Meta-Analysis in Epidemiology II.7

Maria Blettner, Peter Schlattmann

7.1 Introduction

The use of meta-analyses in order to synthesise 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 randomised clinical trials and nearly the same proportion (27%) were from observational studies, including 12% papers in which the aetiology 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. While methods for meta-analyses in order to summarise or synthesise evidence from randomised controlled clinical trials have been continuously developed during the last years, and methods are now summarised in several text books for example Sutton et al. (2000), Whitehead (2002) and in a handbook by Egger et al. (2001), 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 randomised 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 summarise 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.

7.2 Different Types of Overviews

Approaches for summarising evidence include four 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 co-operation 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 re-analysed. Often, there is a close co-operation 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 re-analysis 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 re-analysed (CGHFBC 1996). A further international collaboration led by Lubin and colleagues were set up to re-analyse data from eleven large cohort studies on lung cancer and radon among uranium miners. The re-analysis allowed a refined dose-response analysis and provided data for radiation protection issues. Pooled re-analyses are mostly performed by combining data from studies of the same type only. For example Hung et al. (2003) re-analysed data from all case-control studies in which the role of genetic polymorphisms for lung cancer in non-smokers were 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 methodologic aspects for a pooled analysis of data from cohort studies was recently published by Bennett (2003).

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 standardised 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 multicentre randomised clinical trials, important heterogeneity between the study centres 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 centres) 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 siutation sometimes neighbourhood 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 tumour study initiated by the IARC including data from eight different countries (see Schlehofer et al. 1999).

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 (cf. Chap. III.8 of this handbook).

7.3 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 leukaemia 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 categorised 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 leukaemia.

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 re-analysis. 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.

Steps in Performing a Meta-Analysis 7.4

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 (cf. Chap. III.8 of this handbook).

Step 1. Define a clear and focused topic for the review: As for any other investigation, a clear protocol in which the research hypothesis, i.e. 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, e.g. to exclude ecological studies. Friedenreich et al. (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 8933 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 key-words 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, e.g. 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 can not 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 summarising 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 intervals 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). 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.

Statistical Analysis 7.5

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 epidemiologic data. 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 summarise 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 the appendix.

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

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.

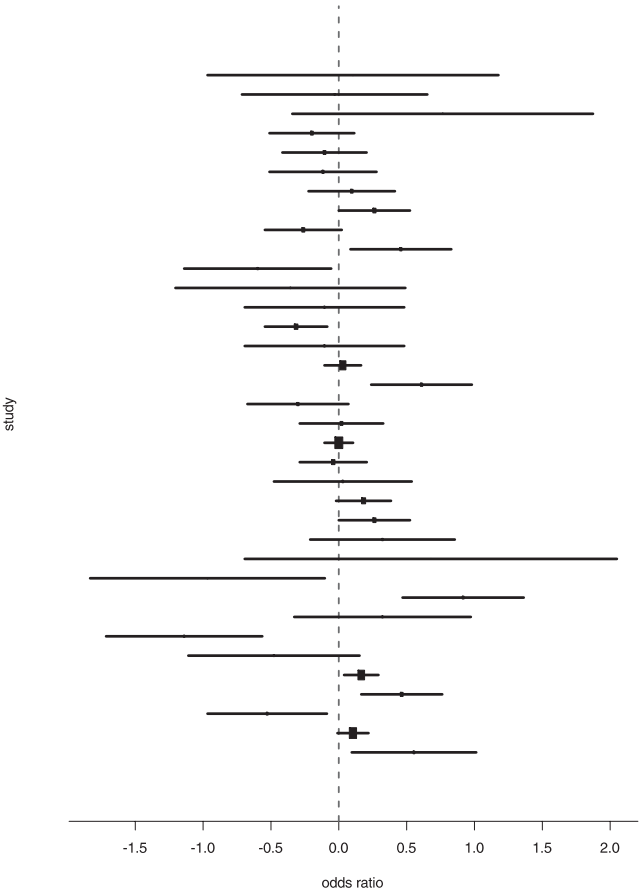

Figure 7.1. Confidence interval plot of the breast cancer data

Figure 7.1 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.

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 randomised 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). For epidemiologic observational studies additional problems exist, because often a large number of variables will be collected in questionnaires as potential confounders. If one or several of these potential confounders yield significant or important results, they 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. As a result prior to further statistical analyses publication bias should be investigated.

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 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.

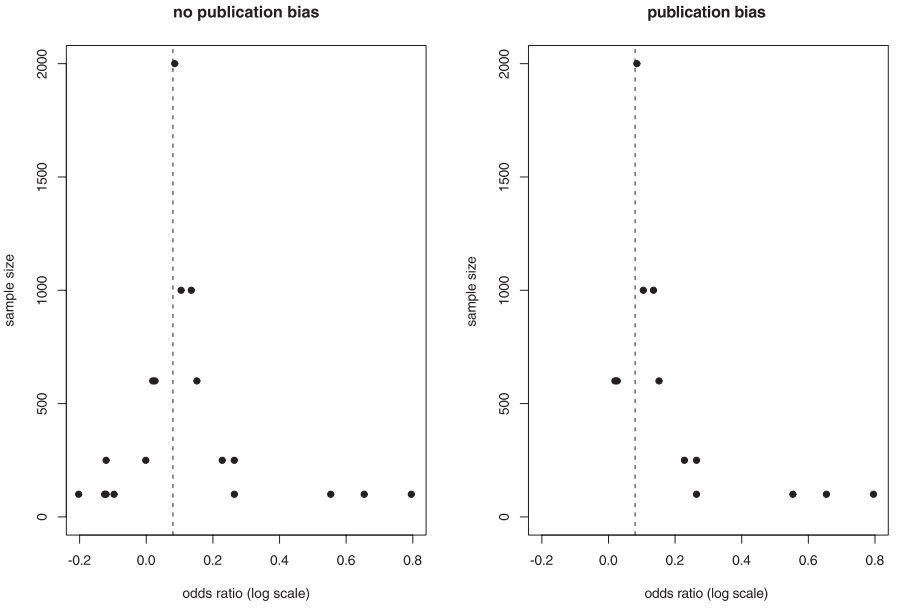

Figure 7.2. Examples of funnel plots based on simulated data with (*right figure*) and without publication bias present (*left figure*). The *dotted line* shows the true effect

Figure 7.2 shows examples of funnel plots. The left subplot of Fig. 7.2 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. 7.2 the presence of publication bias would be assumed.

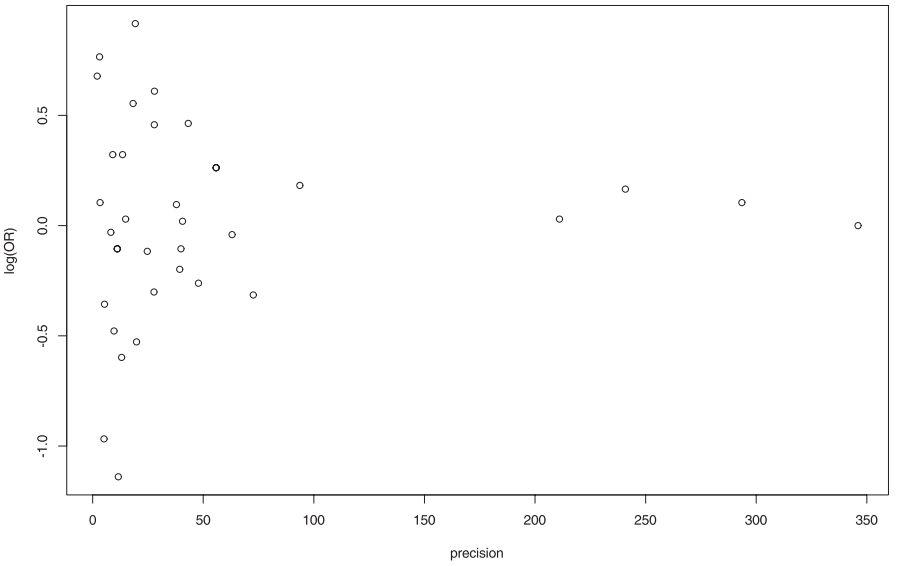

Figure 7.3. Funnel plot of the breast cancer data

Figure 7.3 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 Macaskil et al. (2001) favoured 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.

$$\hat{\theta}_i = \alpha + \beta \frac{1}{\sigma_i^2} + \varepsilon_i \,, \quad i = 1, \ldots, k \,, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right) \,. \tag{7.1}$$

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, i.e. 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. 7.4 shows no apparent slope. Likewise the model output (not shown) does not indicate the presence of publication bias for the data at hand.

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 variances from each study have to be abstracted from publications or calculated if data are available. Then, a combined

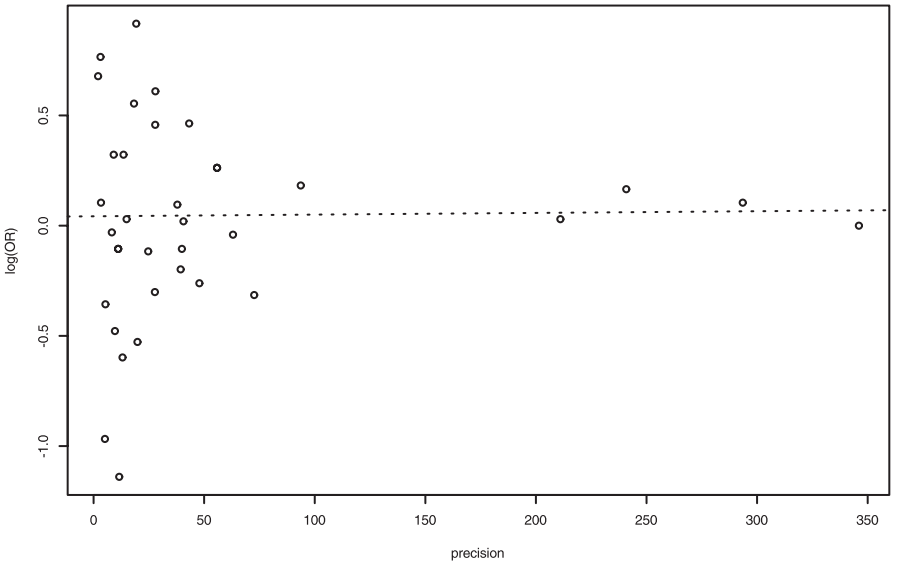

Figure 7.4. Funnel regression plot of the breast cancer data

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 Chap. I.2 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

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i \hat{\theta}_i}{\sum_{i=1}^{k} w_i}, \tag{7.2}$$

$$w_i = \frac{1}{\sigma_i^2}. \tag{7.3}$$

The variance is given by

$$\text{var}\left(\hat{\theta}\right) = \frac{1}{\sum_{i=1}^{k} 1/\sigma_i^2}. \tag{7.4}$$

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 percent 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, i.e. 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 7.5 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).

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 (Petitti 1994; Paul and Donner 1989). A simple test for heterogeneity is based on the following test statistic:

$$\chi^2_{het} = \sum_{i=1}^{k} \frac{(\hat{\theta} - \hat{\theta}_i)^2}{\sigma_i^2} \sim \chi^2_{k-1} \,, \tag{7.5}$$

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^2_{het} exceeds the $1-\alpha$ quantile of χ^2_{k-1} denoted as $\chi^2_{k-1,1-\alpha}$. For the data at hand we clearly conclude that there is heterogeneity present ($\chi^2_{het} = 116.076$, 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. The first part of the appendix 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

$$\theta_i = \theta + \varepsilon_i \,, \quad i = 1, \ldots, k \,, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right) \,. \tag{7.6}$$

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

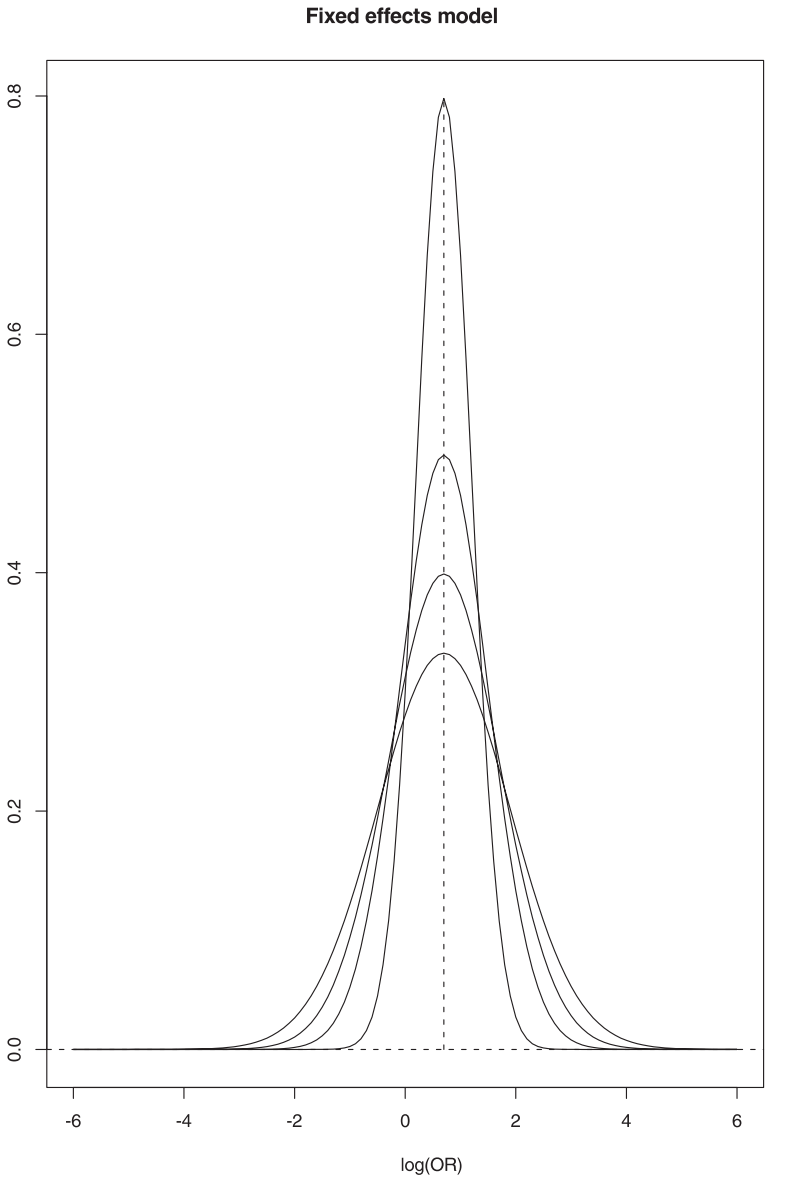

Figure 7.5. Fixed effects model: Common effect with different study variances

(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. 7.6. 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-

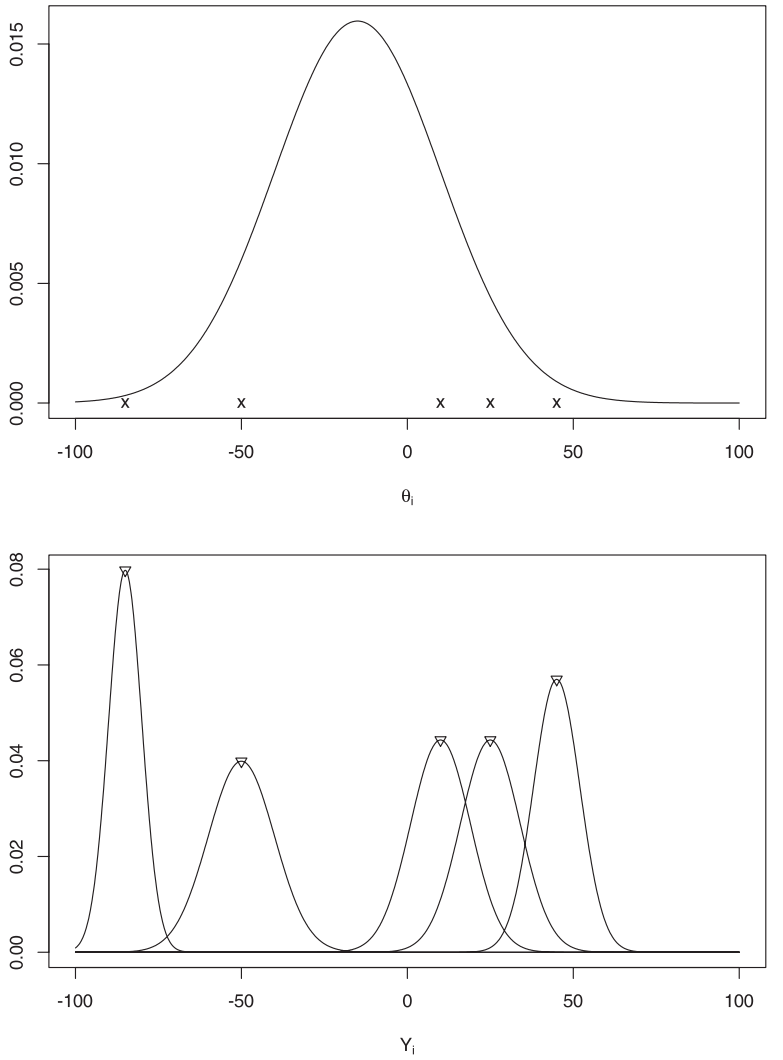

Figure 7.6. Random effects model: Variable effects drawn from a population of study effects

homogeneity between the effects of different studies. This leads to the following model:

$$\theta_i = \theta + b_i + \varepsilon_i \,, \quad i = 1, \ldots, k \,, \quad b_i \sim N\left(0, \tau^2\right) \,, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right) \,. \tag{7.7}$$

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. 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 (7.7) 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 (7.5) to its expected value. This expectation is calculated under the assumption of a random effects model and given by $E\left(\chi^2_{het}\right) = k-1+\tau^2\left(\sum w_i - (\sum w_i^2)/(\sum w_i)\right)$. The weights w_i are those defined in (7.3). Equating χ^2_{het} to its expectation and solving for τ^2 gives:

$$\hat{\tau}^2 = \left[\chi^2_{het} - (k-1)\right] \Big/ \left(\sum w_i - \frac{\sum w_i^2}{\sum w_i}\right) . \tag{7.8}$$

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

$$\hat{\theta}_{DL} = \frac{\sum_{i=1}^k w_i^* \hat{\theta}_i}{\sum_{i=1}^k w_i^*} , \tag{7.9}$$

$$\text{with} \quad w_i^{*^2} = \frac{1}{\sigma_i^{*^2}} = \frac{1}{\hat{\tau}^2 + \sigma_i^2} \quad \text{we obtain} \tag{7.10}$$

$$\hat{\theta}_{DL} = \frac{\sum_{i=1}^k \hat{\theta}_i / \left(\hat{\tau}^2 + \sigma_i^2\right)}{\sum_{i=1}^k 1 / \left(\hat{\tau}^2 + \sigma_i^2\right)} . \tag{7.11}$$

The variance of this estimator is given by:

$$\text{var}(\hat{\theta}_{DL}) = \frac{1}{\sum_{i=1}^k 1/\sigma_i^{*^2}} , \tag{7.12}$$

$$= \frac{1}{\sum_{i=1}^k 1/(\hat{\tau}^2 + \sigma_i^2)} . \tag{7.13}$$

The between study variance τ^2 can also be interpreted as a measure for the heterogeneity between studies. It should be noted that in general random effects methods 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 random effects model weight decreases for larger studies compared to the fixed effects model). 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.

However, two comments are in order. First, pooling in the presence of heterogeneity may be seriously misleading. Heterogeneity between studies should yield careful investigation of the sources of the differences. If there is 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). The second is in terms of statistical methodology. Within this approach the study specific variances are assumed to be known constants. That is the reason why this approach can lead to a considerable bias when pooling estimates using the DerSimonian–Laird estimator as demonstrated by Böhning et al. (2002).

Besides the moment based method by DerSimonian and Laird 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 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 BIC $= -2 \times \log$ 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 $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are coming from q subpopulations $\theta_j, j = 1, \ldots, q$. Again assuming that the effect of each individual study follows a normal distribution

$$f\left(\hat{\theta}_i, \theta_j, \sigma_i^2\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \mathrm{e}^{-(\hat{\theta}_i - \theta_j)^2/(2\sigma_i^2)}, \quad j = 1, \ldots, q\,. \tag{7.14}$$

we obtain a finite mixture model

$$f(\hat{\theta}_i, P) = \sum_{j=1}^{q} f\left(\hat{\theta}_i, \theta_j, \sigma_i^2\right) p_j\,. \tag{7.15}$$

The parameters of the distribution P

$$P \equiv \begin{bmatrix} \theta_1 & \ldots & \theta_q \\ p_1 & \ldots & p_q \end{bmatrix} \quad \text{with} \quad p_j \geq 0 \quad j = 1, \ldots, q\,, \tag{7.16}$$

$$p_1 + \ldots + p_q = 1\,. \tag{7.17}$$

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). 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
```

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 shows 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 finds a harmful effect of hormone replacement therapy. It is noteworthy that a proportion of studies finds 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 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 7.1 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 7.1 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 7.1 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.

Table 7.1. Model comparison for the breast cancer data

Method	Residual Hetero.	Estimates (SE) Intercept	Het. ($\hat{\tau}^2$)	log Lik.	BIC
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

Meta-Regression. An important method for investigating heterogeneity is sensitivity analysis, e.g. 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:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \,, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right) \,, \quad i = 1, \ldots, k \,. \tag{7.18}$$

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 form case-control studies. Clearly, after adjustment for covariates the question remains if there is still residual heterogeneity present. Again we can analyse the data using a random effects model in this case with a random intercept.

$$\theta_i = \beta_0 + \beta_1 x_{i1} + b_i + \varepsilon_i \,, \quad b_i \sim N\left(0, \tau^2\right) \,, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right) \,. \tag{7.19}$$

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 7.2. Comparison of fixed and random effects models

Method	Residual Hetero.	Estimates (SE) Intercept	Slope	Het. ($\hat{\tau}^2$)	-log Lik.	BIC
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

Table 7.2 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 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 7.2 the model with the covariate does not provide an improved fit of

the data. The log-likelihood is only slighty larger and penalising 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.

Interpretation of the Results of Meta-Analysis of Observational Studies 7.6

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.

Bias 7.6.1

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 an 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 for example Copas and Shi (2001)). 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 only some of the relevant studies will be identified (Dickersin et al. 1994). For epidemiological observational studies additional problems exist, because often a large number of variables will be collected in questionnaires as potential confounders (Blettner et al. 1999). If one or several of these potential confounders yield significant or important results, they 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.

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.

7.6.2 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 re-analysis of the single studies. Obviously, that requires the original data and a MAP is needed for this purpose.

7.6.3 Heterogeneity

In epidemiological research different study designs are in use and none of them can be considered as a gold standard as the randomised clinical trial for therapy studies. Therefore it is necessary to evaluate the comparability of the single designs before summarising 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. (2001) pointed out several examples in which results from case-control studies differ from those of cohort

studies. E.g. 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 pre-menopausal breast cancer. They include 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 pre-menopausal 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.

Conclusions 7.7

Despite the many problems, there is an immense need to summarise 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 co-operation with the U.S. 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.

Appendix

7.A Data and Computer Code and Output

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

```
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;
```

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
                Sum of
Source      DF  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 the 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 (7.5). 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;
   /* Specifes study as 'classificaton 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;
```

7.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 variable             */
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

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, on behalf of 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 et al. (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, Zochetti 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

CGHFBC-Collaborative Group on Hormonal Factors in 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, McLlroy W, Oxman AD (1993) Should unpublished data be included in meta-analyses? Current conflictions and controversies. JAMA 21:2749–2753

Copas JB, Shi JQ (2001) A sensitivity analysis for publication bias in systematic review. Stat Methods 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, Altman DG (2001) Systematic reviews in health care. Meta-analysis in context, 2nd edn. BMJ Publishing Group, London

Egger M, Davey Smith G, Schneider M (2001) Systematic reviews of observational studies. In: Egger M, Davey Smith G, Altman DG Systematic reviews in health care. Meta-analysis in context, 2nd edn, BMJ Publishing Group, London, pp 211–227

Egger M, Juni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 7:1–76

Egger M, Schneider M, Davey Smith G (1998) Meta-analysis: Spurious precision? Meta-analysis of observational studies. BMJ 361:140–144

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 CM, Braut RF, Riboli E (1994) Influence of methodologic factors in a pooled analysis of 13 case-control studies of colorectal cancer and dietary fiber. Epidemiology 5:66–79

Glass GV (1977) Integrating findings: The meta-analysis of research. Rev Res Educ 5:351–379

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 metaanalytic methods. Am J Epidemiol 140:290–296

Hamajima N, Hirose K, Tajima K et al. (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

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. In: Gail MH, Benichou J Encyclopedia of epidemiologic methods, John Wiley and Sons, Chichester, pp 94–100

Hung RJ, Boffetta P, Brockmoller J et al. (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: A review. 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

Lubin JH, Boice Jr JD, Edling C, Hornung RW, Howe G, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM, et al (1995) Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. Health Phys 69:494–500

Macaskill P, 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

Morris RD (1994) Meta-analysis in cancer epidemiology. Environ Health Perspect 102 Suppl 8:61–66

Normand SL (1999) Meta-analysis: Formulating, evaluating, combining, and reporting. Stat Med 18:321–359

Olkin I (1994) 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 $k2 \times 2$ tables. Stat Med 8:1455–1468

Pettiti DB (1994) Meta-analysis, decision analysis and cost-effectiveness analysis: Methods for quantitative synthesis in medicine. Oxford University Press, Oxford

Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, StefanickML, 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

Sauerbrei W, Blettner M, Royston P (2001) Letter to White IR (1999): On alcohol consumption and all-cause mortality. J Clin Epidemiol 54:537–540

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) Meta-analysis/shmeta-analysis. Am J Epidemiol 140:771–778

Shapiro S (1994b) Is there is or is there ain't no baby?: Dr Shapiro replies to Drs Petitti and Greenland. Am J Epidemiol 140:788–791

Sillero-Arenas M, Delgado-Rodriguez M, Rodigues-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 et al. (2002) Dietary fat and risk of lung cancer in a pooled analysis of prospective studies. Cancer Epidemiol Biomarkers Prev 11:987–992

Steinberg KK, Smith SJ, Stroup DF, Olkin I, Lee N, Williamson GD, Thacker SB (1997) 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. In: Cooper H, Hedges LV The Handbook of 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. JAMA 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, Abram KR, Jones DR, Sheldon TA, Song F (2000) Methods for meta-analysis in medical research. John Wiley and Sons, Chichester

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-analyis: multivariate approach and meta-regression. Stat Med 59:589–624

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. John Wiley and Sons, Chichester

Zeegers 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