

Maria Blettner, Ulrike Krahn, and Peter Schlattmann

Contents

36.1	Introduction	1378
36.2	Reasons for Meta-Analysis in Epidemiology	1378
36.3	Different Types of Overviews	1380
36.4	Steps in Performing a Meta-Analysis	1381
36.5	Statistical Analysis	1383
36.5.1	Single Study Results	1384
36.5.2	Publication Bias	1384
36.5.3	Estimation of a Summary Effect	1387
36.5.4	Heterogeneity	1389
36.6	Interpretation of the Results of Meta-Analysis of Observational Studies	1397
36.6.1	Bias	1398
36.6.2	Confounding	1398
36.6.3	Heterogeneity	1398
36.7	Network Meta-Analysis	1399
36.7.1	Statistical Methods	1400
36.7.2	Limitations	1402
36.8	Conclusions	1404
Appendices	1404
A.	Data	1404
B.	Elementary Analysis with SAS	1405
C.	SAS Code for the Random Effects Model	1407
References	1407

M. Blettner (✉) • U. Krahn
Institute for Medical Biostatistics, Epidemiology and Informatics, Johannes Gutenberg
University, Mainz, Germany

P. Schlattmann
Institute of Medical Statistics, Computer Sciences and Documentation, Jena University Hospital,
Jena, Germany

36.1 Introduction

The use of meta-analyses in order to synthesize the evidence from epidemiological studies has become more and more popular recently. It has been estimated by Egger et al. (1998) that from articles retrieved by MEDLINE with the medical subject heading (MeSH) term “meta-analysis” some 33% reported results of a meta-analysis from randomized clinical trials and nearly the same proportion (27%) were from observational studies, including 12% papers in which the etiology of a disease was investigated. The remaining papers include methodological publications or review articles. Reasons for the popularity of meta-analyses are the growing information in the scientific literature and the need of timely decisions for risk assessment or in public health. Methods for meta-analyses in order to summarize or synthesize evidence from randomized controlled clinical trials have been continuously developed during the last years. In 1993, the Cochrane Collaboration was established as an international organization, which provides systematic reviews to evaluate healthcare interventions. They have published a handbook (Higgins and Green 2009) with detailed information on how to conduct systematic reviews of randomized clinical trials. While methods for meta-analyses of randomized clinical trials are now also summarized in several text books, for example, Sutton et al. (2000) and Whitehead (2002), and in a handbook by Egger et al. (2001a) and Dickersin (2002) argued that statistical methods for meta-analyses of epidemiological studies are still behind in comparison to the progress that has been made for randomized clinical trials. The use of meta-analyses for epidemiological research caused many controversial discussions; see, for example, Blettner et al. (1999), Berlin (1995), Greenland (1994), Feinstein (1995), Olkin (1994), Shapiro (1994a, b), or Weed (1997) for a detailed overview of the arguments. The most prominent arguments against meta-analyses are the fundamental issues of confounding, selection bias, as well as the large variety and heterogeneity of study designs and data collection procedures in epidemiological research. Despite these controversies, results from meta-analyses are often cited and used for decisions. They are often seen as the fundamentals for risk assessment. They are also performed to summarize the current state of knowledge often prior to designing new studies.

This chapter will first describe reasons for meta-analyses in epidemiological research and then illustrate how to perform a meta-analysis with the focus on meta-analysis of published data. Furthermore, network meta-analyses are introduced, which are a new methodological tool.

36.2 Reasons for Meta-Analysis in Epidemiology

One major issue in assessing causality in epidemiology is “consistency” as pointed out by Hill in 1965. The extent to which an observed association is similar in

different studies, with different designs, using different methods of data collection and exposure assessment, by different investigators, and in different regions or countries is an essential criterion for causality. If different studies with inconsistent results are known, there is a need for understanding the differences. Reasons may be small sample sizes of individual studies (chance), different methods of exposure assessment (measurement errors), different statistical analyses (e.g., adjustment for confounding), or the use of different study populations (selection bias). Also, Thompson et al. (1997) showed that different baseline risks may cause heterogeneity. The goal of a meta-analysis is then to investigate whether the available evidence is consistent and/or to which degree inconsistent results can be explained by random variation or by systematic differences between design, setting, or analysis of the study as has been pointed out by Weed (2000).

Meta-analyses are often performed to obtain a combined estimator of the quantitative effect of the risk factor such as the relative risk (*RR*) or the odds ratio (*OR*). As single studies are often far too small to obtain reliable risk estimates, the combination of data of several studies may lead to more precise effect estimates and increased statistical power. This is mainly true if the exposure leads only to a small increase (or decrease) in risk or if the disease or the exposure of interest is rare. One example is the risk of developing lung cancer after the exposure to passive smoking where relative risk estimates in the order of 1.2 have been observed; see Boffetta (2002) for a summary of the epidemiological evidence. Another typical example is the association between childhood leukemia and exposure to electromagnetic fields. Meinert and Michaelis (1996) have performed a meta-analysis of the available case-control studies as the results of the investigations were inconsistent. Although many huge case-control studies have been performed in the last decade, in each single study, only a few children were categorized as “highly exposed.” In most publications, a small but non-significant increase in risk was found, but no single study had enough power to exclude that there is no association between EMF exposure and childhood leukemia.

Sometimes, meta-analyses are also used to investigate more complex dose-response functions. For example, Tweedie and Mengersen (1995) investigated the dose-response relationship of exposure to passive smoking and lung cancer. A meta-analysis was also undertaken by Longnecker et al. (1988) to study the dose-response of alcohol consumption and breast cancer risk. However, results were limited as not enough data were present in several of the included publications. Interestingly, a large group of investigators led by Hamajima et al. (2002) has recently used individual patient data from 53 studies including nearly 60,000 cases for a reanalysis. It has been shown by Sauerbrei et al. (2001) in a critique that meta-analysis from aggregated data may be too limited to perform a dose-response analysis. A major limitation is that different categories are used in different publications. Thus, dose-response analyses are restricted to published values. Meta-analyses of published data have their main merits for exploring heterogeneity between studies and to provide crude quantitative estimates but probably less for investigating complex dose-response relationships.

36.3 Different Types of Overviews

Approaches for summarizing evidence include five different types of overviews: first, traditional narrative reviews that provide a qualitative but not a quantitative assessment of published results. Methods and guidelines for reviews have been recently published by Weed (1997).

Second, meta-analyses from literature (MAL) which are generally performed from freely available publications without the need of cooperation and without agreement of the authors from the original studies. They are comparable to a narrative review in many respects but include quantitative estimate(s) of the effect of interest. One recent example is a meta-analysis by Zeeger et al. (2003) of studies investigating some familial clustering of prostate cancer. Another meta-analysis has been recently published by Allam et al. (2003) on the association between Parkinson disease, smoking, and family history.

Third, meta-analyses with individual patient data (MAP) in which individual data from published and sometimes also unpublished studies are reanalyzed. Often, there is a close cooperation between the researcher performing the meta-analysis and the investigators of the individual studies. The new analysis may include specific inclusion criteria for patients and controls, new definition of the exposure and confounder variables, and new statistical modeling. This reanalysis may overcome some but not all of the problems of meta-analyses of published data (Blettner et al. 1999). They have been performed in epidemiological research for many years. One of the largest investigations of this form was a recent investigation on breast cancer and oral contraceptive use, where data from 54 case-control studies were pooled and reanalyzed (CGHFBC 1996). A further international collaboration led by Lubin and colleagues were set up to reanalyze data from 11 large cohort studies on lung cancer and radon among uranium miners. The reanalysis allowed a refined dose-response analysis and provided data for radiation protection issues. Pooled reanalyses are mostly performed by combining data from studies of the same type only. For example, Hung et al. (2003) reanalyzed data from all case-control studies in which the role of genetic polymorphisms for lung cancer in non-smokers was investigated. The role of diet for lung cancer was recently reviewed by Smith-Warner et al. (2002) in a qualitative and quantitative way by combining cohort studies. An overview of methodological aspects for a pooled analysis of data from cohort studies was published by Bennett (2003). There are methods under development which allow to combine individual and aggregate data. See, for example, the work by Riley et al. (2007) or Sutton et al. (2008).

Fourth, prospectively planned pooled meta-analyses of several studies in which pooling is already a part of the protocol. Data collection procedures, definitions of variables are as far as possible standardized for the individual studies. The statistical analysis has many similarities with the meta-analysis based on individual data. A major difference, however, is that joint planning of the data collection and analysis increase the homogeneity of the included data sets. However, in contrast to multicenter randomized clinical trials, important heterogeneity between the study centers still may exist. This heterogeneity may arise from differences

in populations, in the relevant confounding variables (e.g., race may only be a confounder in some centers) and potentially differences in ascertainment of controls. For example, complete listings of population controls are available in some but not all countries. In the latter situation sometimes neighborhood controls are used. Mainly in occupational epidemiology, those studies are rather common, many of them were initiated by international bodies such as the International Agency for Research on Cancer (IARC) as the international pooled analysis by Boffetta et al. (1997) of cancer mortality among persons exposed to man-made mineral fiber. Another example for a prospectively planned pooled meta-analysis is given by a large brain tumor study initiated by the IARC including data from eight different countries (see Schlehofer et al. 1999).

Fifth, network meta-analyses which is a method that is increasingly used to estimate comparative effectiveness of treatments not compared directly in randomized controlled trials. This technique is introduced in more detail in [Sect. 36.7](#).

Steinberg et al. (1997) compared the effort required and the results obtained of MAL and MAP with an application to ovarian cancer. Certainly, MAL are easier to perform, cheaper, and faster than MAP. Their credibility may be more questionable as discussed by many authors; see, for example, Blettner et al. (1999) or Egger et al. (1998). Statistical issues of pooling data from case-control studies have been investigated by Stukel et al. (2001) recently. The authors proposed a two-step approach and showed conditions under which the two-step approach gives similar results in comparison to the pooled analysis including all data. Here, the two-step approach implies to estimate first the odds ratio for each study in the usual way. Then in the second step, a combined estimator using either a fixed or random effects model is calculated.

36.4 Steps in Performing a Meta-Analysis

Each type of overview needs a clear study protocol that describes the research question and the design, including how studies are identified and selected, the statistical methods to use, and how the results will be reported. This protocol should also include the exact definition of the disease of interest, the risk factors, and the potential confounding variables that have to be considered. In accordance with Friedenreich (1993) and Jones (1992), the following steps are needed for a meta-analysis/pooled analysis.

Step 1. Define a clear and focused topic for the review: As for any other investigation, a clear protocol in which the research hypothesis, that is, the objectives of the meta-analysis are described, is mandatory. This protocol should include the exact definition of the disease of interest, the risk factors, and the potential confounding variables that have to be considered. The protocol should also include details on the steps that are described below, including specification of techniques for location of the studies, the statistical analysis, and the proposed publications.

Step 2. Establish inclusion and exclusion: It is important to define in advance which studies should be included into the meta-analysis. These criteria may include restrictions on the publication year as older studies may not be comparable to newer ones, on the design of the investigation, for example, to exclude ecological studies. Friedenreich (1994) has also proposed quality criteria to evaluate each study. Whether these criteria, however, should be used as inclusion criteria is discussed controversially. Another decision is whether studies that are only published as abstracts or internal communications should be included (Cook et al. 1993). A rule for the inclusion or exclusion of papers with repeated publication of the data is required. For example, for cohort studies, often several publications with different follow-up periods can be found. As one out of many examples, a German study among rubber workers by Straif et al. (1999, 2000) can be mentioned. In one paper, 11,633 workers were included, while the second paper is based on a subcohort of only 8,933 persons. Which results are more appropriate for the meta-analysis?

Step 3. Locate all studies (published and unpublished) that are relevant to the topic: Since the existence of electronic databases, retrieval of published studies has become much easier. Mainly systems like MEDLINE or CANCERLIT from the National Library of Medicine are valuable sources to locate publications. However, as Dickersin et al. (1994) showed for some examples as little as 50% of the publications were found by electronic searches. Therefore, there is a need to extend the search by manual checks of the reference lists of retrieved papers, monographs, books, and if possible by personal communications with researchers in the field. A clear goal of the search has to be to identify all relevant studies on the topic that meet the inclusion criteria. Egger et al. (2003) have pointed out that the completeness of the literature search is an important feature of the meta-analysis to avoid publication bias or selection bias. Of course, the publication should include the search strategies as well as the keywords and the databases used for electronic searches.

Step 4. Abstract information from the publications: The data collection step in a meta-analysis needs as much care as in other studies. In the meta-analysis, the unit of observation is the publication, and defined variables have to be abstracted from the publication (Stock 1995). In epidemiological studies, the key parameter is often the relative risk or odds ratio. Additionally, standard error, sample size, treatment of confounders, and other characteristics of the study design and data collection procedure need to be abstracted to assess the quality of the study. This is also important for subgroup analyses or for a sensitivity analysis. An abstract form has to be created before abstracting data. This form should be tested like other instruments in a pilot phase. Unfortunately, it may not always be possible to abstract the required estimates directly, for example, standard errors are not presented and have to be calculated based on confidence intervals (Greenland 1987). It may be necessary to contact the investigators to obtain further information if results are not published in sufficient detail. Abstracting and classification of study characteristics is the most time-consuming part of the meta-analysis. It has been recommended to blind the data abstractors, although

some authors argue that blinding may not have a major influence on the results; for further discussion, see Berlin et al. (1997). Additionally, the rater may be acquainted with some of the studies and blinding cannot be performed. Another requirement is that two persons should perform the abstraction in parallel. When a meta-analysis with original data is performed, the major task is to obtain data from all project managers in a compatible way. Our experience shows that this is possible in principle but time consuming as data may not be available on modern electronic devices and often adaptations between database systems are required.

Step 5. Descriptive analysis: A first step in summarizing the results should be an extensive description of the single papers, including tabulation of relevant elements of each study, such as sample size, data collection procedures, confounder variables, means of statistical analysis, study design, publication year, performing year, geographical setting, etc. This request is also included in the guidelines for publications of meta-analysis that were published by Stroup et al. (2000).

Step 6. Statistical analysis: This includes the analysis of the heterogeneity of the study-specific effects, the calculation of a pooled estimate and the confidence interval as well as a sensitivity analysis. Details are given in the next section on statistical methods.

Step 7. Interpretation of the results: The importance of the sources and magnitude of different biases should be taken into account when interpreting the results. Combining several studies will often give small confidence intervals and suggest a false precision (Egger et al. 1998), but estimates may be biased. For clinical studies, Thompson (1994) has pointed out that the investigation of the heterogeneity between studies will generally give more insight than inspecting the confidence interval of the pooled estimate. This is even more true for a meta-analysis from epidemiological studies. Additionally, the possible effects of publication bias (see below) need to be considered carefully (Copas and Shi 2001).

Step 8. Publication: Guidelines for reporting meta-analyses of observational studies have been published by Stroup et al. (2000). More recently, a large group of methodologists and clinicians revised an existing guidance and checklist focused on the reporting of meta-analyses of randomized controlled trials named PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Moher et al. 2009). These guidelines are quite useful for preparing the publication and are also supported by most editors of major medical journals. Especially the detailed description of methods is required so that the analysis could be replicated by others.

36.5 Statistical Analysis

The statistical analysis of aggregated data from published studies was first developed in the fields of psychology and education (Glass 1977; Smith and Glass 1977). These methods have been adopted since the mid-1980s in medicine primarily

for randomized clinical trials and are also used for epidemiological data. A recent overview may be found in the review by Sutton and Higgins (2008). We will give a brief outline of some issues of the analysis using an example based on a meta-analysis performed by Sillero-Arenas et al. (1992). This study was one of the first meta-analyses which tried to summarize quantitatively the association between hormone replacement therapy (HRT) and breast cancer in women. Sillero-Arenas et al. based their meta-analysis on 23 case-control and 13 cohort studies. The data extracted from their paper are given in [Appendix A](#) of this chapter.

The statistical analysis of MAP is more complex and not covered here.

36.5.1 Single Study Results

A first step of the statistical analysis is the description of the characteristics and the results of each study. Tabulations and simple graphical methods should be employed to visualize the results of the single studies. Plotting the odds ratios and their confidence intervals (so-called forest plot) is a simple way to spot obvious differences between the study results.

[Figure 36.1](#) shows a forest plot of 36 studies investigating the association of HRT and breast cancer in women. Obviously there is a high variability of effects between studies present. Later we will describe how to account for heterogeneity of studies quantitatively.

36.5.2 Publication Bias

An important problem of meta-analysis is publication bias. This bias has received a lot of attention particularly in the area of clinical trials. Publication bias occurs when studies that have non-significant or negative results are published less frequently than positive studies. For randomized clinical trials, it has been shown that even with a computer-aided literature search not all of the relevant studies will be identified (Dickersin et al. 1994). In epidemiological studies, additional problems exist, because often a large number of variables will be collected as potential confounders (Blettner et al. 1999), but only significant or important results are mentioned in the abstract. Therefore, the respective publication may be overlooked for variables which are not mentioned in the abstract. Results of important variables may be published in additional papers, which have often not been planned in advance. In general, publication bias yields a non-negligible overestimation of the risk estimate. Consequently, publication bias should be investigated prior to further statistical analyses.

A simple graphical tool to detect publication bias is the so-called funnel plot. The basic idea is that studies which do not show an effect and which are not statistically significant are less likely to be published. If the sample size or alternatively the precision (i.e., the inverse of the variance) is plotted against the effect, a hole in lower left quadrant is expected. [Figure 36.2](#) shows examples of funnel plots. The left

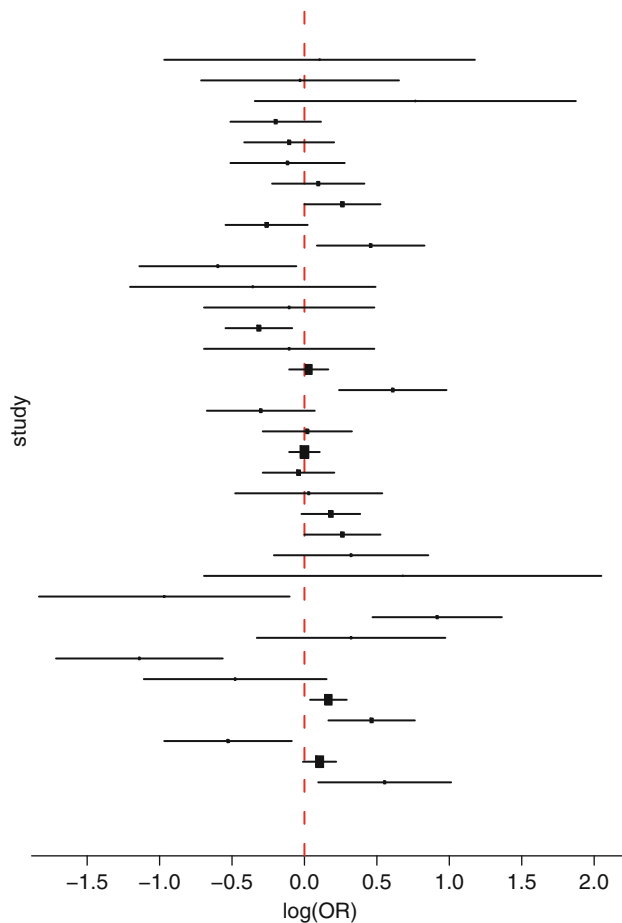


Fig. 36.1 Confidence interval plot of the breast cancer data

subplot of Fig. 36.2 shows a funnel plot with no indication of publication bias. The right subplot shows a so-called apparent hole in the lower left corner. In the case of the right subplot of Fig. 36.2, the presence of publication bias would be assumed.

Figure 36.3 shows a funnel plot for the breast cancer data. No apparent hole in the lower left corner is present. Thus, based on this figure, no publication bias would be assumed.

For a quantitative investigation of publication bias, several methods are available. This may be based on statistical tests; see, for example, Begg and Mazumdar (1994) or Schwarzer et al. (2002). A recent simulation study performed by Macaskill et al. (2001) favored the use of regression methods. The basic idea is to regress the estimated effect sizes $\hat{\theta}_i$ directly on the sample size or the inverse variance σ_i^{-2} as predictor. An alternative idea is to use Egger's regression test (Egger et al. 1997)

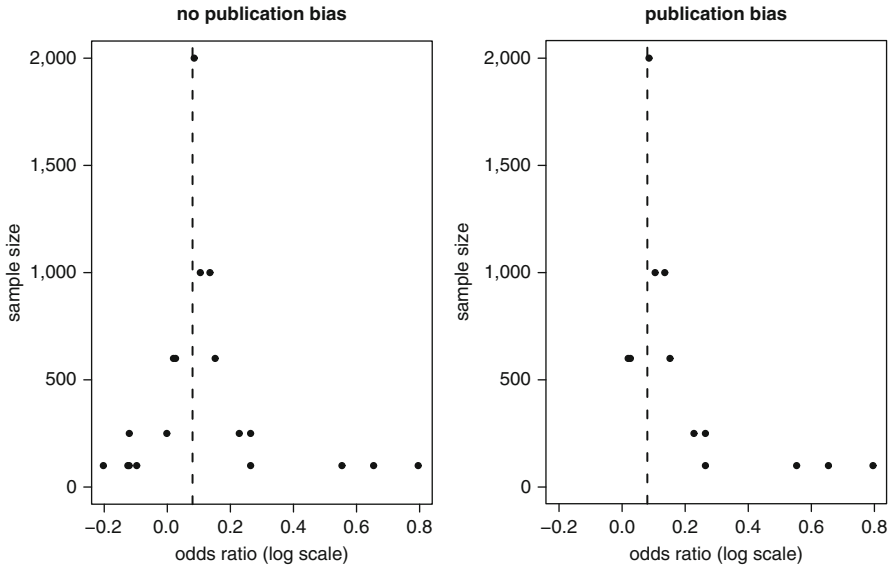


Fig. 36.2 Examples of funnel plots based on simulated data with (*right figure*) and without publication bias present (*left figure*). The dotted line shows the true effect

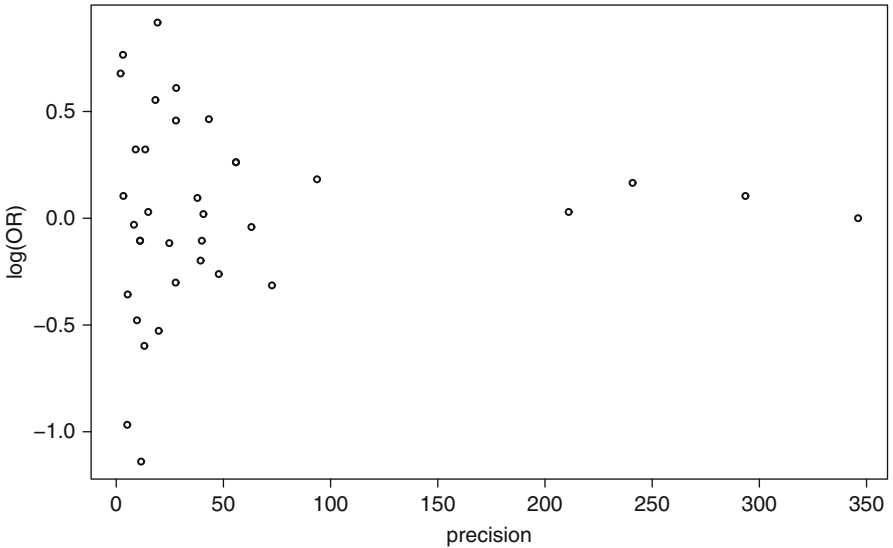


Fig. 36.3 Funnel plot of the breast cancer data

which uses standardized study-specific effects as dependent variable and the corresponding precision as an independent variable. Simulation studies by Macaskill et al. (2001) and by Peters et al. (2006) indicate that the method proposed by Macaskill is superior to Egger's method. Thus, our analysis is restricted to the method by Macaskill which leads to the following regression model

$$\hat{\theta}_i = \alpha + \beta \frac{1}{\sigma_i^2} + \epsilon_i, \quad i = 1, \dots, k, \quad \epsilon_i \sim N(0, \sigma_i^2).$$

Here, the number of studies to be pooled is denoted by k . In this setting, it is assumed that the estimated treatment effects are independently normally distributed. With no publication bias present, the regression line should be parallel to the x -axis, that is, the slope should be zero. A non-zero slope would suggest an association between sample size or inverse variance, possibly due to publication bias. The estimated regression line in Fig. 36.4 shows no apparent slope. Likewise, the model output (not shown) does not indicate the presence of publication bias for the data at hand.

36.5.3 Estimation of a Summary Effect

Frequently, one of the aims of a meta-analysis is to provide an estimate of the overall effect of all studies combined. Methods for pooling depend on the data available. In general, a two-step procedure has to be applied. First, the risk estimates and

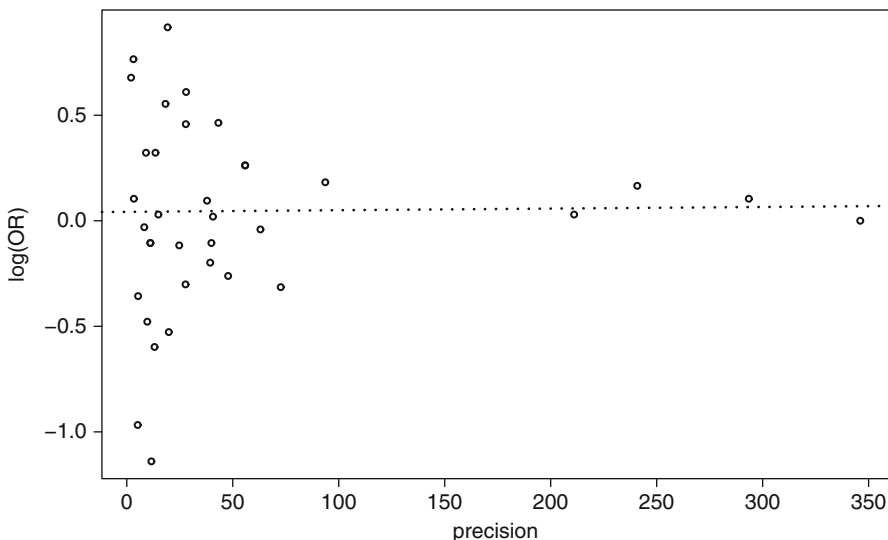


Fig. 36.4 Funnel regression plot of the breast cancer data

variances from each study have to be abstracted from publications or calculated if data are available. Then, a combined estimate is obtained as a (variance based) weighted average of the individual estimates. The methods for pooling based on the 2×2 table include the approaches by Mantel-Haenszel and Peto (see Pettiti 1994 for details). If data are not available in a 2×2 table, but as an estimate from a more complex model (such as an adjusted relative risk estimate), the Woolf approach can be adopted using the estimates and their (published or calculated) variance resulting from the regression model. This results in a weighted average of the log odds ratios $\hat{\theta}_i$ of the individual studies where the weights w_i are given by the inverse of the study-specific variance estimates $\hat{\sigma}_i^2$. For a discussion of risk measures, see chapter ► [Rates, Risks, Measures of Association and Impact](#) of this handbook. Please note that the study-specific variance is assumed to be fixed and known, although they are based on estimates of the study-specific variances. As a result, the uncertainty associated with the estimation of σ_i^2 is ignored. Thus, in the following, the σ_i are treated as constants and the “hat” notation is omitted. The estimate of the summary effect of all studies is then given by

$$\hat{\theta} = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i},$$

$$w_i = \frac{1}{\sigma_i^2}.$$

The variance is given by

$$\text{var}(\hat{\theta}) = \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^2}}.$$

Applying this approach to the HRT data leads to a pooled risk estimate of 0.05598 with an estimated variance equal to 0.00051. Transforming this back to the original scale leads to an odds ratio of 1.058 with a 95% confidence interval of (1.012, 1.11). Thus, we would conclude combining all studies that there is a small harmful effect of hormone replacement therapy.

The major assumption here is that of a fixed model, that is, it is assumed that the underlying true exposure effect in each study is the same. The overall variation and, therefore, the confidence intervals will reflect only the random variation within each study but not any potential heterogeneity between the studies.

[Figure 36.5](#) displays this idea. Whether pooling of the data is appropriate should be decided after investigating the heterogeneity of the study results. If the results vary substantially, no pooled estimator should be presented or only estimators for selected subgroups should be calculated (e.g., combining results from case-control studies only).

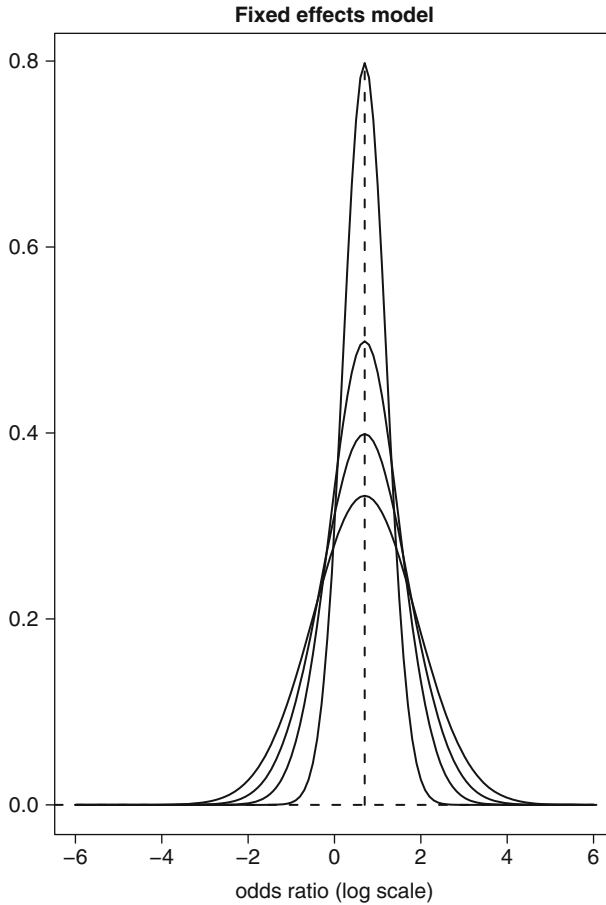


Fig. 36.5 Fixed effects model: common effect with different study variances

36.5.4 Heterogeneity

The investigation of heterogeneity between the different studies is a main task in each review or meta-analysis (Thompson 1994). For the quantitative assessment of heterogeneity, several statistical tests are available (Pettiti 1994; Paul and Donner 1989). A simple test for heterogeneity is based on the following test statistic:

$$\chi_{\text{het}}^2 = \sum_{i=1}^k \frac{(\hat{\theta} - \hat{\theta}_i)^2}{\sigma_i^2} \sim \chi_{k-1}^2 \quad (36.1)$$

which under the null hypothesis of heterogeneity follows a χ^2 distribution with $k - 1$ degrees of freedom. Hence, the null hypothesis is rejected if χ_{het}^2 exceeds the $1 - \alpha$

quantile of χ_{k-1}^2 denoted as $\chi_{k-1,1-\alpha}^2$. For the data at hand, we clearly conclude that there is heterogeneity present ($\chi_{\text{het}}^2 = 116.076$, $\text{df} = 35$, p -value: 0.00000). Thus, using a combined estimate is at least questionable. Pooling the individual studies and performing this test can be done with any statistical package capable of weighted least squares regression. In part B of the Appendix of this chapter shows a SAS program which provides the results obtained so far. A major limitation of formal heterogeneity tests like the one presented before is, however, their low statistical power to detect any heterogeneity present.

A more powerful method is given by model-based approaches. A model-based approach has the advantage that it can be used to test specific alternatives and thus has a higher power to detect heterogeneity. So far, we considered the following simple fixed effects model:

$$\theta_i = \theta + \varepsilon_i, \quad i = 1, \dots, k, \quad \varepsilon_i \sim N(0, \sigma_i^2).$$

Obviously, this model is not able to account for any heterogeneity, since deviations from θ_i and θ are assumed to be explained only by random error.

Thus, alternatively a random effects model should be considered. This model incorporates variation between studies. It is assumed that each study has its own (true) exposure effect and that there is a random distribution of these true exposure effects around a central effect. This idea is presented in Fig. 36.6. Frequently, it is assumed that the individual study effects follow a normal distribution with mean θ_i and variance σ_i^2 and the random distribution of the true effects is again a normal distribution with variance τ^2 . In other words, the random effects model allows non-homogeneity between the effects of different studies. This leads to the following model:

$$\theta_i = \theta + b_i + \varepsilon_i, \quad i = 1, \dots, k, \quad b_i \sim N(0, \tau^2), \quad \varepsilon_i \sim N(0, \sigma_i^2). \quad (36.2)$$

The observed effects from the different studies are used to estimate the parameters describing the fixed and random effects. This may be done using maximum-likelihood procedures. In the following sections, two further methods of estimating the heterogeneity variance in a random effects model are presented.

36.5.4.1 The Heterogeneity Variance Estimator Proposed by DerSimonian and Laird

The widely used approach by DerSimonian and Laird (1986) applies a method of moments to obtain an estimate of τ^2 . Taking the expectation of (36.2) leads to $E(\theta_i) = \theta$ and calculating the variance leads to $\text{var}(\theta_i) = \text{var}(b_i) + \text{var}(\varepsilon_i) = \tau^2 + \sigma_i^2 = \sigma_i^{*2}$ assuming that b_i and ε_i are independent. The heterogeneity variance τ^2 is unknown and has to be estimated from the data. The method by DerSimonian and Laird equates the heterogeneity test statistic (36.1) to its expected value. This expectation is calculated under the assumption of a random effects model and given

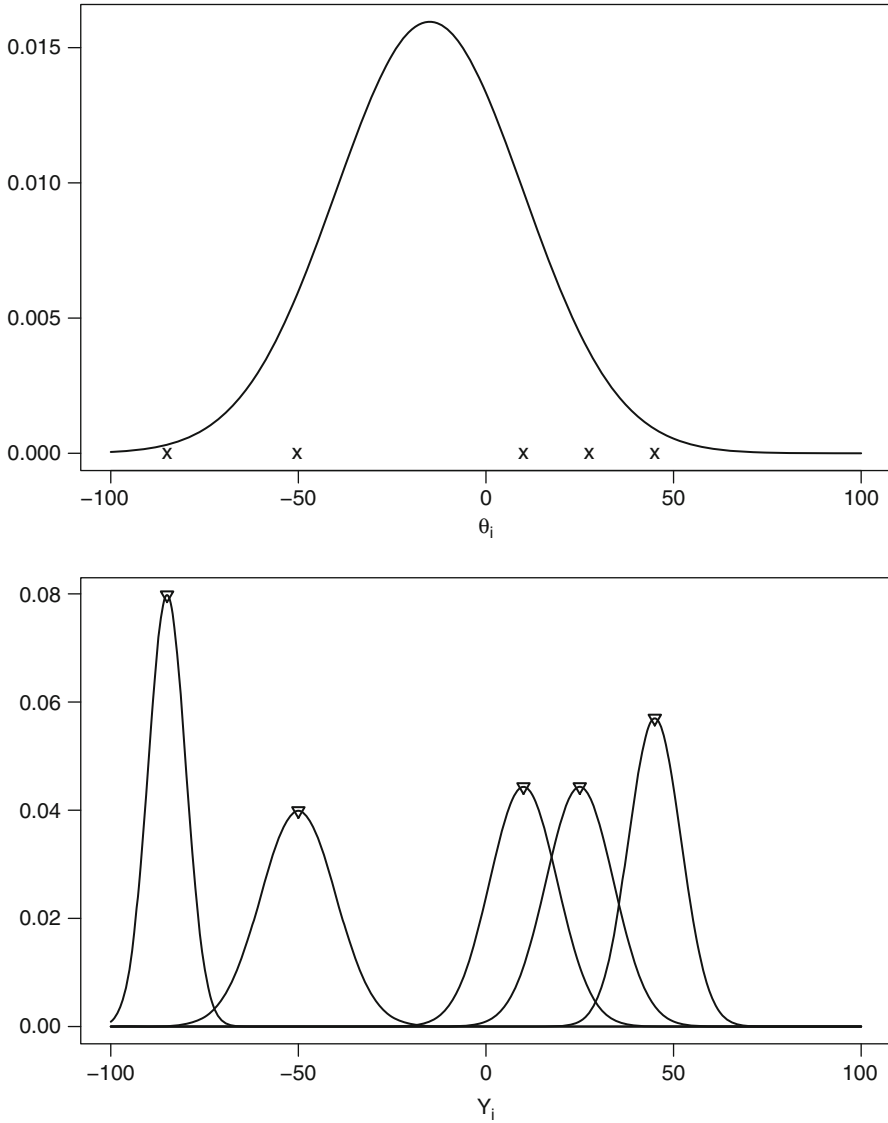


Fig. 36.6 Random effects model: variable effects drawn from a population of study effects

by $E(\chi^2_{\text{het}}) = k - 1 + \tau^2 \left(\sum w_i - \frac{\sum w_i^2}{\sum w_i} \right)$. The weights w_i are those defined in (36.1). Equating χ^2_{het} to its expectation and solving for τ^2 gives

$$\hat{\tau}^2 = [\chi^2_{\text{het}} - (k - 1)] / \left(\sum w_i - \frac{\sum w_i^2}{\sum w_i} \right).$$

In case $\chi_{\text{het}}^2 < k - 1$, the estimator $\hat{\tau}^2$ is truncated to zero. Thus, the pooled estimator $\hat{\theta}_{\text{DL}}$ under heterogeneity can be obtained as weighted average:

$$\hat{\theta}_{\text{DL}} = \frac{\sum_{i=1}^k w_i^* \hat{\theta}_i}{\sum_{i=1}^k w_i^*} \quad (36.3)$$

with $w_i^* = \frac{1}{\sigma_i^{*2}} = \frac{1}{\hat{\tau}^2 + \sigma_i^2}$ we obtain (36.4)

$$\hat{\theta}_{\text{DL}} = \frac{\sum_{i=1}^k \hat{\theta}_i / (\hat{\tau}^2 + \sigma_i^2)}{\sum_{i=1}^k 1 / (\hat{\tau}^2 + \sigma_i^2)}.$$

The variance of this estimator is given by:

$$\begin{aligned} \text{var}(\hat{\theta}_{\text{DL}}) &= \frac{1}{\sum_{i=1}^k \frac{1}{\sigma_i^{*2}}} \\ &= \frac{1}{\sum_{i=1}^k \frac{1}{\hat{\tau}^2 + \sigma_i^2}}. \end{aligned}$$

The between-study variance τ^2 can also be interpreted as a measure for the heterogeneity between studies. For our example, we obtain a pooled DerSimonian-Laird estimate of 0.0337 with heterogeneity variance equal to 0.0453. The variance of the pooled estimator is given by 0.0024. Transformed back to the original scale, we obtain an odds ratio of $OR = 1.034$ with 95% CI (0.939, 1.139). Based on this analysis, we would conclude that after adjusting for heterogeneity, this meta-analysis does not provide evidence for an association between HRT replacement therapy and breast cancer in women. It should be noted that within this approach, the study-specific variances are assumed to be known constants. This can lead to a considerable bias when pooling estimates using the DerSimonian-Laird estimator as demonstrated by Böhning et al. (2002).

36.5.4.2 Another Estimator of τ^2 : The Simple Heterogeneity Variance Estimator

Sidik and Jonkman (2005) proposed a simple method for the estimation of the heterogeneity variance τ^2 . As before, they consider a model as in (36.2) with $b_i \sim N(0, \tau^2)$. Now, for the purpose of their estimation method random effects model, they reparameterize the variance as

$$\text{var}(\hat{\theta}_i) = \tau^2 + \sigma_i^2 = \tau^2 \frac{\sigma_i^2 + \tau^2}{\tau^2} = \tau^2 (r_i + 1)$$

with $r_i = \sigma_i^2 / \tau^2$. Then, the problem of estimating τ^2 is cast into the framework of simple linear regression:

$$E(\hat{\theta}) = X\theta$$

$$\text{var}(\hat{\theta}_1) = \tau^2 V.$$

Here, X is a vector of 1s with dimension $k \times 1$, and V is a diagonal matrix with elements $r_1 + 1, \dots, r_k + 1$. In the framework of the usual weighted least squares, an estimator of θ is obtained as

$$\hat{\theta}_v = \frac{\sum_{i=1}^k v_i^{-1} \hat{\theta}_i}{\sum_{i=1}^k v_i^{-1}}$$

with $v_i = r_i + 1$. Please note that this is equivalent to (36.3) and (36.4). The advantage of casting the problem into the usual weighted least squares approach is that an estimator of τ^2 can be obtained as the usual weighted residual sum of squares as follows:

$$\hat{\tau}^2 = \frac{1}{k-1} \sum_{i=1}^k v_i^{-1} (\hat{\theta}_i - \hat{\theta}_v)^2. \quad (36.5)$$

Of course, an estimator of τ is needed to compute the ratios r_i of the within-study variance σ_i^2 and the between-study variance τ^2 . Here, Sidik and Jonkman propose to use the empirical variance of the study-specific estimates $\hat{\theta}_i$

$$\hat{\tau}_0^2 = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_i - \bar{\theta})^2.$$

Plugging this into (36.5) leads to

$$\hat{\tau}_{SH}^2 = \frac{1}{k-1} \sum_{i=1}^k \hat{v}^{-1} (\hat{\theta}_i - \hat{\theta}_v)^2$$

with $\hat{v}_i = \hat{r}_i + 1$ and $\hat{r}_i = \sigma_i^2 / \hat{\tau}_0^2$. Please note that this estimator is strictly positive in contrast to the DerSimonian-Laird estimator. This estimator will be referred to as the heterogeneity variance estimator SH. For the data at hand, an estimate of τ^2 equals 0.199.

36.5.4.3 Heterogeneity Variance Estimation by Likelihood-Based Methods

Besides the moment-based method by DerSimonian and Laird or the simple heterogeneity variance estimator, estimates of τ^2 can be obtained using likelihood-based

methods. See, for example, the tutorials by Normand (1999) and van Houwelingen et al. (2002) for more details. The Appendix B and C of this chapter gives a SAS code to estimate the fixed and random effects models based on likelihood methods with the SAS program *proc mixed*. Estimates based on likelihood methods offer the advantage that they provide the option to formally test which model is appropriate for the data by applying the likelihood ratio test or penalized criteria such as the Bayesian information criterion (BIC). The BIC is obtained by the formula $\text{BIC} = -2 \times \log \text{Likelihood} + \log(k) \times q$ where q is the number of parameters in the model and k denotes the number of studies.

When using random effects models, another topic of interest is the form of the random effects' distribution. Besides a parametric distribution for the random effects, a discrete distribution may be assumed. Here, we suppose that the study-specific estimators $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ are coming from q subpopulations $\theta_j, j = 1, \dots, q$. Again, assuming that the effect of each individual study follows a normal distribution

$$f(\hat{\theta}_i, \theta_j, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(\hat{\theta}_i - \theta_j)^2}{2\sigma_i^2}}, \quad j = 1, \dots, q,$$

we obtain a finite mixture model

$$f(\hat{\theta}_i, P) = \sum_{j=1}^q f(\hat{\theta}_i, \theta_j, \sigma_i^2) p_j.$$

The parameters of the distribution P

$$P \equiv \begin{bmatrix} \theta_1 & \dots & \theta_q \\ p_1 & \dots & p_q \end{bmatrix} \quad \text{with} \quad p_j \geq 0, \quad j = 1, \dots, q,$$

$$p_1 + \dots + p_q = 1,$$

need to be estimated from the data. The mixing weights p_j denote the a priori probability of an observation of belonging to a certain subpopulation with parameter θ_j . Please note that also the number of components q needs to be estimated as well. Estimation may be done with the program C.A.MAN (Schlattmann and Böhning 1993; Böhning et al. 1998) or more recently with the R package CAMAN (Schlattmann 2009). For the HRT data, we find a solution with three components which gives an acceptable fit to the data

weight:	0.2804	parameter:	-0.3365
weight:	0.5671	parameter:	0.0778
weight:	0.1524	parameter:	0.5446

Log-Likelihood at iterate :	-17.6306
-----------------------------	----------

Table 36.1 Model comparison for the breast cancer data

Method	Residual	Estimates (SE)		log Lik.	BIC
	Hetero	Intercept	Het. ($\hat{\tau}^2$)		
Fixed	None	0.056 (0.023)	–	–33.19	70.0
Mixed	Additive	0.027 (0.061)	0.086	–18.65	44.4
FM	Additive		0.079	–17.63	53.2

Here, the weights correspond to the mixing weights p_j , and the parameter corresponds to the subpopulation mean θ_j . These results imply that about 28% of the studies show a protective effect of HRT, whereas the majority of the studies show a harmful effect. About 57% of the studies show an increased log(risk) of 0.08, and 15% of the studies show a log(odds ratio) of 0.54. Thus, using a finite mixture model (FM), we find again considerable heterogeneity where the majority of studies find a harmful effect of hormone replacement therapy. It is noteworthy that a proportion of studies find a beneficial effect. Of course, this needs to be investigated further. One way to do this would be to classify the individual studies using the finite mixture model. Doing so, we find that, for example, study nine from the data given in the Appendix A belongs to this category. This is a case-control study for which no information about confounder adjustment is available. This would be a starting point for a sensitivity analysis. Table 36.1 gives an overview about the models fitted so far. These include the fixed effects model with a BIC value of 70.0, the mixed effects model using a normal distribution for the random effects with a BIC value of 44.4. The finite mixture model (FM) has a BIC value of 53.2. Thus, based on Table 36.1, it is quite obvious that a fixed effects model does not fit the data very well and that a random effects model should be used. Of course, the question remains which random effects model to choose for the analysis. Based on the BIC criterion given in Table 36.1, one would choose the parametric mixture model provided the assumption of a normal distribution of the random effects is justifiable. This can be investigated, for example, by a normal quantile-quantile plot of the estimated individual random effects given by the parametric model. For the data at hand, the assumption of normally distributed random effects appears reasonable; thus, we would choose the parametric mixture model.

36.5.4.4 Further Aspects of Heterogeneity

It should be noted that in general random effects models yield larger variance and confidence intervals than fixed effects models because a between-study component τ^2 is added to the variance. If the heterogeneity between the studies is large, τ^2 will dominate the weights and all studies will be weighted more equally (in the random effects model weight decreases for larger studies compared to the fixed effects model).

Furthermore, pooling in the presence of heterogeneity may be seriously misleading. Heterogeneity between studies warrants careful investigation of the sources of the differences. If there are a sufficient number of different studies available,

further analyses, such as “meta-regression,” may be used to examine the sources of heterogeneity (Greenland 1987, 1994). One specific problem that occurs when binary endpoints are considered is that different risk measurements are used. While in case-control studies, an odds ratio is estimated, cohort studies yield an estimate for the relative risk. Although for studies in which rare events are investigated, odds ratios and relative risk estimates are very similar, this is not the case in studies where diseases with a higher prevalence are investigated. Odds ratio and relative risk differ and should not be combined. The problem that arises if studies with different designs are combined has not been well studied. We believe that no pooled estimate should be calculated combining data from case-control and cohort studies. There are many sources of heterogeneity, not only the size of risk estimates.

36.5.4.5 Meta-Regression

An important method for investigating heterogeneity is sensitivity analysis, for example, to calculate pooled estimators only for subgroups of studies (according to study type, quality of the study, period of publication, etc.) to investigate variations of the odds ratio. An extension of this approach is meta-regression as proposed by Greenland (1987); see also Thompson and Sharp (1999). The principal idea of meta-regression is once heterogeneity is detected to identify sources of heterogeneity by inclusion of known covariates.

For the breast cancer meta-analysis example a potential covariate is study type, case-control studies may show different results than cohort studies due to different exposure assessment. For our data, case-control studies are coded as $x_{i1} = 0$, and cohort studies are coded as $x_{i1} = 1$.

The fixed effect model is now

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_i^2), \quad i = 1, \dots, k.$$

Here, we find that cohort studies identify an association between HRT and breast cancer based on the regression equation $\hat{\theta}_i = 0.0015 + 0.145$ for a cohort study. Obviously, cohort studies come to results different from case-control studies. Clearly, after adjustment for covariates, the question remains if there is still residual heterogeneity present. Again, we can analyze the data using a random effects model in this case with a random intercept:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + b_i + \epsilon_i, \quad b_i \sim N(0, \tau^2), \quad \epsilon_i \sim N(0, \sigma_i^2).$$

For this model, the regression equation for the fixed effects gives now for a cohort study $\hat{\theta}_i = -0.009 + 0.1080$, and the corresponding heterogeneity variance is estimated as $\hat{\tau}^2 = 0.079$.

Table 36.2 compares fixed and random effects models for the HRT data. The table shows models with and without an estimate for the slope. Model selection can be based again on the BIC criterion. Apparently based on the BIC criterion, both fixed

Table 36.2 Comparison of fixed and random effects models

Method	Residual	Estimates (SE)		Het. ($\hat{\tau}^2$)	-log Lik.	BIC
	Hetero	Intercept	Slope			
Fixed	None	0.056 (0.023)	–	–	–33.85	70.0
Fixed	None	0.0014 (0.029)	0.145 (0.046)	–	–28.36	63.9
Mixed	Additive	0.027 (0.061)	–	0.086	–18.65	44.4
Mixed	Additive	–0.009 (0.072)	0.108 (0.126)	0.079	–18.25	47.3

effects models do not fit the data very well since their BIC values are considerably higher than those of the random effects models. Please note that if only the fixed effects models would be considered, this meta-analysis would show that cohort studies show a harmful effect. Comparing the mixed effects models in [Table 36.2](#), the model with the covariate does not provide an improved fit of the data. The log-likelihood is only slightly larger, and penalizing the number of parameters leads to a larger BIC value for the mixed effect model with the covariate. Another interesting point is to compare the heterogeneity variance estimated by both models. Here, there is no substantial portion of heterogeneity explained by the covariate, since the heterogeneity variance is reduced to 0.079 from 0.086. From a statistical point of view, further covariates need to be identified and included into the model. From a public health point of view, the conclusion is perhaps less straightforward. Although inclusion of the covariate study type does not explain the heterogeneity of the studies very well, we find that cohort studies find a harmful effect. One might argue that although these results are far from perfect, they should not be ignored as absence of evidence does not imply evidence of absence. Looking back at these data in the light of the results from the woman health initiative (WHI) study (Rossouw et al. 2002), it becomes clear that caution is required in the analysis and interpretation of meta-analyses of observational studies. The major finding of the WHI study was that the group of subjects undergoing treatment with combined HRT in the form of Prempro (0.625 mg/day conjugated equine estrogens (CEE) +2.5 mg/day medroxyprogesterone acetate) was found to have increased risk of breast cancer (hazard ratio = 1.26, 95% CI: 1.00–1.59) and no apparent cardiac benefit. This is contradictory to the prior belief that HRT provides cardiovascular benefit. As a result, although several benefits were considered, these interim findings at 5 years were deemed sufficiently troubling to stop this arm of the trial at 5.2 years.

36.6 Interpretation of the Results of Meta-Analysis of Observational Studies

The example from above shows that the interpretation of the results of a meta-analysis should not only discuss the pooled estimator and the confidence interval but should focus on the examination of the heterogeneity between the results of the studies. Strength and weaknesses as well as potential bias should be discussed.

36.6.1 Bias

For epidemiological studies in general, the main problem is not the lack of precision and the random error but the fact that results may be distorted by different sources of bias or confounding; for a general overview of the problem of bias, see Hill and Kleinbaum (2000). That means that the standard error (or the size of the study) may not be the best indicator for the weight of a study. If more or better data are collected on a smaller amount of subjects, results may be more accurate than in a large study with insufficient information on the risk factors or on confounders. The assessment of bias in individual studies is therefore crucial for the overall interpretation.

The central problem of meta-analyses of clinical trials is publication bias that has already been a topic in a paper by Berlin et al. as early as 1989 and is still a topic of recent methodological investigations (see, e.g., Copas and Shi 2001). This bias has received a lot of attention particularly in the area of clinical trials. As mentioned in Sect. 36.5.2, publication bias yields in general a non-negligible overestimation of the risk estimate. However, as Morris (1994) has pointed out, there exist little systematic investigations of the magnitude of the problem for epidemiological studies. A major worry is that non-significant results are neither mentioned in the title nor in the abstract and publications and may be lost in the retrieval process.

36.6.2 Confounding

Another problem arises because different studies adjust for different confounding factors. It is well known that the estimated effect of a factor of interest is (strongly) influenced by the inclusion or exclusion of other factors, in the statistical model if these factors have an influence on the outcome and if they are correlated with the risk factor of interest. Combining estimates from several studies with different ways of adjusting for confounders yields biased results. Using literature data only, crude estimates may be available for some of the studies, model-based estimates for others. However, as the adjustment for confounders is an important issue for the assessment of an effect in each single study, it is obvious that combining these different estimates in a meta-analysis may not give meaningful results. It is necessary to use “similar” confounders in each study to adjust the estimated effect of interest in the single studies. In general, that would require a reanalysis of the single studies. Obviously, that requires the original data and a MAP is needed for this purpose.

36.6.3 Heterogeneity

In epidemiological research, different study designs are in use, and none of them can be considered as a gold standard as the randomized clinical trial for therapy

studies. Therefore, it is necessary to evaluate the comparability of the single designs before summarizing the results. Often, case-control studies, cohort studies, and cross-sectional studies are used to investigate the same questions, and results of those studies need to be combined. Egger et al. (2001b) pointed out several examples in which results from case-control studies differ from those of cohort studies. For example, in a paper by Boyd et al. (1993), it was noted that cohort studies show no association between breast cancer and saturated fat intake, while the same meta-analysis using results from case-control studies only revealed an increased, statistically significant risk. Other reasons for heterogeneity may be different uses of data collection methods, different control selection (e.g., hospital vs. population controls), and differences in case ascertaining. Differences could be explored in a formal sensitivity analysis but also by graphical methods (funnel plot). However, meta-analyses from published data provide only limited information if the reasons for heterogeneity shall be investigated in depth.

The problem of heterogeneity can be well demonstrated with nearly any example of published meta-analysis. For example, Ursin et al. (1995) investigated the influence of the body mass index (BMI) on the development of premenopausal breast cancer. They included 23 studies of which 19 are case-control studies and 4 are cohort studies. Some of these studies were designed to investigate BMI as risk factor, others measured BMI as confounders in studies investigating other risk factors. It can only be speculated that the number of unpublished studies in which BMI was mainly considered as a confounder and did not show a strong influence on premenopausal breast cancer is non-negligible and that this issue may result in some bias. As is usual practice in epidemiological studies, relative risks were provided for several categories of BMI. To overcome this problem, the authors estimated a regression coefficient for the relative risk as a function of the BMI; however, several critical assumptions are necessary for this type of approach. The authors found severe heterogeneity across all studies combined (the p -value of a corresponding test was almost zero). An influence of the type of study (cohort study or case-control study) was apparent. Therefore, no overall summary is presented for case-control and cohort studies combined. One reason for the heterogeneity may be the variation in adjustment for confounders. Adjustment for confounders other than age was used only in 10 out of the 23 studies.

36.7 Network Meta-Analysis

Network meta-analyses are a recent development in meta-analyses, especially for randomized clinical trials, and are also referred to as multiple treatments meta-analyses. They are applied when the difference between two or more medical treatments should be investigated, but no or only few head-to-head comparison studies are available. In this situation, studies comparing the treatments of interest with a common comparator can be used for relative effect inferences. For example, there are randomized studies comparing treatment B versus standard treatment

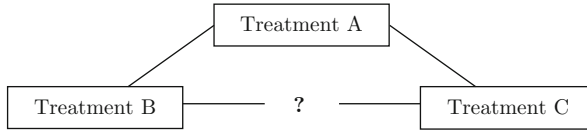


Fig. 36.7 Indirect comparison between treatment *B* and treatment *C*

A and furthermore randomized studies comparing treatment *C* versus treatment *A* (shown graphically in Fig. 36.7). A previously unresearched effect of treatment *B* versus treatment *C* can then be evaluated by an indirect comparison using the available studies.

36.7.1 Statistical Methods

In the setting of Fig. 36.7, let $\hat{\theta}_{BA}$ and $\hat{\theta}_{CA}$ be the estimated summary effects of the corresponding studies. These effects, for example, log odds ratios for binary outcome or differences in means for continuous data, can be combined in a network meta-analysis as follows:

$$\hat{\theta}_{BC} = \hat{\theta}_{BA} - \hat{\theta}_{CA}.$$

Extending this network, different paths for the same treatment comparison can also exist.

If further randomized studies were available comparing treatment *D* with treatments *B* and *C*, respectively (see Fig. 36.8), the effect θ_{BC} could be estimated on a second way:

$$\hat{\theta}_{BC} = \hat{\theta}_{BD} - \hat{\theta}_{CD}.$$

The network of different treatment comparisons can be arbitrarily extended to use all available evidence. Accordingly, several treatments can be compared simultaneously based on pairwise comparative or multiarm studies. If direct and indirect comparisons are included in a network meta-analysis, it is also named as mixed treatment comparisons (MTC) meta-analysis (Lu and Ades 2004). For example, Elliott and Meyer (2007a) investigated the effect of five antihypertensive drugs and placebo on incident diabetes mellitus based on 22 clinical trials with comparisons between two and three treatments. Figure 36.9 shows the network with the summary estimates for pairwise comparisons of traditional meta-analyses and each included study. None of these 22 trials performed a direct comparison between angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARB).

The effect estimates used for indirect comparisons or the combination of different network paths may be based on fixed or random effects models. In the simplest case, independent estimates of different paths can be synthesized by an average, weighted by the inverse of their respective variances. Furthermore, Bayesian methods and hierarchical models can be used to analyze indirect comparisons as well as more complex data structures such as multiarm trials, especially when

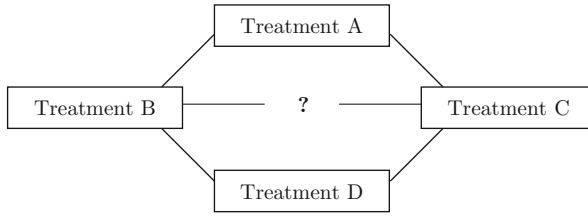


Fig. 36.8 Simple network meta-analysis

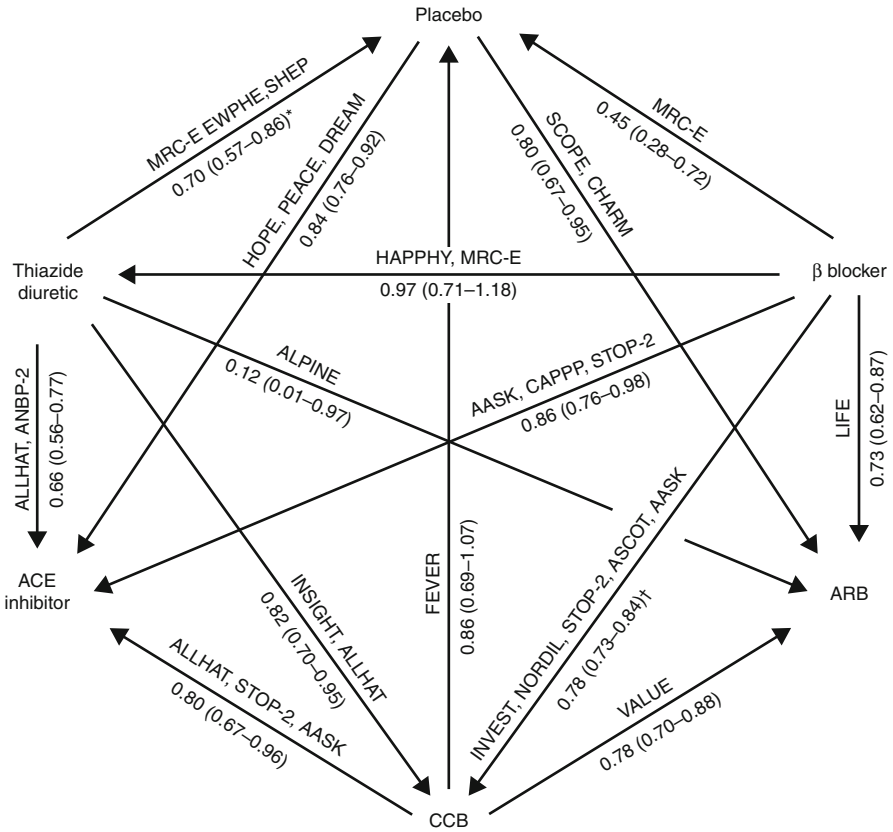


Fig. 36.9 A network of antihypertensive drug clinical trials (Reprinted from Elliott and Meyer 2007b, with permission from Elsevier)

additional information is available (Lu and Ades 2004). Adjustments for covariates and the use of meta-regression methods are also possible in network meta-analyses. Nixon et al. (2007) presented an MTC Bayesian meta-regression model to estimate the efficacy of biological treatments in rheumatoid arthritis adjusted for the both study-level covariates, disease duration and severity of disease.

36.7.2 Limitations

The reasons for potential bias, as already discussed for common meta-analysis, can also lead to distortions in network meta-analysis. But since an indirect comparison requires at least two meta-analyses, there is assumedly more scope for error.

For a valid indirect comparison, the studies should be methodologically and clinically comparable. Also, it is important to synthesize the evidence from different randomized clinical trials by comparing their treatment effects, such as log odds ratios. Considering only single treatment arms of the original controlled studies would break the randomization (Bucher et al. 1997). For example, the observations of treatment arm *B* and treatment arm *C* of the original by treatment *A* controlled studies (see Fig. 36.7) should not be pooled. Although the data originate from randomized controlled trials, they are reduced to the equivalent of those that come from an observational study. The comparison of effectiveness can therefore be biased by several confounders. Estimates of indirect comparisons based only on single treatment arms of randomized controlled trials are sometimes referred to as “unadjusted indirect estimates” and should be avoided.

In a network meta-analysis, not only the heterogeneity in treatment effects between studies of each pairwise contrast should be discussed, but also different network paths for the same comparison. For example, the estimations $\hat{\theta}_{BA} - \hat{\theta}_{CA}$ and $\hat{\theta}_{BD} - \hat{\theta}_{CD}$ in Fig. 36.8 can disagree. Particularly in MTC meta-analyses, inconsistencies in treatment effects obtained from direct and indirect comparisons of the same treatments may exist (Song et al. 2003). Baker and Kramer (2002) highlighted examples where the transitivity assumption between a direct and an indirect comparison is not given, for example, due to differences in patient or trial characteristics. In addition, imperfections in study design or analysis may lead to inconsistencies. For example, the observed treatment effect in open-label or ineffectively blinded studies might depend on whether the treatment is new or used as an active control which may alter the expectation of researchers and patients. This may in turn influence the observed treatment effect (see Lumley 2002). Song et al. (2009) surveyed systematic reviews published between 2000 and 2007 with indirect comparisons and investigated basic assumptions and methodological problems in the application. The assumption of consistency was explicitly mentioned in only 18 of the 30 included reviews where direct and indirect estimates were compared or combined.

The investigation of inconsistency, or also known as incoherence, is possible when there are closed loops such as in the connected polygonal network in Fig. 36.9. Independent estimates of different network paths can be tested, for example, with the simple χ^2 -test like in the testing of heterogeneity (see Sect. 36.5.4 and Caldwell et al. 2009). Lumley (2002) offered a quantitative estimation procedure using linear mixed models, in which not only the heterogeneity but also the inconsistency can be fitted as a further random effect. Lu and Ades (2006) set the estimation of inconsistency in the framework of Bayesian hierarchical models. However, large inconsistency needs, as large heterogeneity, a closer look for possible reasons. The adjustment for covariates and the use of meta-regression methods are another

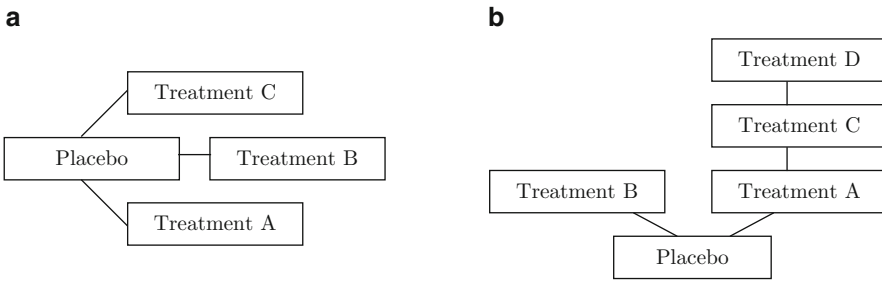


Fig. 36.10 Different network geometries: (a) radiating star (b) linear structure

possibility to address the inconsistency and also the heterogeneity within comparisons (Cooper et al. 2009). But since inconsistency is a property of loops, it is not easily possible to ascertain which specific contrast is responsible for the inconsistency, and it is an open question to which extent unexplained inconsistency is acceptable in an MTC meta-analysis. For a radiating or linear network meta-analysis structure, the estimation of inconsistency is impossible. Pattern (a) in Fig. 36.10 arises, for example, if there are only placebo-controlled studies and the active treatments are not compared to each other. If new treatments are tested only against the current standard, pattern (b) in Fig. 36.10 might result. Fewer cross-links and longer paths, for example, between treatment *B* and treatment *C* in Fig. 36.10, lead to greater uncertainty of the treatment effect estimation. But the information in these networks can be used, for example, in the planning of future trials to directly test the comparison where no direct evidence exists or to enhance the network. Salanti et al. (2008) discussed network geometry and asymmetry in this context, namely, that specific treatments or comparisons are represented more heavily than others.

Comparing the estimation accuracy between direct and indirect comparisons shows that under the assumption of statistical independence of different studies and consistent estimation accuracy, the variance of the effect estimator in indirect comparisons is greater than in direct comparisons. Assuming, for example, that the effect estimates of all studies have the same variance σ^2 , the variance of the fixed-effect inverse variance estimator of the direct comparisons based on $2k$ studies would be

$$\text{var}(\hat{\theta}_{BC\text{direct}}) = \sigma^2/2k.$$

However, the variance of the indirect comparison based on k studies for each comparison would be

$$\text{var}(\hat{\theta}_{BC\text{indirect}}) = \text{var}(\hat{\theta}_{BA\text{direct}}) + \text{var}(\hat{\theta}_{CA\text{direct}}) = \sigma^2/k + \sigma^2/k = 2\sigma^2/k.$$

Thus, for the indirect approach, four times as many studies are needed to have the same power as the direct comparisons provide (Glenny et al. 2005).

Consequently, indirect comparisons are not preferable to direct comparisons, but they are useful in planning direct comparisons and where head-to-head comparisons

are not possible. Both can be combined to include all the existing evidence and to obtain more precise estimates when direct comparisons are not sufficiently available (Higgins and Whitehead 1996). Song et al. (2008) hypothesized as to whether adjusted indirect comparisons may provide less biased results than direct comparisons under certain circumstances. For example, when an indirect comparison is based on studies with the same extent of bias, adjusted indirect comparison could counterbalance this. However, discrepancies between both approaches should be carefully investigated for potential explanations whenever possible.

36.8 Conclusions

Despite the many problems, there is an immense need to summarize current knowledge, for example, to assess the consequence of human exposure to environmental exposure. For this task, all available data and information will be needed, and meta-analysis is becoming increasingly influential. Particularly where the previously conducted epidemiological studies have provided inconsistent results, a meta-analysis may give some insight. As discussed, a major impediment for meta-analysis of epidemiological data is the heterogeneity across studies in their design, data collection methods, and analyses performed. The statistical combination of risk estimates should not be the central component of a meta-analysis using published data. An expert group in cooperation with the US Environmental Protection Agency was recently established to discuss the use of meta-analyses in environmental health studies. One of the objectives of this group was also to develop a consensus on “when meta-analysis should or should not be used” (Blair et al. 1995). There is always a danger that meta-analysis of observational studies produces precise looking estimates which are severely biased. This should be kept in mind as more and more public health regulators and decision-makers may rely on the results of a meta-analysis.

Appendices

A. Data

The listing shows the effect measure on the log-scale, the corresponding variance and the study type of each of the 36 studies analyzed in the meta-analysis by Sillero-Arenas et al. (1992):

```
data sillar;
input study or est type;
cards;
1 0.10436 0.299111 0
2 -0.03046 0.121392 0
3 0.76547 0.319547 0
```

```
4 -0.19845 0.025400 0
5 -0.10536 0.025041 0
6 -0.11653 0.040469 0
7 0.09531 0.026399 0
8 0.26236 0.017918 0
9 -0.26136 0.020901 0
10 0.45742 0.035877 0
11 -0.59784 0.076356 0
12 -0.35667 0.186879 0
13 -0.10536 0.089935 0
14 -0.31471 0.013772 0
15 -0.10536 0.089935 0
16 0.02956 0.004738 0
17 0.60977 0.035781 0
18 -0.30111 0.036069 0
19 0.01980 0.024611 0
20 0.00000 0.002890 0
21 -0.04082 0.015863 0
22 0.02956 0.067069 0
23 0.18232 0.010677 0
24 0.26236 0.017918 1
25 0.32208 0.073896 1
26 0.67803 0.489415 1
27 -0.96758 0.194768 1
28 0.91629 0.051846 1
29 0.32208 0.110179 1
30 -1.13943 0.086173 1
31 -0.47804 0.103522 1
32 0.16551 0.004152 1
33 0.46373 0.023150 1
34 -0.52763 0.050384 1
35 0.10436 0.003407 1
36 0.55389 0.054740 1
run;
```

B. Elementary Analysis with SAS

SAS code for the elementary analysis using weighted least squares:

```
/* calculation of weights */
data sillar;
set sillar;
weight =1./est;
run;
```

```

/* intercept only */
proc glm data=sillar; /* use proc GLM with data set sillar */
model logor=/solution inverse ; /* Show solution */
/* Show inverse of weighted design matrix */
weight weight; /* weights 1./variance */
run;

```

This gives the following shortened output:

```

The GLM Procedure
Dependent Variable: logor
Weight: weight

```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.1683128	6.1683128	1.86	0.1813
Error	35	116.0756869	3.3164482		
Un.Total	36	122.2439997			

Parameter	df	Estimate	SE	t Value	Pr > t
Intercept	1	0.0559813731	0.04104847	1.36	0.1813

Please note that for performing a meta-analysis the standard error given by the program must be divided by the root mean square error in order to obtain the standard error of the pooled estimate. In order to avoid additional calculations the SAS output giving the inverse of the weighted design matrix gives the desired variance. The test of heterogeneity is given by the residual sum of squares as indicated by formula (36.4). This result can also be obtained using the SAS code for the fixed effect model based on maximum likelihood:

```

proc mixed method=ml data=sillar; /* Use proc mixed (ML
estimation) */
class study; /* Specifies study as 'classification
variable' */
model or=/ s cl; /* Intercept only model, show
solution and CI */
repeated /group =study; /* Each trial has its own
within trial variance */
parms /parmsdata=sillar /* The parmsdata option reads
in the variable EST indicating the variances from
the data set sillar.sd2 */
eqcons=1 to 36; /* The within study variances
are known and fixed */
run;

```

C. SAS Code for the Random Effects Model

The SAS procedure `proc mixed` requires the following manipulations of the data

```
data covvars;          /* data set containing the variances */
set sillar;
keep est;
run;
data start;           /* include the starting value for the */
input est;           /* heterogeneity variance */
cards;
0.0
run;
data start;           /* Combine both data sets */
set start covvars;
run;
```

Obtain the model with `proc mixed`

```
proc mixed method=ml cl data=sillar; /* CL asks for confidence intervals */
/* of covariance parameters */
class study;          /* Study is classification variable */
model or= / s cl;    /* Intercept only model, Fixed solution and CI */
random int /subject=study ; /* Study is specified as random effect */
repeated /group =study; /* Each study has its own variance */
parms /parmsdata= start /* start contains starting value a. trial vars. */
eqcons=2 to 37;      /* entries 2 to 37 are the fixed study vars. */
run;
```

References

- Allam MF, Del Castillo AS, Navajas RF (2003) Parkinson's disease, smoking and family history: meta-analysis. *Eur J Neurol* 10:59–62
- Baker SG, Kramer BS (2002) The transitive fallacy for randomized trials: if a bests b and b bests c in separate trials, is a better than c? *BMC Med Res Methodol* 2(1):13
- Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50:1088–1101
- Bennett DA (2003) Review of analytical methods for prospective cohort studies using time to event data: single studies and implications for meta-analysis. *Stat Methods Med Res* 12:297–319
- Berlin JA (1995) Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 142:383–387
- Berlin JA, Begg CB, Louis TN (1989) An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 84:381–392
- Berlin JA, The University of Pennsylvania Meta-Analysis Blinding Study Group (1997) Does blinding of readers affect the results of meta-analyses? *Lancet* 350:185–186
- Blair A, Burg J, Floran J, Gibb H, Greenland S, Morris R, Raabe G, Savitz D, Teta J, Wartenberg D, Wong O, Zimmerman R (1995) Guidelines for application of meta-analysis in environmental epidemiology. *Regul Toxicol Pharmacol* 22:189–197
- Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C (1999) Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 28:1–9
- Boffetta P (2002) Involuntary smoking and lung cancer. *Scand J Work Environ Health* 28(Suppl 2):30–40

- Boffetta P, Saracci R, Andersen A, Bertazzi PA, Chang-Claude J, Cherrie J, Ferro G, Frentzel-Beyme R, Hansen J, Plato N, Teppo L, Westerholm P, Winter PD, Zocchetti C (1997) Cancer mortality among man-made vitreous fiber production workers. *Epidemiology* 8:259–268
- Böhning D, Dietz E, Schlattmann P (1998) Recent developments in computer assisted mixture analysis. *Biometrics* 54:283–303
- Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A (2002) Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator. *Biostatistics* 3:445–457
- Boyd NF, Martin LJ, Noffel M, Lockwood GA, Trichler DL (1993) A meta-analysis of studies of dietary fat and breast cancer. *Br J Cancer* 68:627–636
- Bucher HC, Guyatt GH, Griffith LE, Walter SD (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 50(6):683–691
- Caldwell DM, Welton NJ, Ades AE (2009) Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol*. doi:10.1016/j.jclinepi.2009.08.025
- CGHFBC Collaborative Group On Hormonal Factors Breast Cancer (1996) Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 347:1713–1727
- Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McLroy W, Oxman AD (1993) Should unpublished data be included in meta-analyses? current confictions and controversies. *J Am Med Assoc* 21:2749–2753
- Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ (2009) Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 28(14):1861–1881
- Copas JB, Shi JQ (2001) A sensitivity analysis for publication bias in systematic review. *Med Res* 10:251–265
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7:177–188
- Dickersin K (2002) Systematic reviews in epidemiology: why are we so far behind? *Int J Epidemiol* 31:6–12
- Dickersin K, Scherer R, Lefebvre C (1994) Identifying relevant studies for systematic reviews. *BMJ* 309:1286–1291
- Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a single, graphical test. *BMJ* 315:629–634
- Egger M, Schneider M, Davey Smith G (1998) Spurious precision? meta-analysis of observational studies. *BMJ* 316:140–144
- Egger M, Davey Smith G, Altman DG (2001a) Systematic reviews in health care. Meta-analysis in context, 2nd edn. BMJ Publishing Group, London
- Egger M, Davey Smith G, Schneider M (2001b) Systematic reviews of observational studies. *BMJ Publishing Group, London*, pp 211–227
- Egger M, Juni P, Bartlett C, Hohenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? An empirical study. *Health Technol Assess* 7:1–76
- Elliott WJ, Meyer PM (2007a) Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 369(9557):201–207
- Elliott WJ, Meyer PM (2007b) Incident diabetes in clinical trials of antihypertensive drugs. *Lancet* 369:1514–1515
- Feinstein AR (1995) Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 48:71–79
- Friedenreich CM (1993) Methods for pooled analyses of epidemiologic studies. *Epidemiology* 4:295–302
- Friedenreich C (1994) Influence of methodologic factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber. *Epidemiology* 5:66–67
- Glass GV (1977) Integrating findings: the meta-analysis of research. *Rev Res* 5:3–8

- Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, Damico R, Bradburn M (2005) Indirect comparisons of competing interventions. *Health Technol Assess* 9(26):1–134
- Greenland S (1987) Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 9:1–302
- Greenland S (1994) Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 140:290–296
- Hamajima N, Hirose K, Tajima K (2002) Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer* 87:1234–1245
- Higgins JPT, Green S (2009) *Cochrane handbook for systematic reviews of interventions*, Version 5.0.2. The Cochrane Collaboration, London
- Higgins JP, Whitehead A (1996) Borrowing strength from external trials in a meta-analysis. *Stat Med* 15(24):2733–2749
- Hill AB (1965) The environment and disease: association or causation? *Proc R Soc Med* 58:295–300
- Hill HA, Kleinbaum DG (2000) Bias in observational studies. Wiley, Chichester, pp 94–100
- Hung RJ, Boffetta P, Brockmoller J (2003) CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in caucasian non-smokers: a pooled analysis. *Carcinogenesis* 24:875–882
- Jones DR (1992) Meta-analysis of observational epidemiologic studies in a consistent form. *J R Soc Med* 85:165–168
- Longnecker MP, Berlin JA, Orza MJ, Chalmers TC (1988) A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 260:652–656
- Lu G, Ades AE (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 23(20):3105–3124
- Lu G, Ades AE (2006) Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* 101(474):447–459
- Lubin JH, Boice JD Jr, Edling C, Hornung RW, Howe G, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM, Timarche M, Woodward A, Yao SX (1995) Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Phys* 69:494–500
- Lumley T (2002) Network meta-analysis for indirect treatment comparisons. *Stat Med* 21(16):2313–2324
- Macaskill PS, Walter SD, Irwig L (2001) A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 20:641–654
- Meinert R, Michaelis J (1996) Meta-analyses of studies on the association between electromagnetic fields and childhood cancer. *Radiat Environ Biophys* 35:11–18
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *The Cochrane Collaboration*. *PLoS Med* 6(7):e1000097. doi:10.1371/journal.pmed.1000097
- Nixon RM, Bansback N, Brennan A (2007) Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* 26(6):1237–1254
- Normand SL (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 18:321–359
- Olkin I (1994) Invited commentary: Re: “a critical look at some popular meta-analytic methods”. *Am J Epidemiol* 140:297–299
- Paul SR, Donner A (1989) A comparison of tests of homogeneity of odds ratios in k 2x2 tables. *Stat Med* 8:1455–1468
- Peters J, Sutton A, Jones D, Abrams K, Rushton L (2006) Comparison of two methods to detect publication bias in meta-analysis. *The Cochrane Collaboration*. *JAMA* 295(6):676–680
- Pettiti DB (1994) *Meta-analysis, decision analysis and cost-effectiveness analysis*. Oxford University Press, New York
- Morris RD (1994) Meta-analysis in cancer epidemiology. *Environ Health Perspect* 102:61–66

- Riley RD, Simmonds MC, Look MP (2007) Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *J Clin Epidemiol* 60(5):431–439. doi:10.1016/j.jclinepi.2006.09.009
- Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J, Writing Group for the Women's Health Initiative Investigators (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *JAMA* 288:321–333
- Salanti G, Higgins JP, Ades AE, Ioannidis JP (2008) Evaluation of networks of randomized trials. *Stat Methods Med Res* 17(3):279–301
- Sauerbrei W, Blettner M, Royston P (2001) Letter to White, IR (1999): "The level of alcohol consumption at which all-cause mortality is least". *J Clin Epidemiol* 54:537–538
- Schlattmann P (2009) Medical applications of finite mixture models. *Statistics for biology and health*. Springer, Berlin, 246p. EUR 64.15
- Schlattmann P, Böhning D (1993) Computer packages c.a.man (computer assisted mixture analysis) and dismap. *Stat Med* 12:1965
- Schlehofer B, Blettner M, Preston-Martin S, Niehoff D, Wahrendorf J, Arslan A, Ahlbom A, Choi WN, Giles GG, Howe GR, Little J, Menegoz F, Ryan P (1999) Role of medical history in brain tumour development. Results from the international adult brain tumour study. *Int J Cancer* 82:155–160
- Schwarzer G, Antes G, Schumacher M (2002) Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Stat Med* 21:2465–2477
- Shapiro S (1994a) Is there is or is there ain't no baby? Dr. Shapiro replies to Drs. Petitti and Greenland. *Am J Epidemiol* 140:788–791
- Shapiro S (1994b) Meta-analysis/shmeta-analysis. *Am J Epidemiol* 140:771–778
- Sidik K, Jonkman JN (2005) Simple heterogeneity variance estimation for meta-analysis. *Appl Stat* 54:367–384
- Sillero-Arenas M, Delgado-Rodriguez M, Rodriguesw-Canteras R, Bueno-Cavanillas A, Galvez-Vargas R (1992) Menopausal hormone replacement therapy and breast cancer: a meta-analysis. *Obstet Gynecol* 79:286–294
- Smith ML, Glass GV (1977) Meta-analysis of psychotherapy outcome studies. *Am Psychol* 32(9):752–760
- Smith-Warner SA, Ritz J, Hunter DJ, Albanes D (2002) Dietary fat and risk of lung cancer in a pooled analysis of prospective studies. *Cancer Epidemiol Biomark Prev* 11:987–992
- Song F, Altman DG, Glenny AM, Deeks JJ (2003) Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 326(7387):472
- Song F, Harvey I, Lilford R (2008) Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol* 61(5):455–463
- Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG (2009) Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 338:b1147. doi:10.1136/bmj.b1147
- Steinberg KK, Smith SJ, Lee N, Stroup DF, Olkin I, Williamson GD (1997) A comparison of meta-analysis to pooled analysis: an application to ovarian cancer. *Am J Epidemiol* 145:1917–1925
- Stock WA (1995) Systematic coding for research synthesis. The Russell Sage Foundation, New York, pp 1–2
- Straif K, Chambless L, Weiland SK, Wienke A, Bungers M, Taeger D, Keil U (1999) Occupational risk factors for mortality from stomach and lung cancer among rubber workers: an analysis using internal controls and refined exposure assessment. *Int J Epidemiol* 28:1037–1043
- Straif K, Keil U, Taeger D, Holthenrich D, Sun Y, Bungers M, Weiland SK (2000) Exposure to nitrosamines, carbon black, asbestos, and talc and mortality from stomach, lung, and laryngeal cancer in a cohort of rubber workers. *Am J Epidemiol* 152:297–306

- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology group. *J Am Med Assoc* 283:2008–2012
- Stukel TA, Demidenko E, Dykes J, Karagas MR (2001) Two-stage methods for the analysis of pooled data. *Stat Med* 20:2115–2130
- Sutton AJ, Higgins JP (2008) Recent developments in meta-analysis. *Stat Med* 27(5):625–650
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000) *Methods for meta-analysis in medical research*. Wiley, Chichester/New York
- Sutton AJ, Kendrick D, Coupland CA (2008) Meta-analysis of individual- and aggregate-level data. *Stat Med* 27(5):651–669
- Thompson SG (1994) Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 309:1351–1355
- Thompson SG, Sharp SJ (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 18:2693–2708
- Thompson SG, Smith TC, Sharp SJ (1997) Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med* 16:2741–2758
- Tweedie RL, Mengersen KL (1995) Meta-analytic approaches to dose-response relationships, with application in studies of lung cancer and exposure to environmental tobacco smoke. *Stat Med* 14:545–569
- Ursin G, Longenecker MP, Haile RW, Greenland S (1995) A meta-analysis of body mass index and risk of premenopausal breast cancer. *Epidemiology* 6:137–141
- van Houwelingen HC, Arends LC, Stijnen T (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 21:589–624
- Weed DL (1997) Methodologic guidelines for review papers. *JNCI* 89:6–7
- Weed DL (2000) Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *Int J Epidemiol* 29:387–390
- Whitehead A (2002) *Meta-analysis of controlled clinical trials*. Wiley, Chichester
- Zeeger MP, Jellema A, Ostrer H (2003) Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis. *Cancer* 97:1894–1903