Citation:

Schwarzer, G., Antes, G. & Schumacher, M. (2003). Statistical tests for the detection of bias in meta-analysis. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 71-78). Hogrefe & Huber.

5

Statistical Tests for the Detection of Bias in Meta-Analysis

Guido Schwarzer[1,3]
Gerd Antes[2,3]
Martin Schumacher[2]

[1] Freiburg Center for Data Analysis and Modelling
University of Freiburg

[2] Institute of Medical Biometry and Medical Informatics
University Hospital of Freiburg

[3] German Cochrane Centre
University Hospital of Freiburg

Summary

The result of a meta-analysis as part of a systematic review critically depends on the extent to which relevant information about the particular research question can be retrieved. Biases are especially to be expected due to the selective publication of significant results (publication bias).

For the investigation of biases in meta-analyses, both (informal) graphical as well as statistical methods are used. Within the framework of a simulation study, two tests for biases are compared; a rank-correlation test (Begg & Mazumdar, 1994) and a test based on a linear regression approach (Egger, Smith, Schneider, & Minder, 1997).

5.1 INTRODUCTION

The use of meta-analysis to combine the results of several independent trials is still increasing in the medical field. The validity of a meta-analysis may be affected by various sources of bias, for example, publication bias (Begg & Berlin, 1988; Egger, Smith, et al., 1997) and language bias (Egger, Zellweger-Zahner, et al., 1997). An analysis of bias should be a part of any systematic review. Both statistical tests and graphical methods have been proposed for this purpose. In this chapter, we describe these methods in some detail by the use of a simulated dataset with binary outcome data, which are common in medical applications.

Throughout the chapter, we utilize the following notation. Let t_i denote the estimated effect (e.g., the log relative risk or the log odds ratio) in trial i, $i = 1, \ldots, k$ with $E(t_i) = \mu$ and $\mathrm{Var}(t_i) = \sigma_i^2$. The estimated variance of t_i is denoted by v_i. The inverse variance method is used to derive an overall treatment effect

$$\bar{t} = \frac{\sum_{j=1}^{k}(t_j/v_j)}{\sum_{j=1}^{k}(1/v_j)},$$

where k is the number of trials involved in the meta-analysis (Cooper & Hedges, 1994).

We referred to a survey conducted at the German Cochrane Centre to generate sample sizes of individual trials. All issues from 1948 to 1998 of eight German medical journals were examined and information from all published primary randomized clinical trials was extracted (Galandi, personal communication). We fitted a log-normal distribution to this dataset restricted to trials with a total sample size of at least 30 patients. This log-normal distribution with mean 3.678 and variance 1.146 was used to generate sample sizes. A forest plot of 20 such generated trials with an underlying relative risk of 0.5 is displayed in Figure 5.1. The variance estimates v_i were calculated according to Fleiss (1993). Due to the small sample sizes, many trials in this specific meta-analysis have equal relative risk estimates, for example, five trials result in an estimated relative risk of 0.5. The estimated overall treatment effect is 0.542 with a 95% confidence interval [0.387; 0.757], which is in good agreement with the true treatment effect. This simulated dataset is used for illustrative purposes in the sequel.

5.2 GRAPHICAL METHODS FOR THE DETECTION OF BIAS IN META-ANALYSIS

A funnel plot is the most often used graphical method to check informally the presence of bias in meta-analysis. Beside this method, a radial plot can be used for this purpose, too.

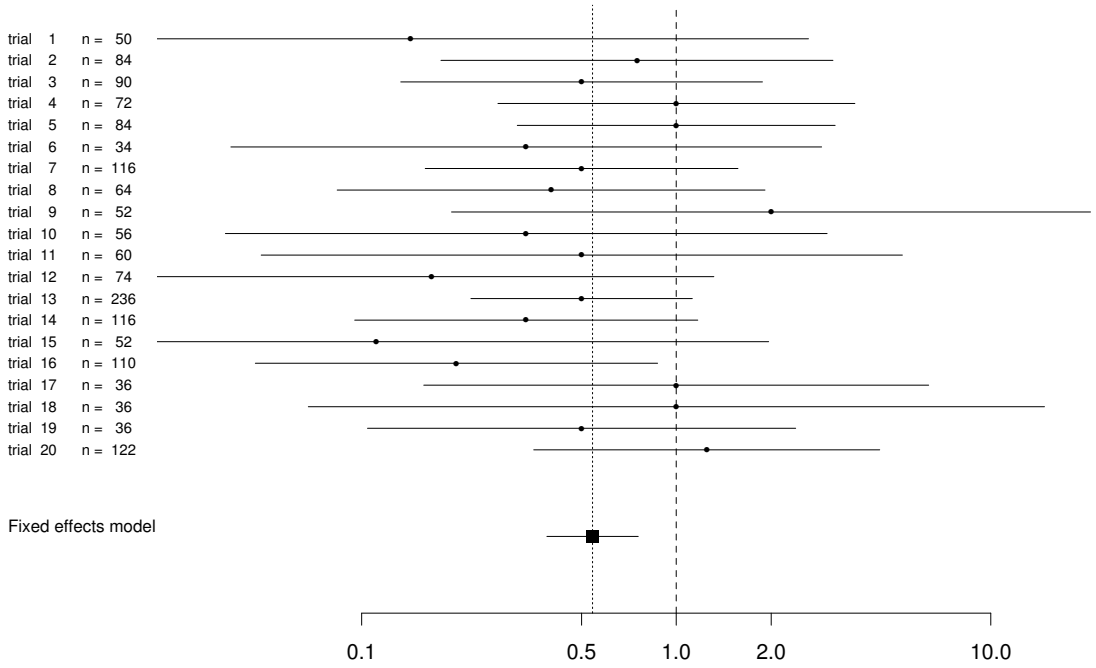

Figure 5.1 Forest plot of simulated meta-analysis; relative risk as measure of treatment effect.

5.3 FUNNEL PLOT

A scatterplot of the estimated treatment effect $x_i = t_i$ and a measure of the precision of t_i is called a funnel plot (Light & Pillemer, 1984). Typically, the sample size $y_i = n_i$ or the inverse of the estimated variance $y_i = 1/v_i$ is used as a measure of precision. For both measures, the display looks like a funnel if no publication bias and between-trial heterogeneity exists showing decreasing fluttering from bottom to top of the graph. Asymmetry in the funnel plot is taken as an indication of bias in the meta-analysis.

A variant of the funnel plot with standard error as measure of precision is displayed in Figure 5.2. The display should look like a triangle centered at the true treatment effect when the standard error is used as measure of precision. This kind of display has been chosen by the statistical methods group of the Cochrane Collaboration as the preferred variant.

We introduced a simple form of bias in the simulated meta-analysis as indicated by the plotting symbol in Figure 5.2. A funnel plot for the published trials (denoted by "s") can be derived from Figure 5.2 because the position of x_i and y_i which contain only trial specific information does not change. A meta-analysis considering only the published trials results in an estimated overall treatment effect of 0.385 with 95% confidence interval from [0.2546; 0.5821]. The asymmetry in the funnel plot is obvious if trials marked with "n" are not considered.

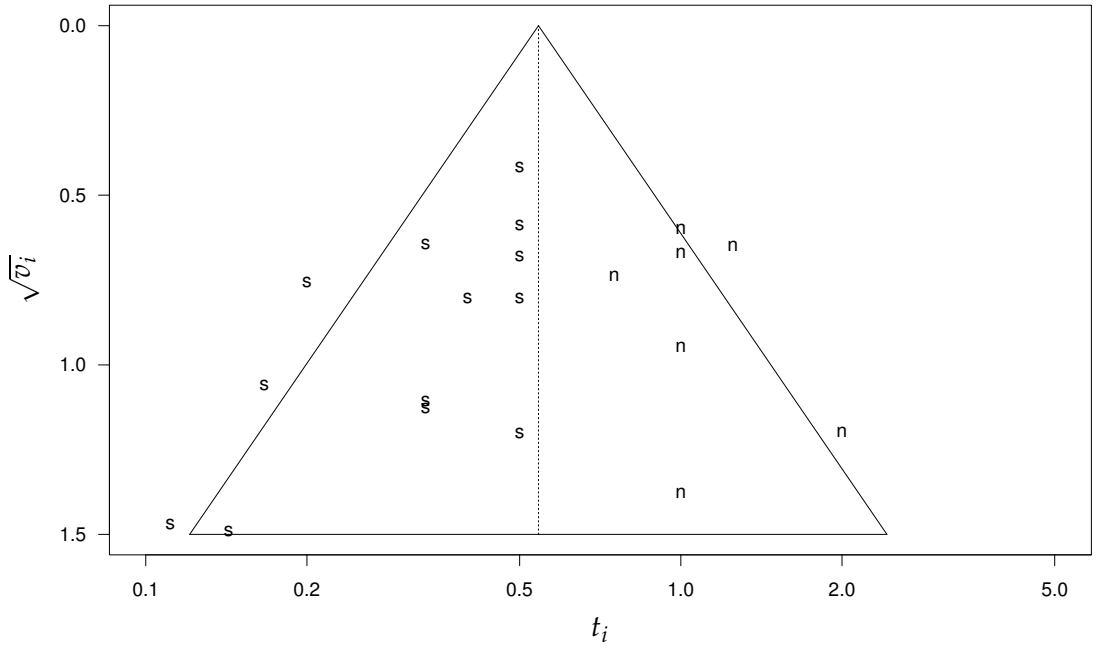

Figure 5.2 Funnel plot of simulated meta-analysis; relative risk as measure of treatment effect; s = trial published, n = trial not published.

5.4 RADIAL PLOT

Galbraith (1988b) introduced the radial plot in order to display several point estimates with different standard errors in a single graph. An additional paper focused on medical applications and the use of the log odds ratio as effect measure of interest (Galbraith, 1988a).

A scatterplot of $x_i = 1/\sigma_i$ and $y_i = t_i/\sigma_i$ is called a radial plot. A radial plot has the following properties (Galbraith, 1988b):

a) $\text{Var}(y_i) = 1$

b) $t_i =$ slope of the line through (0,0) and (x_i, y_i)

c) Points are close to zero on the x-axis for large σ_i

d) Estimated overall effect $\bar{t} =$ slope in linear regression model: $y_i = \beta \cdot x_i + \epsilon_i$.

Due to properties b) and d), a circular scale is typically displayed on the right-hand side of a radial plot showing the estimated treatment effect. Sometimes $y_i^* = (t_i - \bar{t})/\sigma_i$ is plotted against x_i to get a better visual discrimination. In this case, the estimated overall effect coincides with the horizontal axis and departures from the overall effect are more obvious. In practice, the variances σ_i^2 are unknown and have to be estimated.

A radial plot of the simulated meta-analysis is depicted in Figure 5.3. Again, a plot for the published trials can be derived from this figure by omitting trials marked with n because the position of x_i and y_i does not change; a different

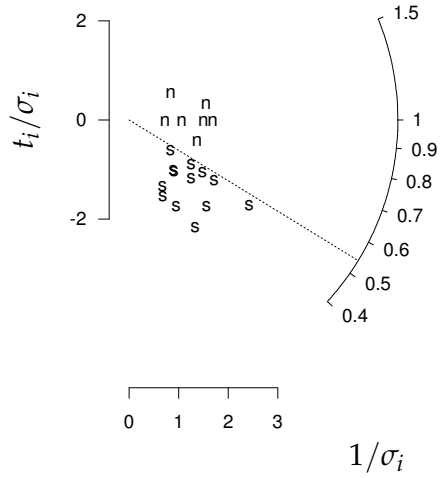

Figure 5.3 Radial plot of simulated meta-analysis; relative risk as measure of treatment effect; s = trial published, n = trial not published; overall treatment effect is indicated by the dashed line.

estimated overall treatment effect has to be considered for the published trials. A gap in the upper left part indicates the presence of bias if only published trials are considered. However, this is more obvious in the funnel plot.

5.5 STATISTICAL TESTS FOR THE DETECTION OF BIAS IN META-ANALYSIS

At least two test procedures for the detection of bias in meta-analysis enjoy some popularity. Begg and Mazumdar (1994) proposed a rank correlation test; Egger, Smith, et al. (1997) introduced a test based on a linear regression of the standard normal deviate on precision which is strongly connected to a radial plot. The estimated variance of the treatment effect in each single trial v_i is of central importance in both tests.

5.5.1 Begg and Mazumdar Test

Begg and Mazumdar (1994) proposed an adjusted rank correlation test for the detection of bias in a meta-analysis and evaluated the power of this test in a simulation study assuming a normal distribution for t_i. The test is based on the correlation between the standardized effect measure

$$t_i^* = \frac{(t_i - \bar{t})}{\sqrt{v_i^*}} \quad \text{with} \quad v_i^* = v_i - \frac{1}{\sum_{j=1}^{k} v_j^{-1}}$$

and the variance v_i. Kendall's tau is used as correlation measure. Let x denote the number of pairs of trials with standardized effects and variances ranked in the same order, that is, ($t_i^* > t_j^*$ and $v_i > v_j$) or ($t_i^* < t_j^*$ and $v_i < v_j$), where

$i \neq j$. The number of pairs ranked in the opposite order are denoted by y. The normalized test statistic for the case that no ties are present neither within t_i^* nor v_i is

$$z = \frac{(x - y)}{\sqrt{\frac{k(k-1)(2k+5)}{18}}},$$

where k is the number of trials involved in the analysis. A modified version for tied observations can be found in Armitage and Berry (1994). The test statistic z has an asymptotic standard normal distribution if the variances σ^2 are known.

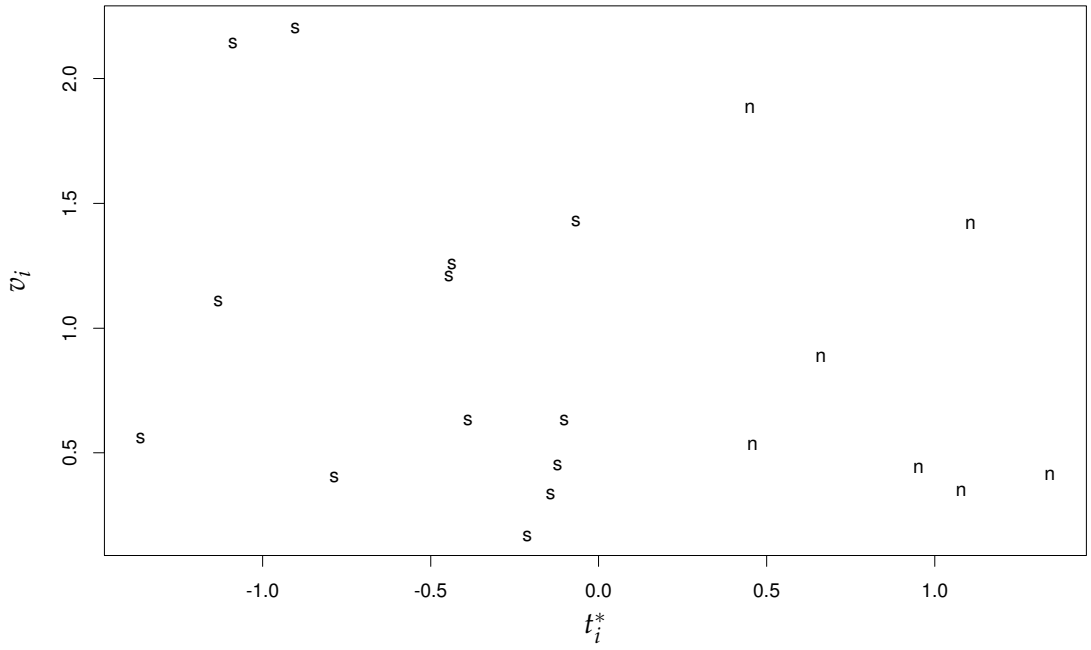

Figure 5.4 Graphical display of the rank correlation test for the detection of bias in the simulated meta-analysis; s = trial published, n = trial not published.

A scatterplot of t_i^* and v_i for the simulated dataset is depicted in Figure 5.4. No correlation between t_i^* and v_i is apparent if all trials are considered. This impression is supported by the result of the rank correlation test. The difference $x - y$ is -28 with a standard error of 30.8 resulting in a p-value of $p = .31$. The shape of the display for the published trials is different from Figure 5.4 because a different overall treatment effect $\bar{t}$ is utilized to calculate t_i^*. The difference is -30 with a standard error of 16.4 resulting in a p-value of $p = .067$ if only published trials are considered.

5.5.2 Egger Test

The test proposed by Egger, Smith, et al. (1997) for the detection of bias in meta-analysis is based on a linear regression of y_i on x_i: $y_i = \alpha + \beta \cdot x_i + \epsilon_i$. In contrast to the radial plot, the regression line is not constrained to run through the origin. A test for the detection of bias is constructed by testing the null-

hypothesis of a zero intercept. The approach is justified by the intuitive argument that, in the presence of publication bias, small trials with non-significant or negative results are less likely to get published. Thus, points close to zero on the x-axis do not scatter randomly around the overall effect resulting in a non-zero intercept, that is, a departure from property d) of a radial plot. The test procedure is implicitly based on the assumption that linearity still holds in the presence of bias.

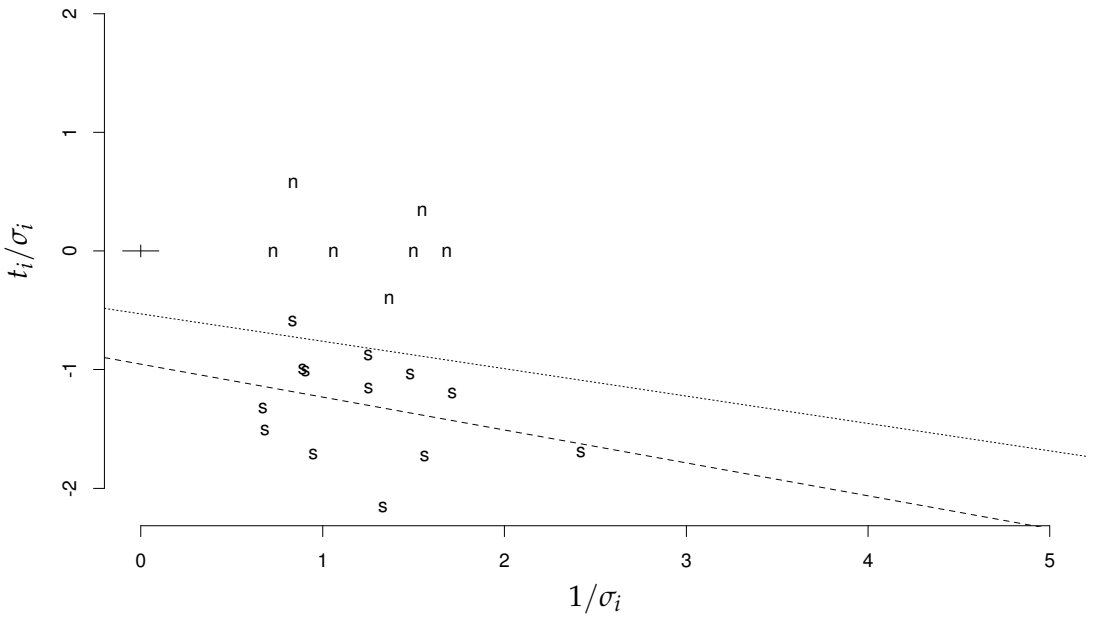

Figure 5.5 Result of the Egger test for the detection of bias in the simulated dataset; x_i and y_i according to a radial plot; regression lines displayed both for all trials and subset of published trials; s = trial published, n = trial not published.

The result of the Egger test for the simulated meta-analysis is displayed in Figure 5.5. The estimated intercept, if all trials are considered, is $\widehat{\alpha} = -0.53$ with a standard error (SE) of $SE(\widehat{\alpha}) = 0.532$ compared to a t-distribution with 18 df, resulting in a p-value of .33. A clear indication of bias is given if only published trials are considered: $\widehat{\alpha} = -0.95$ with $SE(\widehat{\alpha}) = 0.33$ resulting in a p-value of .015 (compared to a t-distribution with 11 degrees of freedom).

5.6 CONCLUDING REMARKS

In this chapter, we described two statistical tests on bias in meta-analysis which have been developed recently. Both tests are implicitly based on the assumption that the variances σ_i^2 are known. The statistical properties of these tests in practical relevant situations are still unknown. Further research is needed, especially with regard to the usefulness in meta-analysis with binary outcome data and in the case of heterogeneity.

We utilized a very simple method to generate bias in meta-analysis by arbitrarily omitting trial results. In order to compare the tests on bias in meta-analysis in simulations, more sophisticated mechanisms to generate bias are needed. This simulation model could be based on an approach described in Copas (1999) linking the probability of trial publication to both sample size and magnitude of observed treatment effect.

REFERENCES

Armitage, P., & Berry, G. (1994). *Statistical methods in medical research* (3rd ed.). Oxford: Blackwell.

Begg, C. B., & Berlin, J. A. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society, Series A*, *151*, 419–445.

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

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Copas, J. (1999). What works?: Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A*, *162*, 95–109.

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

Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomised controlled trials published in English and German. *The Lancet*, *350*, 326–329.

Fleiss, J. L. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research*, *2*, 121–145.

Galbraith, R. F. (1988a). A note on estimated odds ratios from several clinical trials. *Statistics in Medicine*, *7*, 889–894.

Galbraith, R. F. (1988b). Graphical display of estimates having differing standard errors. *Technometrics*, *30*, 271–281.

Light, R. J., & Pillemer, D. B. (1984). *Summing up. The science of reviewing research.* Cambridge: Harvard University Press.