been demonstrated that discrepancies between large trials and corresponding meta-analyses and heterogeneity in meta-analyses may also be largely dependent on the arbitrary choice of the method used to construct the classic funnel plot. In particular, the shape of the plot in the absence of bias changes with the choice of axes and it has been suggested that funnel plots of meta-analyses should generally be limited to using standard error as the measure of study size and ratio measures of treatment effect. Even when this is adhered to, the visual and quantitative assessment of asymmetry is flawed. It has been suggested that funnel plot asymmetry detected using measures of impact such as the risk difference (measures that are correlated with baseline risk) may be artefactual and thus funnel plots and related tests using risk differences should not be undertaken.

### *Egger's Regression*

The most popular formal statistical test of funnel plot asymmetry is the Egger's test. Its power is limited, particularly for moderate amounts of bias or meta-analyses based on a small number of small studies. Egger's regression is essentially a linear regression on the standardized ESs ( $Z_i$ ) with precision ( $1/\sigma_i^2$ ) as predictor where  $\theta_i$  is the ES and the standardized ES is then given by

$$z_i = \frac{\theta_i}{\sigma_i}$$

Egger's regression is then

$$z_i = \alpha + \beta \frac{1}{\sigma_i}$$

With no publication bias present, the intercept ( $\alpha$ ) should not be significantly different from zero. This is similar to regression of a Galbraith plot not constrained to the origin (see above).

### ***Doi Plot***

Another plot that is more objective uses the approach of a linear ranking to assess study asymmetry using the same scale for the ES on which its standard error exists. Essentially, each subject in every trial within the meta-analysis is assigned the ES of their trial and ranked serially. As all subjects in a trial have the same ES, they will have the same rank and thus each trial has a single final rank based on the number of subjects ( $N$ ) in the study. However, because  $N$  does not capture the trials' information content completely (the number of observed events in each arm of a study is often more important in driving the precision of the estimate than the study size per se), an updated  $N$  (designated  $N'$ ) is used to incorporate this. The final ranking is then converted to a percentile and then a  $z$ -score using the method detailed below.

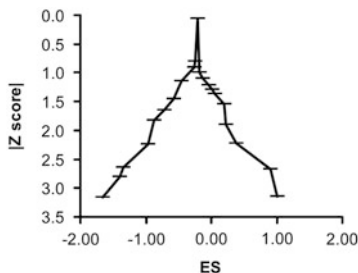
First,  $N'$  is generated as follows:

$$N'_i = \text{int} \left( N_i \times \frac{(\max\{\text{SE}_i^2 N_i\})}{\text{SE}_i^2 N_i} \right)$$

where SE is the standard error of the ES. If there are  $k$  studies in a meta-analysis numbered serially as  $i = 1, \dots, k$  each with an ES and study-adjusted patient-information study size ( $N'$ ), the  $k$  studies can then be ranked by ES and the  $N'$  subjects in these  $k$  trials are serially numbered consecutively. The last subject number in each study ( $A_i$ ) is determined by summing the  $N'_i$  across trials with ES smaller than or equal to the ES under consideration then (using indicator functions):

$$A_i = \sum \{ (N'_1 \times 1_{\{\text{ES}_1 \leq \text{ES}_i\}}), \dots, (N'_k \times 1_{\{\text{ES}_k \leq \text{ES}_i\}}) \}$$

If we assign all subjects in a trial to the ES of their trial, the final rank ( $R_i$ ) of each study based on ES and number of subjects is computed as follows:



**Fig. 15.6** The Doi plot for the same studies as in Fig. 15.5. The ES is the LnOR. The *dashes* represent each study with the absolute Z score on the y-axis plotted against the effect size on the x-axis. Absence of publication bias is indicated by a similar slope of each limb of the plot away from the vertical plane. With publication bias, both limbs will slope differently with respect to the vertical plane

$$R_i = \frac{\max\{(A_1 \times 1_{\{ES_1 < ES_i\}}), \dots, (A_k \times 1_{\{ES_k < ES_i\}})\} + A_i}{2}$$

$R_i$  is then converted into a percentile ( $P_i$ ) as follows:

$$P_i = \frac{(R_i - 0.5)}{\sum_{i=1}^k N_i}$$

Finally the percentile is converted into a z-score [ $z = \text{NORMINV}(P_i, 0, 1)$ ].

This new measure of precision is now the absolute value of the z-score and the ES is then plotted against this absolute value of the z-score to create the new mountain plot. With symmetrical studies, the most precise trials will define the mid-point around which results should scatter, and thus they will be close to mid-rank and will be close to zero on the z-score axis. Smaller less precise trials will produce an ES that scatters increasingly widely, and the absolute z-score will gradually increase for both smaller and larger ES's on either side of that of the precise trials. Thus, a symmetrical triangle is created with a z-score close to zero at its peak. If the trials are homogeneous and not affected by selection or other forms of bias, the plot will therefore resemble a symmetrical triangle with the studies themselves making up the limbs of the plot (Fig. 15.6).

## Bibliography

Bax L, Ikeda N, Fukui N, Yaju Y, Tsuruta H, Moons KG (2009) More than numbers: the power of graphs in meta-analysis. *Am J Epidemiol* 169:249–255

Charlson F, Siskind D, Doi SA, McCallum E, Broome A, Lie DC (2012) ECT efficacy and treatment course: a systematic review and meta-analysis of twice vs thrice weekly schedules. *J Affect Disord* 138:1–8

- Lewis JA, Ellis SH (1982) A statistical appraisal of post-infarction beta-blocker trials. *Primary Cardiol suppl* 1:31–37
- Rücker G, Schwarzer G, Carpenter JR, Schumacher M (2008) Undue reliance on I<sup>2</sup> in assessing heterogeneity may mislead. *BMC Med Res Method* 8:79
- Rucker G, Carpenter JR, Schwarzer G (2011) Detecting and adjusting for small-study effects in meta-analysis. *Biom J* 53:351–368
- Sterne JA, Egger M (2001) Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 54:1046–1055
- Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JP (2011) Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343:d4002
- Takkouche B, Cadarso-Suarez C, Spiegelman D (1999) Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Am J Epidemiol* 150:206–215
- Tang JL, Liu JL (2000) Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 53:477–484
- Terrin N, Schmid CH, Lau J (2005) In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 58:894–901
- Whiting P, Harbord R, Kleijnen J (2005) No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 5:19
- Yusuf S, Collins R, Peto R, Furberg C, Stampfer MJ, Goldhaber SZ, Hennekens CH (1985) Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. *Eur Heart J* 6:556–585