Chapter 2
Fixed Effect and Random Effects Meta-Analysis

In this chapter we describe the two main methods of meta-analysis, fixed effect model and random effects model, and how to perform the analysis in R. For both models the inverse variance method is introduced for estimation. The pros and cons of these methods in various contexts have been debated at length in the literature [9, 28, 29, 41], without any conclusive resolution. Here, we briefly describe each model, and how it is estimated in the R package **meta** [33, 34].[1]

An estimated treatment effect and its variance from each study are sufficient to apply the inverse variance method. Therefore, this method is sometimes called the generic inverse variance method. For the random effects model, various methods to estimate the between-study variance, the Hartung–Knapp adjustment and prediction intervals are briefly described.

We also show how to use R to generate forest plots. Along the way, we will show how the tabular and graphical summaries usually included in Cochrane reviews can be generated in R. We give examples using both base R and functions provided by our R package **meta**. The various methods of meta-analysis are best illustrated using base R; furthermore some basic R knowledge is gained from working with fundamental R functions. The R code using functions from the R package **meta** shows how routine manipulations and calculations can be automated. In practice a meta-analyst would like to do the analyses using the more sophisticated functions in the R package **meta**. Accordingly, readers not interested in the mathematical details could run over the examples using base R functions.

We will use a continuous outcome to introduce both fixed effect and random effects model. Accordingly, we start by describing the two most common effect measures for continuous outcomes, mean difference and standardised mean

---

[1]If you did not already install R package **meta** do so using R command `install.packages("meta")`.

difference. In Sect. 2.6, the generic inverse variance method is applied in meta-analyses with survival outcome, cross-over trials and adjusted estimates from regression models.

2.1 Effect Measures for Continuous Outcomes

Meta-analysis typically focuses on comparing two interventions, which we refer to as *experimental* and *control*. When the response is continuous (i.e. quantitative) typically the mean, standard deviation and sample size are reported for each group. Let $\hat{\mu}_{ek}, s_{ek}^2, n_{ek}$ and $\hat{\mu}_{ck}, s_{ck}^2, n_{ck}$ denote the observed mean, standard deviation and sample size for study k, $k = 1, \ldots, K$ (see Table 2.1).

We consider two different types of effect measures for continuous outcomes: mean difference and standardised mean difference. The mean difference is typically used when all studies report the outcome on the same scale. On the other hand, the standardised mean difference can be used when studies measure the outcome on different scales, e.g. different depression scales like the Hamilton Depression Rating Scale or the Hospital Anxiety and Depression Scale.

2.1.1 Mean Difference

For study k, the estimated mean difference is

$$\hat{\mu}_k = \hat{\mu}_{ek} - \hat{\mu}_{ck}, \tag{2.1}$$

Table 2.1 Variable names in R datasets for meta-analyses of continuous responses

Variable name	Notation	Description
author		First author of study
year		Year study published (if available)
Ne	n_e	Number of patients in the experimental (i.e. active) treatment arm
Me	$\hat{\mu}_e$	Mean response in the experimental treatment arm
Se	s_e	Standard deviation of the response in the experimental treatment arm
Nc	n_c	Number of patients in the control (often equivalent to placebo) arm
Mc	$\hat{\mu}_c$	Mean response in the control arm
Sc	s_c	Standard deviation of the response in the control arm

with variance estimate[2]

$$\widehat{\mathrm{Var}}\left(\hat{\mu}_k\right) = \frac{s_{ek}^2}{n_{ek}} + \frac{s_{ck}^2}{n_{ck}} .\tag{2.2}$$

An approximate two-sided $(1-\alpha)$ confidence interval for the mean difference is given by

$$(\hat{\mu}_{ek} - \hat{\mu}_{ck}) \ \pm \ z_{1-\frac{\alpha}{2}} \ \sqrt{\frac{s_{ek}^2}{n_{ek}} + \frac{s_{ck}^2}{n_{ck}}}\tag{2.3}$$

with $z_{1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution. For the usual 95 % confidence interval, $z_{1-\frac{0.05}{2}} = z_{0.975} = 1.96$, i.e. the 97.5 % point of the standard normal distribution.

Example 2.1 We return to the meta-analysis by Spooner et al. [37] comparing Nedocromil sodium with placebo for preventing exercise-induced bronchoconstriction which we already used in Chap. 1. Outcome of interest is the maximum fall in the forced expiratory volume in 1 second (FEV_1) over the course of follow-up, expressed as a percentage. Accordingly, all studies report the same outcome and the use of the mean difference is warranted.

The raw data consist of eight variables with headings in Table 2.1. Code to read in the data, together with the data, are shown in Fig. 1.2. From the data we see that the meta-analysis contains 17 studies, with sample sizes ranging between 16 (Shaw 1985; DeBenedictis 1995) and 48 (Novembre 1994f).

For each study (labelled by first author and date) mean values, standard deviations and sample sizes are given in Fig. 1.2. Thus for study 1 (Boner 1988) the estimated mean difference is $13.54 - 20.77 = -7.23$ and for study 2 (Boner 1989) it is $15.70 - 22.70 = -7.00$ (see Fig. 1.4). Accordingly, the maximum fall in FEV_1 is on average about 7 % in Boner 1988 and Boner 1989. For study 1 (Boner 1988) the 95 % confidence interval (2.3) is

$$(13.54 - 20.77) \ \pm \ 1.96 \ \sqrt{\frac{13.85^2}{13} + \frac{21.46^2}{13}} \quad \text{giving} \quad (-21.11, 6.65).$$

We can use base R to calculate mean difference and 95 % confidence interval for the Boner 1988 trial (assuming that the file dataset01.csv is in the current working directory; see Sect. 1.2 for details):

```
> # 1. Read in the data
> data1 <- read.csv("dataset01.csv", as.is=TRUE)
> # 2. Calculate mean difference and its standard error for
```

---

[2]Note we could use a pooled estimate of the sample variance, but this assumes that the response variance is the same in the two groups which will not be true in general.

```
> #    study 1 (Boner 1988) of dataset data1:
> MD <- with(data1[1,], Me - Mc)
> seMD <- with(data1[1,], sqrt(Se^2/Ne + Sc^2/Nc))
> # 3. Print mean difference and limits of 95% confidence
> #   interval using round function to show only two digits:
> round(c(MD, MD + c(-1,1) * qnorm(1-(0.05/2)) * seMD), 2)
[1]  -7.23 -21.11   6.65
```

The values for mean difference, lower and upper limit of the 95 % confidence interval are identical to those calculated manually.

We can also use the `metacont` function from R package **meta** to calculate mean difference and confidence interval:

```
> with(data1[1, ],
+     print(metacont(Ne, Me, Se, Nc, Mc, Sc),
+           digits=2))
    MD        95%-CI     z  p-value
 -7.23 [-21.11; 6.65] -1.02   0.3074

Details:
- Inverse variance method
```

We get the same result by using the `metacont` function with argument `sm="MD"` (i.e. summary measure is the Mean Difference) as this is the default setting.

Note, the printout states that the inverse variance method has been used which strictly speaking refers to the method of meta-analysis, i.e. a setting with at least two studies. For a single study this simply means that Eqs. (2.1)–(2.3) have been used in the calculation of the mean difference and its confidence interval.

Instead of using the `with` function, a more convenient way is to use the `metacont` function with arguments `data` and `subset`.

```
> print(metacont(Ne, Me, Se, Nc, Mc, Sc,
+                data=data1, subset=1), digits=2)
    MD        95%-CI     z  p-value
 -7.23 [-21.11; 6.65] -1.02   0.3074
*** Output truncated ***
```

In addition to mean difference and its 95 % confidence interval, the `metacont` function reports z-score and *p*-value for the test of an overall treatment effect. These quantities can be calculated using base R functions `pnorm` and `abs` as follows:

```
> zscore <- MD/seMD
> round(c(zscore, 2*pnorm(abs(zscore), lower.tail=FALSE
[1] -1.0206  0.3074
```

When calling `metacont` we are matching up the first argument `n.e` of the `metacont` function with the variable `Ne` of the `Boner 1988` trial; and similarly for the other arguments. In order to access the data of the `Boner 1988` trial we use the argument `subset=1` which selects the first row of the dataset `data1`. A more

general way to select the `Boner 1988` trial which is not relying on the order of the dataset is `subset=(author=="Boner"&year=="1988")`.[3]

The argument `subset` can also be used to exclude some studies, e.g., `subset=-2` selects all but the second trial, `subset=author!="Boner"` excludes all trials from the author Boner, and `subset=!(author=="Boner"` `&year=="1988")` excludes the `Boner 1988` trial.[4]   □

2.1.2 Standardised Mean Difference

In the bronchoconstriction meta-analysis used in Example 2.1 all studies measured the outcome of interest on the same scale, so an overall effect can be estimated directly by pooling the mean differences in the individual studies. However, in many settings different studies use different outcome scales, e.g. different depression scales or quality of life scales. In such cases we cannot pool the effect estimates (mean differences) directly. Instead, we calculate a dimensionless effect measure from every study and use this for pooling. A very popular dimensionless effect measure is the standardised mean difference which is the study's mean difference divided by a standard deviation based either on a single treatment group or both treatment groups.

There are a number of formulae in the literature for calculating a standardised mean difference and its variance; see Chapter 16 of Cooper and Hedges [3] for a summary. The `metacont` function from R package **meta** uses the same estimator as RevMan 5 [40], i.e. a version of the standardised mean difference which is called Hedges's g [15, 16] based on the pooled sample variance. This standardised mean difference for study k is calculated as:

$$\hat{g}_k = \left(1 - \frac{3}{4\,n_k - 9}\right) \frac{\hat{\mu}_{ek} - \hat{\mu}_{ck}}{\sqrt{\left((n_{ek} - 1)s_{ek}^2 + (n_{ck} - 1)s_{ck}^2\right)/(n_k - 2)}} \tag{2.4}$$

where $n_k = n_{ek} + n_{ck}$ and the factor $1 - 3/(4\,n_k - 9)$ corrects for the bias in the estimated standard error. To calculate a confidence interval for $\hat{g}_k$, we need its variance; again following RevMan 5 this is calculated as [18, page 80, equation (8)]

$$\widehat{\mathrm{Var}}\,(\hat{g}_k) = \frac{n_k}{n_{ek} \cdot n_{ck}} + \frac{\hat{g}_k^2}{2(n_k - 3.94)}\,. \tag{2.5}$$

---

[3]The parentheses are not mandatory to select `Boner 1988`; we use them only to make the R code more accessible.

[4]The parentheses are mandatory to exclude `Boner 1988` using the variables `author` and `year`.

Once $\hat{g}_k$ and $\widehat{\mathrm{Var}}\,(\hat{g}_k)$ are calculated a two-sided $(1-\alpha)$ confidence interval can be calculated by

$$\hat{g}_k \;\pm\; z_{1-\frac{\alpha}{2}}\; \mathrm{S.E.}\,(\hat{g}_k) \tag{2.6}$$

with standard error $\mathrm{S.E.}\,(\hat{g}_k) = \sqrt{\widehat{\mathrm{Var}}\,(\hat{g}_k)}$ and $z_{1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution.

Example 2.2 Furukawa et al. [10] carried out a systematic review comparing low dosage tricyclic antidepressants with placebo for the treatment of depression. They reported the effect on presence/absence of depression and on depression severity. Here we focus on the latter outcome. Unfortunately, different studies used different scores to measure depression severity, e.g. 19 studies used some version of the Hamilton Depression Rating Scale and five studies used the Montgomery-Åsberg Depression Rating Scale. Accordingly, it is not possible to pool the estimated effects directly.

Figure 2.1 reads in and views the data assuming that the file `dataset02.csv` is in the current working directory; see Sect. 1.2 for details. The large differences in means (columns `Me`, `Mc`) and standard deviations (columns `Se`, `Sc`) within the experimental and control arms are typical of what occurs when different studies use different outcome measures.

For each study (labelled by first author) mean values, standard deviations and sample sizes are given in Fig. 2.1. For study 1 (`Blashki`), the standardised mean difference with its 95 % confidence interval can be calculated using formulae (2.4) to (2.6) in the following way:

$$\hat{g}_1 = \left(1 - \frac{3}{4\,(13+18)-9}\right)\; \frac{6.4 - 11.4}{\sqrt{(12 \cdot 5.4^2 + 17 \cdot 9.6^2)/(13+18-2)}} = -0.60.$$

Further

$$\widehat{\mathrm{Var}}\,(\hat{g}_1) = \frac{13+18}{13 \cdot 18} + \frac{-0.60^2}{2\,(13+18-3.94)} = 0.1391305$$

and thus

$$\mathrm{S.E.}\,(\hat{g}_1) = \sqrt{0.1391305} = 0.373002.$$

The 95 % confidence interval is

$$-0.6 \;\pm\; 1.96 \cdot 0.373002, \quad \text{i.e.} \quad (-1.33, 0.13).$$

```
> # 1. Read in the data:
> data2 <- read.csv("dataset02.csv")
> # 2. As usual, to view an object, type its name:
> data2
              author  Ne     Me    Se Nc    Mc    Sc
1    Blashki(75%150)  13   6.40  5.40 18 11.40  9.60
2     Hormazabal(86)  17  11.00  8.20 16 19.00  8.20
3  Jacobson(75-100)  10  17.50  8.80  6 23.00  8.80
4         Jenkins(75)  7  12.30  9.90  7 20.00 10.50
5     Lecrubier(100)  73  15.70 10.60 73 18.70 10.60
6         Murphy(100)  26   8.50 11.00 28 14.50 11.00
7           Nandi(97)  17  25.50 24.00 10 53.20 11.20
8      Petracca(100)  11   6.20  7.60 10 10.00  7.60
9       Philipp(100) 105  -8.10  3.90 46 -8.50  5.20
10    Rampello(100)  22  13.40  2.30 19 19.70  1.30
11       Reifler(83)  13  12.50  7.60 15 12.50  7.60
12        Rickels(70)  29   1.99  0.77 39  2.54  0.77
13     Robertson(75)  13  11.00  8.20 13 15.00  8.20
14      Rouillon(98)  78  15.80  6.80 71 17.10  7.20
15            Tan(70)  23  -8.50  8.60 23 -8.30  6.00
16 Tetreault(50-100)  11  51.90 18.50 11 74.30 18.50
17      Thompson(75)  11   8.00  8.10 18 10.00  9.70
> # 3. Calculate total sample sizes
> summary(data2$Ne+data2$Nc)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  14.00   26.00   31.00   53.06   54.00  151.00
```

Fig. 2.1 Data from meta analysis by Furukawa et al. [10]. See Table 2.1 for details on the variables in dataset data2

We can calculate the standardised mean difference, its standard error and 95 % confidence interval for study 1 (Blashki) using base R:

```
> # 1. Calculate standardised mean difference (SMD) and
> #    its standard error (seSMD) for study 1 (Blashki) of
> #    dataset data2:
> N <- with(data2[1,], Ne + Nc)
> SMD <- with(data2[1,],
+             (1 - 3/(4 * N - 9)) * (Me - Mc) /
+             sqrt(((Ne - 1) * Se^2 + (Nc - 1) * Sc^2)/(N - 2)))
> seSMD <- with(data2[1,],
+              sqrt(N/(Ne * Nc) + SMD^2/(2 * (N - 3.94))))
> # 2. Print standardised mean difference and limits of 95% CI
> #    interval using round function to show only two digits:
> round(c(SMD, SMD + c(-1,1) * qnorm(1-(0.05/2)) * seSMD), 2)
[1] -0.60 -1.33  0.13
```

We get the same result by using the `metacont` function with argument
`sm="SMD"` (Standardised Mean Difference):

```
> print(metacont(Ne, Me, Se, Nc, Mc, Sc, sm="SMD",
+                data=data2, subset=1), digits=2)
  SMD        95%-CI      z  p-value
 -0.6 [-1.33; 0.13] -1.61   0.1083

Details:
- Inverse variance method
```

Once the standardised mean difference and its variance have been calculated
using the formulae (2.4) and (2.5), the calculations for both fixed effect and random
effects meta-analyses follow exactly as described in the next section.   □

2.2 Fixed Effect Model

The fixed effect model assumes that the estimated effects from the component
studies in a meta-analysis come from a single homogeneous population. In order
to calculate an overall estimate, we therefore average the estimates from each study,
allowing for the fact that some estimates are more precise than others (having come
from larger studies).

More formally, let $k = 1, \ldots, K$ index study, $\hat{\theta}_k$ denote the intervention effect
estimate from study k, and θ denote the intervention effect in the population, which
we wish to estimate. Denote by $\hat{\sigma}_k^2$ the sample estimate of $\mathrm{Var}\,(\hat{\theta}_k)$.

The fixed effect model is

$$\hat{\theta}_k = \theta + \sigma_k \epsilon_k, \quad \epsilon_k \overset{\text{i.i.d.}}{\sim} N(0, 1). \tag{2.7}$$

We now consider the fixed effect estimate of θ, denoted by $\hat{\theta}_F$. Given estimates
$(\hat{\theta}_k, \hat{\sigma}_k), k = 1, \ldots, K$, the maximum-likelihood estimate under model (2.7) is

$$\hat{\theta}_F = \frac{\sum\limits_{k=1}^{K} \hat{\theta}_k / \hat{\sigma}_k^2}{\sum\limits_{k=1}^{K} 1 / \hat{\sigma}_k^2} = \frac{\sum\limits_{k=1}^{K} w_k \hat{\theta}_k}{\sum\limits_{k=1}^{K} w_k}. \tag{2.8}$$

Accordingly, $\hat{\theta}_F$ is a weighted average of the individual effect estimates $\hat{\theta}_k$ with
weights $w_k = 1/\hat{\sigma}_k^2$. Therefore, this method is called the *inverse variance method*.

The variance of $\hat{\theta}_F$ is estimated by

$$\widehat{\mathrm{Var}}\,(\hat{\theta}_F) = \frac{1}{\sum\limits_{k=1}^{K} w_k}. \tag{2.9}$$

A $(1 - \alpha)$ confidence interval for $\hat{\theta}_F$ can be calculated by

$$\hat{\theta}_F \ \pm \ z_{1-\frac{\alpha}{2}} \ \text{S.E.} \, (\hat{\theta}_F) \tag{2.10}$$

with standard error S.E. $(\hat{\theta}_F) = \sqrt{\widehat{\text{Var}} \, (\hat{\theta}_F)}$ and $z_{1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution. A corresponding test for an overall treatment effect can be constructed using $\hat{\theta}_F \Big/ \text{S.E.} \, (\hat{\theta}_F)$ as test statistic.

Example 2.3 The fixed effect estimate $\hat{\theta}_F$ and its 95 % confidence interval for the bronchoconstriction meta-analysis are given in Fig. 1.4; here we show how $\hat{\theta}_F$ can be calculated using R. Recall Eqs. (2.1) and (2.2) which give the mean difference $\hat{\mu}_k$ and its variance estimate $\widehat{\text{Var}} \, (\hat{\mu}_k)$. The fixed effect estimate $\hat{\theta}_F$ and its variance can be calculated using the following quantities:

$$\hat{\theta}_k = \hat{\mu}_k$$

$$\hat{\sigma}_k^2 = \widehat{\text{Var}} \, (\hat{\mu}_k).$$

The fixed effect estimate and its variance can be calculated using base R code:

```
> # 1. Calculate mean difference, variance and weights
> MD <- with(data1, Me - Mc)
> varMD <- with(data1, Se^2/Ne + Sc^2/Nc)
> weight <- 1/varMD
> # 2. Calculate the inverse variance estimator
> round(weighted.mean(MD, weight), 4)
[1] -15.514
> # 3. Calculate the variance
> round(1/sum(weight), 4)
[1] 1.4126
```

Note, the standard `weighted.mean` function is used to calculate $\hat{\theta}_F$.

The meta-analysis can be conducted much easier using the `metacont` function which yields identical results:

```
> mc1 <- metacont(Ne, Me, Se, Nc, Mc, Sc,
+                 data=data1,
+                 studlab=paste(author, year))
> round(c(mc1$TE.fixed, mc1$seTE.fixed^2), 4)
[1] -15.5140   1.4126
```

We select `mc1$TE.fixed`, i.e. the Treatment Estimate in the fixed effect model, and its standard error `mc1$seTE.fixed` from the meta-analysis object `mc1`. We can use the command `str(mc1)` to print the whole structure of the meta-analysis object `mc1` and look at the help page of the `metacont` function which describes the individual elements of `mc1`.

A complete printout for the meta-analysis is given in Fig. 2.2. The first thing the output gives is a table whose rows are the component studies in the meta-analysis.

```
                        MD          95%-CI %W(fixed)  %W(random)
Boner 1988             -7.2 [-21.1;   6.7]      2.82        3.08
Boner 1989             -7.0 [-16.2;   2.2]      6.38        6.58
Chudry 1987           -18.4 [-28.8;  -8.0]      5.01        5.29
Comis 1993            -16.8 [-27.8;  -5.8]      4.50        4.78
DeBenedictis 1994a    -13.0 [-22.8;  -3.2]      5.68        5.93
DeBenedictis 1994b    -16.6 [-35.8;   2.6]      1.47        1.64
DeBenedictis 1995     -13.9 [-27.6;  -0.2]      2.87        3.13
Debelic 1986          -18.2 [-30.7;  -5.8]      3.52        3.80
Henriksen 1988        -29.7 [-41.6; -17.8]      3.83        4.11
Konig 1987            -14.2 [-25.0;  -3.4]      4.65        4.93
Morton 1992           -22.5 [-33.5; -11.5]      4.48        4.76
Novembre 1994f        -13.0 [-19.5;  -6.6]     12.98       12.15
Novembre 1994s        -15.1 [-23.8;  -6.4]      7.14        7.28
Oseid 1995            -14.8 [-23.7;  -5.9]      6.82        6.99
Roberts 1985          -20.0 [-36.9;  -3.1]      1.90        2.10
Shaw 1985             -24.2 [-33.2; -15.1]      6.67        6.85
Todaro 1993           -13.4 [-18.7;  -8.1]     19.29       16.58


Number of studies combined: k=17


                      MD          95%-CI     z   p-value
Fixed effect model   -15.5 [-17.8; -13.2] -13.1 < 0.0001
Random effects model -15.6 [-18.1; -13.2] -12.3 < 0.0001

Quantifying heterogeneity:
tau^2 = 2.4374; H = 1.05 [1; 1.35]; I^2 = 8.9% [0%; 45.3%]

Test of heterogeneity:
    Q d.f.  p-value
 17.57   16   0.3496


Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

Fig. 2.2 Output from meta-analysis of the bronchoconstriction meta-analysis [37]. The output starts with a table of the included studies. For each study, the mean difference (MD) with 95 % confidence interval is given, along with weights used for fixed effect and random effects model. There are 17 studies in the example. Next, the results of fixed effect and random effects model are presented with 95 % confidence intervals, z statistic and p-value. Heterogeneity is quantified by the estimated between-study variance τ^2, H and I^2, see Sects. 2.3 and 2.4, and tested using Cochran's Q statistic, see Eq. (2.12). There is not much heterogeneity present in this example. The output ends with details of the methods used, e.g. how τ^2 was estimated, see Sect. 2.3.1

This table is also shown in Fig. 1.4 on the right side of the forest plot. The column MD is the mean difference of the response (maximum change in FEV_1 as a percentage) between the Nedocromil sodium and placebo group. Next comes a 95 % confidence interval for this difference, calculated based on (2.3). The next two columns are the

weights given to the study under the fixed effect (`%W(fixed)`) and random effects model (`%W(random)`).

The weight of study 1 (`Boner 1988`) in the fixed effect meta-analysis is given by the inverse of the variance (2.2) which can be calculated as

$$1 \Big/ \left(\frac{13.85^2}{13} + \frac{21.46^2}{13} \right) = 1/50.18108 = 0.01992783.$$

The percentage weight of study 1 (`Boner 1988`) in the fixed effect meta-analysis reported in Figs. 1.4 and 2.2 is

$$100 \cdot \frac{w_1}{\sum_{i=1}^{17} w_i} = 100 \cdot \frac{0.01992783}{0.7079028} = 2.82\%.$$

We could also use R to calculate these values:

```
> mc1$w.fixed[1]
[1] 0.01992783
> sum(mc1$w.fixed)
[1] 0.7079028
> round(100*mc1$w.fixed[1] / sum(mc1$w.fixed), 2)
[1] 2.82
```

After reporting the number of studies combined in meta-analysis, fixed effect estimate $\hat{\theta}_F$, random effects estimate $\hat{\theta}_R$ (see Sect. 2.3) and their 95% confidence intervals, z and p-values are given in Fig. 2.2. Next come the measures for heterogeneity and a test for heterogeneity (see Sect. 2.4). Finally a note indicates that the "Inverse variance method" has been used. This is in fact the only method for continuous data; but with binary data (see Chap. 3) we shall see there are other alternatives.

A forest plot is shown in Fig. 2.3 which has been produced by the R command

```
> forest(mc1, comb.random=FALSE, xlab=
+          "Difference in mean response (intervention – control)
+ units: maximum % fall in FEV1",
+          xlim=c(-50,10), xlab.pos=-20, smlab.pos=-20)
```

Note the use of the `xlab` option to label the x-axis, and in particular how a line break in the input text creates a line break in the axis label on the graph. The option `xlim=c(-50,10)` is used to specify that the limits of the x-axis are between -50 and 10. The options `xlab.pos` and `smlab.pos` specify the centre of the label on x-axis and the summary measure at the top of the figure; otherwise these texts would be centred around 0.

Note, the meta-analysis could have also been done using the `metagen` function which is the primary function in R package **meta** to conduct a meta-analysis based on the generic inverse variance method.

```
> # 1. Apply generic inverse variance method
> mc1.gen <- metagen(mc1$TE, mc1$seTE, sm="MD")
```

Study	Experimental			Control			Mean difference	MD	95%–CI	W(fixed)
	Total	Mean	SD	Total	Mean	SD				
Boner 1988	13	13.54	13.85	13	20.77	21.46		−7.23	[−21.11; 6.65]	2.8%
Boner 1989	20	15.70	13.10	20	22.70	16.47		−7.00	[−16.22; 2.22]	6.4%
Chudry 1987	12	21.30	13.10	12	39.70	12.90		−18.40	[−28.80; −8.00]	5.0%
Comis 1993	12	14.50	12.20	12	31.30	15.10		−16.80	[−27.78; −5.82]	4.5%
DeBenedictis 1994a	17	14.40	11.10	17	27.40	17.30		−13.00	[−22.77; −3.23]	5.7%
DeBenedictis 1994b	8	14.80	18.60	8	31.40	20.60		−16.60	[−35.83; 2.63]	1.5%
DeBenedictis 1995	13	15.70	16.80	13	29.60	18.90		−13.90	[−27.65; −0.15]	2.9%
Debelic 1986	12	29.83	15.95	12	48.08	15.08		−18.25	[−30.67; −5.83]	3.5%
Henriksen 1988	12	17.50	13.10	12	47.20	16.47		−29.70	[−41.61; −17.79]	3.8%
Konig 1987	12	12.00	14.60	12	26.20	12.30		−14.20	[−25.00; −3.40]	4.7%
Morton 1992	16	15.83	13.43	16	38.36	18.01		−22.53	[−33.54; −11.52]	4.5%
Novembre 1994f	24	15.42	8.35	24	28.46	13.84		−13.04	[−19.51; −6.57]	13.0%
Novembre 1994s	19	11.00	12.40	19	26.10	14.90		−15.10	[−23.82; −6.38]	7.1%
Oseid 1995	20	14.10	9.50	20	28.90	18.00		−14.80	[−23.72; −5.88]	6.8%
Roberts 1985	9	18.90	17.70	9	38.90	18.90		−20.00	[−36.92; −3.08]	1.9%
Shaw 1985	8	10.27	7.02	8	34.43	10.96		−24.16	[−33.18; −15.14]	6.7%
Todaro 1993	13	10.10	8.90	13	23.50	4.00		−13.40	[−18.70; −8.10]	19.3%
Fixed effect model	**240**			**240**				**−15.51**	**[−17.84; −13.18]**	**100%**

Heterogeneity: I–squared=8.9%, tau–squared=2.437, p=0.3496

−50 −40 −30 −20 −10 0 10
Difference in mean response (intervention − control)
units: maximum % fall in FEV1
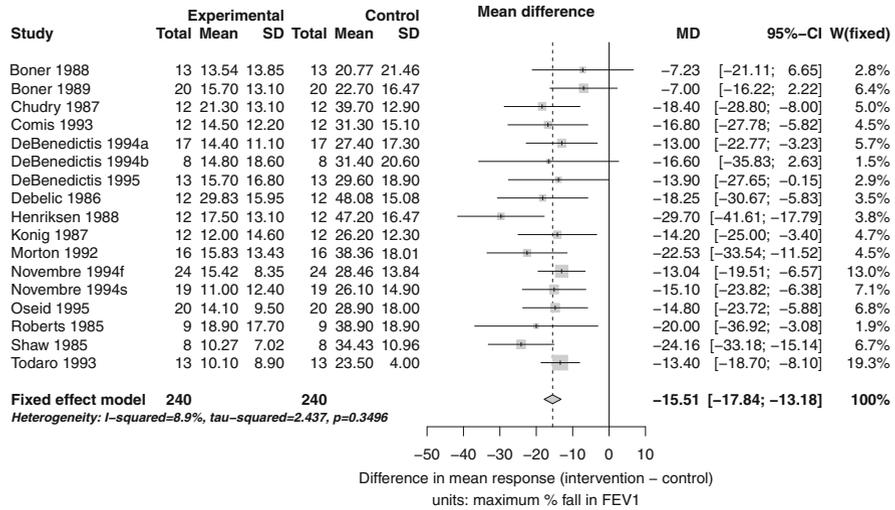
Fig. 2.3 Forest plot for the bronchoconstriction meta-analysis [37]. For details, see text

```
> # 2. Same result
> mc1.gen <- metagen(TE, seTE, data=mc1, sm="MD")
> # 3. Print results for fixed effect and random effects method
> c(mc1$TE.fixed, mc1$TE.random)
[1] −15.51403 −15.64357
> c(mc1.gen$TE.fixed, mc1.gen$TE.random)
[1] −15.51403 −15.64357
```

In steps 1 and 2, the generic inverse variance method is applied using the metagen function; we use the list elements mc1$TE (treatment effect) and mc1$seTE (standard error) as inputs to the metagen function. Output of resulting object mc1.gen is identical to results using the metacont function as exemplified in step 3 for the fixed effect and random effects estimates. Applying the metagen function in this way seems rather artificial, however, as we will see in Sect. 2.6 this function can be used to conduct a meta-analysis for other outcomes. □

Following RevMan 5, the following quantities are used to estimate the standardised mean difference in the fixed effect model:

$$\hat{\theta}_k = \hat{g}_k$$

$$\hat{\sigma}_k^2 = \widehat{\text{Var}}(\hat{g}_k)$$

with $\hat{g}_k$ and $\widehat{\text{Var}}(\hat{g}_k)$ defined in (2.4) and (2.5). These quantities are utilised in formulae (2.8)–(2.10) to calculate the fixed effect estimate of the standardised mean difference.

Example 2.4 For the standardised mean difference, we can calculate the fixed effect
estimate and its variance using base R:

```
> # 1. Calculate standardised mean difference,
> #     variance and weights
> N <- with(data2, Ne + Nc)
> SMD <- with(data2,
+             (1 - 3/(4 * N - 9)) * (Me - Mc)/
+             sqrt(((Ne - 1) * Se^2 + (Nc - 1) * Sc^2)/(N - 2)))
> varSMD <- with(data2,
+               N/(Ne * Nc) + SMD^2/(2 * (N - 3.94)))
> weight <- 1/varSMD
> # 2. Calculate the inverse variance estimator
> round(weighted.mean(SMD, weight), 4)
[1] -0.3915
> # 3. Calculate the variance
> round(1/sum(weight), 4)
[1] 0.0049
```

Again, the meta-analysis can be conducted using the `metacont` function:

```
> mc2 <- metacont(Ne, Me, Se, Nc, Mc, Sc, sm="SMD",
+                 data=data2)
> round(c(mc2$TE.fixed, mc2$seTE.fixed^2), 4)
[1] -0.3915  0.0049
```

A complete summary for the meta-analysis is given in Fig. 2.4.   □

```
> print(summary(mc2), digits=2)
Number of studies combined: k=17


                        SMD         95%-CI      z  p-value
Fixed effect model   -0.39 [-0.53; -0.25] -5.61 < 0.0001
Random effects model -0.59 [-0.87; -0.30] -4.04 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.2309; H = 1.91 [1.5; 2.43]; I^2 = 72.5% [55.4%; 83.1%]

Test of heterogeneity:
    Q d.f.  p-value
 58.27   16 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

Fig. 2.4 Output from meta-analysis of the tricyclic antidepressants for depression [10]. The output
is organised similar to Fig. 2.2, except that information on individual studies is omitted by using
the `summary.meta` function

2.3   Random Effects Model

The random effects model seeks to account for the fact that the study effect estimates $\hat{\theta}_k$ are often more variable than assumed in the fixed effect model. Under the random effects model,

$$\hat{\theta}_k = \theta + u_k + \sigma_k \epsilon_k, \quad \epsilon_k \overset{\text{i.i.d.}}{\sim} N(0, 1); \ u_k \overset{\text{i.i.d.}}{\sim} N(0, \tau^2), \tag{2.11}$$

where the u's and ϵ's are independent. Comparing with (2.7) shows the random effects model has the fixed effect model as a special case when $\tau^2 = 0$. A key assumption of the random effects model is that the u_k we see in our data are not intrinsically associated with study k; if study k was rerun, the new u_k would be an independent draw from $N(0, \tau^2)$. This is known as the *exchangeability assumption*. If we accept this assumption then, compared with the fixed effect model, calculating an overall effect estimate will pay greater attention to the effect estimates from the smaller studies. This difference with the fixed effect model lies at the heart of discussions about whether the random effects model is appropriate. A number of authors have argued that, as small studies are more susceptible to bias, the fixed effect estimate is (almost) always preferable [11, 30].

Under the random effects model there are a number of options for estimating θ, $\text{Var}(\hat{\theta})$ and τ^2. Maximum-likelihood is attractive, but the resulting variance estimates are biased downwards if the number of studies is small. This has led to the widespread use of the method of moments estimate proposed by DerSimonian and Laird [7], which has the attraction that it can be readily calculated when the response is discrete, when maximum-likelihood estimation is less straightforward.

Again, the default settings in the `metacont` function are the same as those in RevMan 5. Define

$$Q = \sum_{k=1}^{K} w_k (\hat{\theta}_k - \hat{\theta}_F)^2 \tag{2.12}$$

the weighted sum of squares about the fixed effect estimate with $w_k = 1/\hat{\sigma}_k^2$. This is usually referred to as either the homogeneity test statistic or the heterogeneity statistic [18, p. 266, 290]. Next define

$$S = \sum_{k=1}^{K} w_k - \frac{\sum_{k=1}^{K} w_k^2}{\sum_{k=1}^{K} w_k}.$$

If $Q < (K - 1)$, then $\hat{\tau}^2$ is set to 0 and the random effects estimate $\hat{\theta}_R$ is set equal to the fixed effect estimate $\hat{\theta}_F$. Otherwise, the DerSimonian–Laird estimator of the

between-study variance is defined as

$$\hat{\tau}^2 = \frac{Q - (K - 1)}{S}$$

and the random effects estimate and its variance are given by

$$\hat{\theta}_R = \frac{\sum_{k=1}^{K} w_k^* \hat{\theta}_k}{\sum_{k=1}^{K} w_k^*} \tag{2.13}$$

$$\widehat{\text{Var}}\left(\hat{\theta}_R\right) = \frac{1}{\sum_{k=1}^{K} w_k^*}. \tag{2.14}$$

with weights $w_k^* = 1/(\hat{\sigma}_k^2 + \hat{\tau}^2)$. The random effects estimator $\hat{\theta}_R$ is a weighted average of the individual effect estimates $\hat{\theta}_k$ with weights $1/(\hat{\sigma}_k^2 + \hat{\tau}^2)$. Accordingly, this method is often called "Inverse variance method", too.

A $(1-\alpha)$ confidence interval for $\hat{\theta}_R$ can be calculated by

$$\hat{\theta}_R \pm z_{1-\frac{\alpha}{2}} \text{ S.E.}\left(\hat{\theta}_R\right) \tag{2.15}$$

with standard error S.E. $\left(\hat{\theta}_R\right) = \sqrt{\widehat{\text{Var}}\left(\hat{\theta}_R\right)}$ and $z_{1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution. A corresponding test for an overall treatment effect can be constructed using $\hat{\theta}_R \big/ \text{S.E.}\left(\hat{\theta}_R\right)$ as test statistic.

Note, formulae (2.13)–(2.15) are used for the standardised mean difference, too.

The method used to estimate the between-study variance τ^2 may have a large impact on the weighting of studies. Several method to estimate τ^2 besides the DerSimonian–Laird method have been published in the literature. These methods will be described in the next Sect. 2.3.1.

Example 2.5 The result for the random effects model fitted to the bronchoconstriction dataset is given in Fig. 2.2. The weight of study 1 (Boner 1988) is

$$100 \cdot \frac{w_1^*}{\sum_{i=1}^{K} w_i^*} = 100 \cdot \frac{0.019005}{0.6179183} = 3.08.$$

The random effects estimate is very similar to the fixed effect estimate ($\hat{\theta}_F = -15.5$, $\hat{\theta}_R = -15.6$); likewise confidence interval limits are similar.   □

Example 2.6 For the depression meta-analysis fixed effect and random effects estimates are rather different ($\hat{\theta}_F = -0.39$, $\hat{\theta}_R = -0.59$), see Fig. 2.4. Furthermore, the confidence interval for the random effects model is much wider. Nevertheless,

both models show a highly statistically significant beneficial effect of tricyclic antidepressants on depression severity. □

2.3.1 Estimation of Between-Study Variance

The following methods to estimate the between-study variance τ^2 are available in the `metagen` and other functions of R package **meta** (argument `method.tau`):

- DerSimonian–Laird estimator [7] (`method.tau="DL"`) (default)
- Paule–Mandel estimator [27] (`method.tau="PM"`)
- Restricted maximum-likelihood estimator [43] (`method.tau="REML"`)
- Maximum-likelihood estimator [43] (`method.tau="ML"`)
- Hunter–Schmidt estimator [22, 43] (`method.tau="HS"`)
- Sidik–Jonkman estimator [35] (`method.tau="SJ"`)
- Hedges estimator [17] (`method.tau="HE"`)
- Empirical Bayes estimator [39] (`method.tau="EB"`).

The DerSimonian–Laird estimator is by far the most popular method, especially in medical research. For example, the DerSimonian–Laird estimator is the only method available in RevMan 5 [40]. Accordingly, this method is the default in R package **meta**.

The properties of these estimators have been evaluated in Monte Carlo simulations [36, 43] as well as analytically [43]. Results of these evaluations are inconsistent, recommending the restricted maximum-likelihood estimator [43] and Sidik–Jonkman or Empirical Bayes estimator [36], respectively.

As a technical note, with exception of the DerSimonian–Laird and the Paule–Mandel methods the `rma.uni` function of R package **metafor** is called internally in the `metagen` function. Thus, it is a good idea to install R package **metafor** to make all estimation methods available.[5] Further details on the various methods are provided in the help page of the `rma.uni` function.

Example 2.7 A forest plot with results for the various estimates of τ^2 in the bronchoconstriction dataset is shown in Fig. 2.5.[6] Results are similar for DerSimonian–Laird, restricted maximum-likelihood and empirical Bayes estimator. Whereas the Sidik–Jonkman estimator is surprisingly large, other estimators (i.e. Paule–Mandel, maximum-likelihood, Hunter–Schmidt and Hedges) are rather small. The very large estimate of τ^2 from the Sidik–Jonkman method cautions against relying exclusively on this approach. □

---

[5]R command: `install.packages("metafor")`.

[6]R code to create the forest plot is given in the web-appendix.

Method	Between–study heterogeneity		MD	95%–CI
Fixed–effect model				
	0		−15.51	[−17.84; −13.18]
Random–effect model				
DerSimonian–Laird	2.44		−15.64	[−18.14; −13.15]
Paule–Mandel	0		−15.51	[−17.84; −13.18]
Restricted maximum–likelihood	2.52		−15.65	[−18.15; −13.15]
Maximum–likelihood	0.06		−15.52	[−17.85; −13.18]
Hunter–Schmidt	0.81		−15.56	[−17.95; −13.18]
Sidik–Jonkman	15.75		−15.96	[−19.10; −12.81]
Hedges	0		−15.51	[−17.84; −13.18]
Empirical Bayes	2.48		−15.65	[−18.14; −13.15]

```
        −20    −18    −16    −14    −12
```
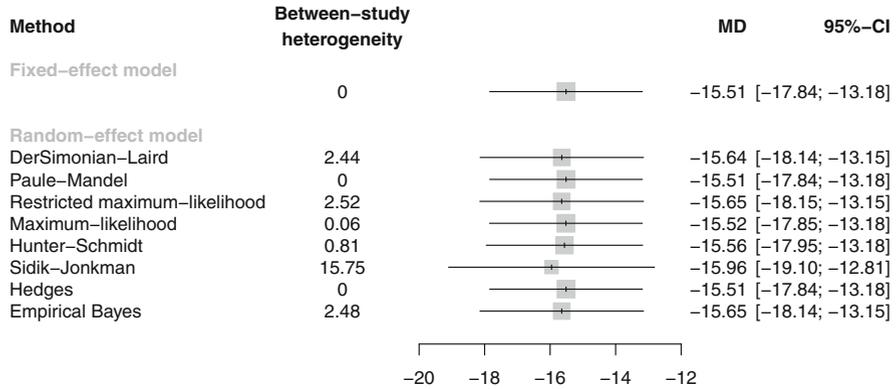
Fig. 2.5 Forest plot for the bronchoconstriction meta-analysis [37] comparing estimation methods for between-study heterogeneity τ^2

2.3.2   Hartung–Knapp Adjustment

Hartung and Knapp [14, 25] introduced a new meta-analysis method based on a refined variance estimator in the random effects model. It has been argued in a recent publication in the Annals of Internal Medicine that the Hartung–Knapp method is preferred over the DerSimonian–Laird method [4].

Instead of using the variance estimate given in Eq. (2.14), Hartung and Knapp propose to use the following variance estimator for $\hat{\theta}_R$:

$$\widehat{\text{Var}}_{\text{HK}}(\hat{\theta}_R) = \frac{1}{K-1} \sum_{k=1}^{K} \frac{w_k^*}{w^*} \left(\hat{\theta}_k - \hat{\theta}_R \right)^2 \qquad (2.16)$$

with weights w_k^* as given in Eq. (2.14) and $w^* = \sum_{k=1}^{K} w_k^*$.

Hartung [13] showed that

$$\frac{\hat{\theta}_R - \theta}{\text{S.E.}_{\text{HK}}(\hat{\theta}_R)}$$

with standard error S.E. $_{\text{HK}}(\hat{\theta}_R) = \sqrt{\widehat{\text{Var}}_{\text{HK}}(\hat{\theta}_R)}$ follows a t-distribution with $K - 1$ degrees of freedom.

Accordingly, a $(1-\alpha)$ confidence interval for $\hat{\theta}_R$ based on the Hartung–Knapp method can be calculated by

$$\hat{\theta}_R \ \pm \ t_{K-1;1-\frac{\alpha}{2}} \ \text{S.E.}_{\text{HK}}(\hat{\theta}_R) \qquad (2.17)$$

with $t_{K-1;1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the t-distribution with $K - 1$ degrees of freedom. A corresponding test for an overall treatment effect can be constructed using $\hat{\theta}_R \big/ \text{S.E.}_{\text{HK}}(\hat{\theta}_R)$ as test statistic.

It has been shown in simulations [25] that a test based on the Hartung–Knapp modification holds the prespecified significance level much better than tests based on S.E. $(\hat{\theta}_F)$ and S.E. $(\hat{\theta}_R)$.

Example 2.8 Results of fixed effect and random effects model to evaluate the use of tricyclic antidepressants for depression [10] are reported in Fig. 2.4.

We can either use the `metacont` function to conduct the Hartung–Knapp adjustment

```
> mc2.hk <- metacont(Ne, Me, Se, Nc, Mc, Sc, sm="SMD",
+                     data=data2, comb.fixed=FALSE,
+                     hakn=TRUE)
```

or the `metagen` function

```
> mc2.hk <- metagen(TE, seTE, data=mc2, comb.fixed=FALSE,
+                    hakn=TRUE)
```

We print the summary of the meta-analysis in the usual way.

```
> print(summary(mc2.hk), digits=2)
Number of studies combined: k=17

                                  95%-CI    t  p-value
Random effects model -0.59 [-0.95; -0.22] -3.4   0.0036

Quantifying heterogeneity:
tau^2 = 0.2309; H = 1.91 [1.5; 2.43]; I^2 = 72.5% [55.4%; 83.1%]

Test of heterogeneity:
    Q d.f.  p-value
 58.27   16 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
- Hartung-Knapp adjustment for random effects model
```

Use of the Hartung–Knapp method yields a much wider 95 % confidence interval as compared to the classic random effects model (see Fig. 2.4): $[-0.95; -0.22]$ versus $[-0.87; -0.30]$. Furthermore, using the test for an overall treatment effect is based on a t-distribution with $K - 1$ degrees of freedom. Accordingly, the p-value is much larger ($p = 0.0036$) as compared to the p-value of the classic random effects method ($p < 0.0001$, see Fig. 2.4). Nonetheless, the test for an overall treatment effect is still highly significant. □

2.3.3   Prediction Intervals

The confidence interval for the random effects estimator $\hat{\theta}_R$ given by Eq. (2.15) describes the uncertainty in the estimation of the mean treatment effect. However, in order to calculate a prediction interval [21] for the treatment effect in a future study from the random effects model (2.11), we need to take into account not only uncertainty in estimating the mean treatment effect but also the between-study variance τ^2.

Such a $(1 - \alpha)$ prediction interval can be calculated as

$$\hat{\theta}_R \;\pm\; t_{K-2,1-\frac{\alpha}{2}} \; \sqrt{\widehat{\mathrm{Var}}\,(\hat{\theta}_R) + \hat{\tau}^2}, \tag{2.18}$$

where we include the estimate of τ in the variance, and $t_{K-2,1-\frac{\alpha}{2}}$ denotes the $1 - \frac{\alpha}{2}$ quantile of the t-distribution with $K - 2$ degrees of freedom.

Example 2.9 In the R package **meta** a prediction interval can be printed in several ways. We can use the argument `prediction=TRUE` in the creation of a meta-analysis object using the `metacont` function.[7] Or, we can specify the `prediction` argument in a `summary`, `forest` or `print` command. In the following R code we use the `prediction` argument in the `summary.meta` command.

```
> print(summary(mc1, prediction=TRUE), digits=2)
Number of studies combined: k=17

                         MD           95%-CI      z  p-value
Fixed effect model    -15.51 [-17.84; -13.18] -13.05 < 0.0001
Random effects model  -15.64 [-18.14; -13.15] -12.30 < 0.0001
Prediction interval          [-19.94; -11.35]
*** Output truncated ***
```

The result for the prediction interval is printed just below the results for the two meta-analysis methods. Note that the point estimate, i.e. the random effects estimate $\hat{\theta}_R$, is not reported for a prediction interval. In the bronchoconstriction meta-analysis the prediction interval is $(-19.94, -11.35)$. Therefore, in a new study we expect an average treatment effect of more than $11\,\%$.

A forest plot showing a prediction interval can be easily generated using the following command:

```
> forest(mc1, prediction=TRUE, col.predict="black")
```

This is shown in Fig. 2.6. The prediction interval is shown as a bar below the two diamonds for the meta-analysis results. We changed the colour of the bar to black; by default, a red bar would be printed.   □

---

[7]We did not do this in the creation of R object `mc1`.

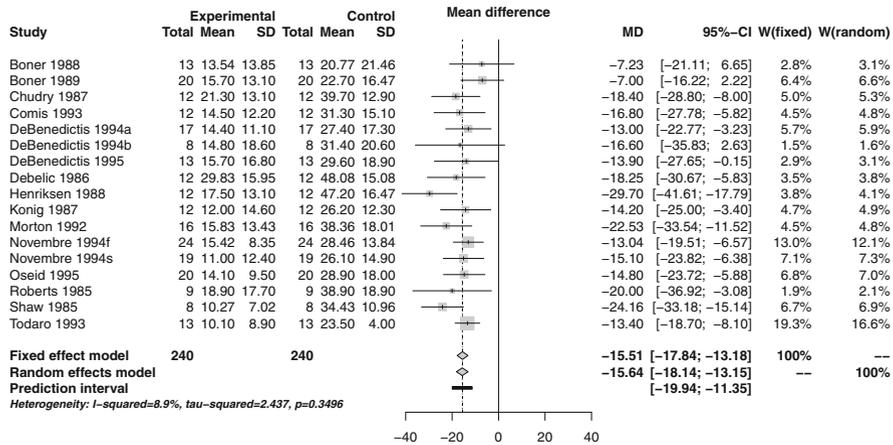| Study	Experimental			Control			Mean difference	MD	95%–CI	W(fixed)	W(random)
	Total	Mean	SD	Total	Mean	SD					
Boner 1988	13	13.54	13.85	13	20.77	21.46		−7.23	[−21.11; 6.65]	2.8%	3.1%
Boner 1989	20	15.70	13.10	20	22.70	16.47		−7.00	[−16.22; 2.22]	6.4%	6.6%
Chudry 1987	12	21.30	13.10	12	39.70	12.90		−18.40	[−28.80; −8.00]	5.0%	5.3%
Comis 1993	12	14.50	12.20	12	31.30	15.10		−16.80	[−27.78; −5.82]	4.5%	4.8%
DeBenedictis 1994a	17	14.40	11.10	17	27.40	17.30		−13.00	[−22.77; −3.23]	5.7%	5.9%
DeBenedictis 1994b	8	14.80	18.60	8	31.40	20.60		−16.60	[−35.83; 2.63]	1.5%	1.6%
DeBenedictis 1995	13	15.70	16.80	13	29.60	18.90		−13.90	[−27.65; −0.15]	2.9%	3.1%
Debelic 1986	12	29.83	15.95	12	48.08	15.08		−18.25	[−30.67; −5.83]	3.5%	3.8%
Henriksen 1988	12	17.50	13.10	12	47.20	16.47		−29.70	[−41.61; −17.79]	3.8%	4.1%
Konig 1987	12	12.00	14.60	12	26.20	12.30		−14.20	[−25.00; −3.40]	4.7%	4.9%
Morton 1992	16	15.83	13.43	16	38.36	18.01		−22.53	[−33.54; −11.52]	4.5%	4.8%
Novembre 1994f	24	15.42	8.35	24	28.46	13.84		−13.04	[−19.51; −6.57]	13.0%	12.1%
Novembre 1994s	19	11.00	12.40	19	26.10	14.90		−15.10	[−23.82; −6.38]	7.1%	7.3%
Oseid 1995	20	14.10	9.50	20	28.90	18.00		−14.80	[−23.72; −5.88]	6.8%	7.0%
Roberts 1985	9	18.90	17.70	9	38.90	18.90		−20.00	[−36.92; −3.08]	1.9%	2.1%
Shaw 1985	8	10.27	7.02	8	34.43	10.96		−24.16	[−33.18; −15.14]	6.7%	6.9%
Todaro 1993	13	10.10	8.90	13	23.50	4.00		−13.40	[−18.70; −8.10]	19.3%	16.6%
Fixed effect model	**240**			**240**				**−15.51**	**[−17.84; −13.18]**	**100%**	**−−**
Random effects model								**−15.64**	**[−18.14; −13.15]**	**−−**	**100%**
Prediction interval									**[−19.94; −11.35]**		
Heterogeneity: I–squared=8.9%, tau–squared=2.437, p=0.3496											

Fig. 2.6 Forest plot for the bronchoconstriction meta-analysis [37] showing a prediction interval which was generated using argument `prediction=TRUE` in the `forest.meta` command

2.4   Tests and Measures of Heterogeneity

There are a number of heterogeneity measures in the literature [19, 32]. The most commonly used measures are calculated by the `metacont` function, and we now briefly describe them. More details on these measures are given in Sect. 4.2.

The first, Q, defined in (2.12), is the weighted sum of squares about the fixed effect estimate $\hat{\theta}_F$. Large values of Q indicate greater heterogeneity between the individual studies in a meta-analysis, and greater values of the between-study heterogeneity τ^2. Under the null hypothesis that $\tau^2 = 0$,

$$Q \sim \chi^2_{K-1},$$

and this can be used to calculate a *p*-value against this null hypothesis.

Two related statistics [20] are commonly quoted:

$$H^2 = \frac{Q}{K-1} \tag{2.19}$$

$$I^2 = \begin{cases} (H^2 - 1)/H^2 & \text{if } Q > (K-1) \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

Under the null hypothesis that $\tau^2 = 0$, Q has mean $K - 1$, so H^2 has mean 1; again large values of H^2 indicate greater heterogeneity. I^2 is a scaled version of H^2, lying between 0 and 1 (or 0% and 100%). Again, large values are consistent with heterogeneity, although for given τ^2, values of I^2 will increase as the sample sizes of the component trials increase [32].

Example 2.10 For the bronchoconstriction meta-analysis, estimates of the measures of heterogeneity ($\tau^2 = 2.44$, $H = 1.05$ [1; 1.35], $I^2 = 8.9\%$ [0%; 45.3%]) and the test for heterogeneity ($Q = 17.57$, p-value $= 0.35$) are given in Fig. 2.2. All these quantities indicate that not much statistical heterogeneity is present. Accordingly, as both fixed effect and random effects are similar and show very strong evidence of an effect, and there is no evidence of heterogeneity, we conclude there is strong evidence Nedocromil sodium ameliorates post-exercise bronchoconstriction. □

Example 2.11 For the depression meta-analysis, estimates of the measures of heterogeneity ($\tau^2 = 0.23$, $H = 1.91$ [1.5; 2.43], $I^2 = 72.5\%$ [55.4%; 83.1%]) and the test for heterogeneity ($Q = 58.27$, p-value < 0.0001) can be found in Fig. 2.4. All these quantities indicate that very large statistical heterogeneity is present. Despite this very large statistical heterogeneity both fixed effect and random effects meta-analysis show a statistically significant beneficial effect of tricyclic antidepressants. Furthermore, only 1 of 17 trials shows a detrimental effect of tricyclic antidepressants. Accordingly, we conclude there is strong evidence for a beneficial effect of tricyclic antidepressants; however, the size of the effect is unclear. □

2.5 Subgroup Analysis

From time to time we need to work with subgroups of studies in a meta-analysis. The various R commands for meta-analysis in the R package **meta** support a `byvar` option, i.e. conduct a subgroup analysis by a variable, which makes this straightforward. We now illustrate its use. More technical details on subgroup analyses are provided in Sect. 4.3.

Example 2.12 Poole and Black [31] report a meta-analysis of mucolytic agents versus placebo for patients with chronic bronchitis and/or chronic obstructive pulmonary disease. The outcome is the mean number of acute exacerbations per month. Acute exacerbation is defined as an increase in cough and in the volume and/or purulence of sputum. As all studies report a mean number of exacerbations, we can work with mean differences, rather than standardised mean differences. R code to read in the data is given in Fig. 2.7. Notice that studies 5 and 12 (`Jackson 1984`, `Grillage 1985`) have zero standard errors.

We do a meta-analysis of the chronic bronchitis data using the following R command:

```
> mc3 <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                 studlab=paste(author, year))
Warning message:
In metacont(Ne, Me, Se, Nc, Mc, Sc, data = data3, :
  Studies with non-positive values for sd.e or sd.c get no weight
  in meta-analysis.
```

```
> # 1. Read in the data:
> data3 <- read.csv("dataset03.csv")
> # 2. As usual, to view an object, type its name:
> data3
          author year  Ne   Me   Se  Nc   Mc   Sc  duration
1   Bontognali 1991  30 0.70 3.76  30 1.27 4.58 <= 3 months
2  Castiglioni 1986 311 0.10 0.21 302 0.20 0.29 <= 3 months
3    Cremonini 1986  21 0.25 0.23  20 0.71 0.29 <= 3 months
4       Grassi 1994  42 0.16 0.29  41 0.45 0.43 <= 3 months
5      Jackson 1984  61 0.11 0.00  60 0.13 0.00 <= 3 months
6      Allegra 1996 223 0.07 0.11 218 0.11 0.14  > 3 months
7     Babolini 1980 254 0.13 0.18 241 0.33 0.27  > 3 months
8        Boman 1983  98 0.20 0.27 105 0.32 0.30  > 3 months
9       Borgia 1981  10 0.05 0.08   9 0.15 0.17  > 3 months
10     Decramer 2005 256 0.10 0.11 267 0.11 0.16  > 3 months
11       Grassi 1976  35 0.14 0.15  34 0.27 0.21  > 3 months
12     Grillage 1985  54 0.10 0.00  55 0.12 0.00  > 3 months
13       Hansen 1994  59 0.11 0.15  70 0.16 0.19  > 3 months
14      Malerba 2004 115 0.06 0.08 119 0.07 0.08  > 3 months
15      McGavin 1985  72 0.42 0.34  76 0.52 0.35  > 3 months
16      Meister 1986  90 0.15 0.15  91 0.20 0.19  > 3 months
17      Meister 1999 122 0.06 0.15 124 0.10 0.15  > 3 months
18      Moretti 2004  63 0.12 0.14  61 0.17 0.17  > 3 months
19        Nowak 1999 147 0.03 0.06 148 0.06 0.12  > 3 months
20      Olivieri 1987 110 0.18 0.31 104 0.33 0.41  > 3 months
21         Parr 1987 243 0.18 0.21 210 0.21 0.21  > 3 months
22         Pela 1999  83 0.17 0.18  80 0.29 0.32  > 3 months
23    Rasmussen 1988  44 0.13 0.21  47 0.14 0.19  > 3 months
```

Fig. 2.7 Reading in data from meta-analysis of mucolytic agents versus placebo for patients with chronic bronchitis and/or chronic obstructive pulmonary disease [31]

A warning has been printed for studies with zero weights. We can verify that these are the `Jackson 1984` and `Grillage 1985` trials:

```
> mc3$studlab[mc3$w.fixed==0]
[1] "Jackson 1984"  "Grillage 1985"
```

The result of the meta-analysis is given by

```
> print(summary(mc3), digits=2)
Number of studies combined: k=21

                        MD         95%-CI      z  p-value
Fixed effect model   -0.05 [-0.05; -0.04] -10.06 < 0.0001
Random effects model -0.08 [-0.11; -0.05]  -5.82 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.0027; H = 2.63 [2.19; 3.15]; I^2 = 85.5% [79.1%; 89.9%]

Test of heterogeneity:
      Q d.f.  p-value
 138.08   20 < 0.0001
Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

The results indicate significant between-study heterogeneity ($Q = 138$, $p <$ 0.0001) with $I^2 = 85.5\%$. Looking at the data (Fig. 2.7), subgroup information is available for study duration: studies whose duration was greater or less than three months.

A subgroup analysis can be done by using argument `byvar` in the original call of the `metacont` function:

```
> mc3s <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                  studlab=paste(author, year),
+                  byvar=duration, print.byvar=FALSE)
```

Another more convenient way is to update the original meta-analysis by using the `update.meta` function from R package **meta**:[8]

```
> mc3s <- update(mc3, byvar=duration, print.byvar=FALSE)
```

The `update.meta` function is a wrapper function for the `metacont` function as well as other R functions discussed in the following chapters. Using the `update.meta` function we only have to specify arguments that should be changed as all other arguments are kept fixed. Note, in order for the `update.meta` function to work the data used in the original function call has to be part of R object `mc3`. This is—by default—the case as argument `keepdata` is equal to `TRUE`. Applying the `update.meta` function to an R object that was created with argument `keepdata=FALSE` would result in a descriptive warning message.

Results of a meta-analysis with subgroups are given by the following R command.

```
> print(summary(mc3s), digits=2)
Number of studies combined: k=21

                       MD         95%-CI       z  p-value
Fixed effect model   -0.05 [-0.05; -0.04] -10.06 < 0.0001
Random effects model -0.08 [-0.11; -0.05]  -5.82 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.0027; H = 2.63 [2.19; 3.15]; I^2 = 85.5% [79.1%; 89.9%]

Test of heterogeneity:
     Q d.f.  p-value
 138.08   20 < 0.0001

Results for subgroups (fixed effect model):
             k    MD         95%-CI    Q    tau^2   I^2
<= 3 months  4 -0.13 [-0.17; -0.09] 22.43  0.035 86.6%
> 3 months  17 -0.04 [-0.05; -0.03] 94.92  0.002 83.1%

Test for subgroup differences (fixed effect model):
                  Q d.f.  p-value
```

---

[8]R function `update` is a generic function like `print` or `summary`.

```
Between groups  20.73    1 < 0.0001
Within groups  117.35   19 < 0.0001

Results for subgroups (random effects model):
               k    MD          95%-CI    Q   tau^2   I^2
<= 3 months    4 -0.28 [-0.50; -0.05] 22.43  0.035 86.6%
> 3 months    17 -0.06 [-0.09; -0.04] 94.92  0.002 83.1%

Test for subgroup differences (random effects model):
                    Q d.f.  p-value
Between groups   3.41    1   0.0647

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

The results for the fixed effect model show that between-group heterogeneity is highly statistically significant ($Q = 20.73$ on 1 degrees of freedom) as well as within-group heterogeneity ($Q = 117.35$, 19 degrees of freedom). Further, the fixed effect estimates (-0.13, short duration; -0.04, long duration) are not that different. While short duration studies seem to have far fewer patients, the effect appears similar; study duration does not appear to be the source of the high degree of heterogeneity in these data. This observation is supported by the results for the random effects model (between-study heterogeneity: $Q = 3.41$, 1 degrees of freedom).

A forest plot with subgroups for length of duration, which is shown in Fig. 2.8, can be produced using the following R command.

```
> forest(mc3s, xlim=c(-0.5, 0.2),
+         xlab="Difference in mean number of acute exacerbations
          per month")
```

The argument `subset` which has been used before to select a single study can also be used to conduct a meta-analysis of a subgroup of studies, e.g. for studies with short study duration:

```
> print(metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                 subset=duration=="<= 3 months",
+                 studlab=paste(author, year)),
+       digits=2)

*** Output truncated ***
Number of studies combined: k=4

                        MD          95%-CI     z  p-value
Fixed effect model   -0.13 [-0.17; -0.09] -6.78 < 0.0001
Random effects model -0.28 [-0.50; -0.05] -2.43   0.0153
*** Output truncated ***
```

Or alternatively using the `update.meta` function:

```
> print(update(mc3, subset=duration=="<= 3 months"),
+       digits=2)
```

Study	Experimental Total	Mean	SD	Control Total	Mean	SD		MD	95%-CI	W(fixed)	W(random)
<= 3 months											
Bontognali 1991	30	0.70	3.76	30	1.27	4.58		−0.57	[−2.69; 1.55]	0.0%	0.0%
Castiglioni 1986	311	0.10	0.21	302	0.20	0.29		−0.10	[−0.14; −0.06]	4.9%	6.2%
Cremonini 1986	21	0.25	0.23	20	0.71	0.29		−0.46	[−0.62; −0.30]	0.3%	2.1%
Grassi 1994	42	0.16	0.29	41	0.45	0.43		−0.29	[−0.45; −0.13]	0.3%	2.1%
Jackson 1984	61	0.11	0.00	60	0.13	0.00		−0.02		0.0%	0.0%
Fixed effect model	465			453				−0.13	[−0.17; −0.09]	5.5%	--
Random effects model								−0.28	[−0.50; −0.05]	--	10.3%
Heterogeneity: I–squared=86.6%, tau–squared=0.035, p<0.0001											
> 3 months											
Allegra 1996	223	0.07	0.11	218	0.11	0.14		−0.04	[−0.06; −0.02]	14.2%	6.7%
Babolini 1980	254	0.13	0.18	241	0.33	0.27		−0.20	[−0.24; −0.16]	4.8%	6.1%
Boman 1983	98	0.20	0.27	105	0.32	0.30		−0.12	[−0.20; −0.04]	1.3%	4.5%
Borgia 1981	10	0.05	0.08	9	0.15	0.17		−0.10	[−0.22; 0.02]	0.5%	3.0%
Decramer 2005	256	0.10	0.11	267	0.11	0.16		−0.01	[−0.03; 0.01]	14.3%	6.7%
Grassi 1976	35	0.14	0.15	34	0.27	0.21		−0.13	[−0.22; −0.04]	1.1%	4.2%
Grillage 1985	54	0.10	0.00	55	0.12	0.00		−0.02		0.0%	0.0%
Hansen 1994	59	0.11	0.15	70	0.16	0.19		−0.05	[−0.11; 0.01]	2.3%	5.4%
Malerba 2004	115	0.06	0.08	119	0.07	0.08		−0.01	[−0.03; 0.01]	18.7%	6.8%
McGavin 1985	72	0.42	0.34	76	0.52	0.35		−0.10	[−0.21; 0.01]	0.6%	3.3%
Meister 1986	90	0.15	0.15	91	0.20	0.19		−0.05	[−0.10; 0.00]	3.2%	5.7%
Meister 1999	122	0.06	0.15	124	0.10	0.15		−0.04	[−0.08; 0.00]	5.6%	6.3%
Moretti 2004	63	0.12	0.14	61	0.17	0.17		−0.05	[−0.10; 0.00]	2.6%	5.5%
Nowak 1999	147	0.03	0.06	148	0.06	0.12		−0.03	[−0.05; −0.01]	16.8%	6.8%
Olivieri 1987	110	0.18	0.31	104	0.33	0.41		−0.15	[−0.25; −0.05]	0.8%	3.7%
Parr 1987	243	0.18	0.21	210	0.21	0.21		−0.03	[−0.07; 0.01]	5.2%	6.2%
Pela 1999	83	0.17	0.18	80	0.29	0.32		−0.12	[−0.20; −0.04]	1.2%	4.4%
Rasmussen 1988	44	0.13	0.21	47	0.14	0.19		−0.01	[−0.09; 0.07]	1.2%	4.3%
Fixed effect model	2078			2059				−0.04	[−0.05; −0.03]	94.5%	--
Random effects model								−0.06	[−0.09; −0.04]	--	89.7%
Heterogeneity: I–squared=83.1%, tau–squared=0.002, p<0.0001											
Fixed effect model	**2543**			**2512**				**−0.05**	**[−0.05; −0.04]**	**100%**	**--**
Random effects model								**−0.08**	**[−0.11; −0.05]**	**--**	**100%**
Heterogeneity: I–squared=85.5%, tau–squared=0.0027, p<0.0001											

−0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2
Difference in mean number of acute exacerbations per month
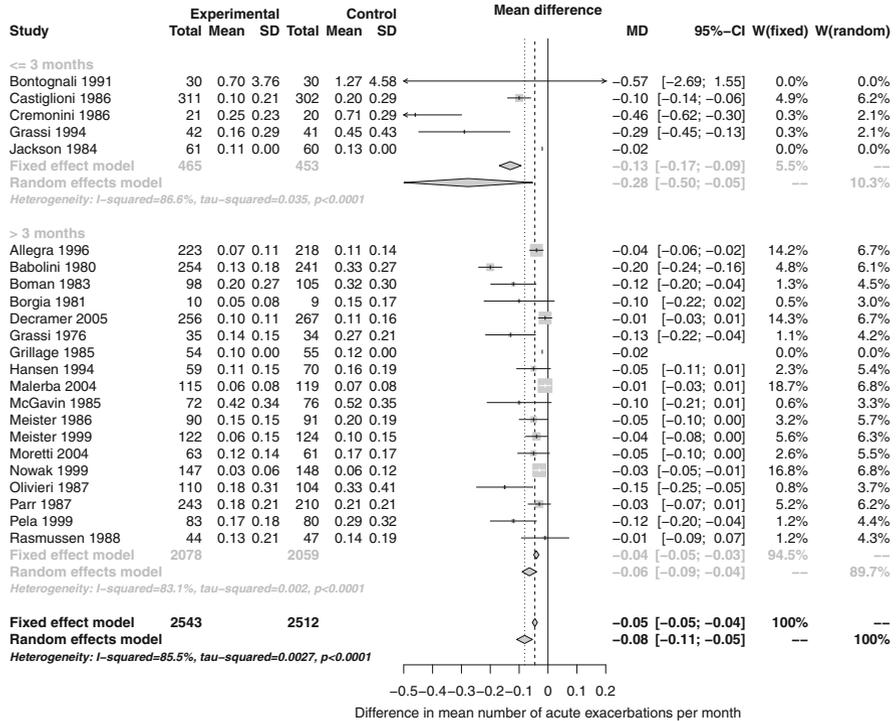
Fig. 2.8 Subgroup analysis for mucolytic agents data [31]

```
*** Output truncated ***
                      MD        95%-CI      z   p-value
Fixed effect model  -0.13 [-0.17; -0.09] -6.78 < 0.0001
Random effects model -0.28 [-0.50; -0.05] -2.43   0.0153
*** Output truncated ***
```

These are exactly the same treatment estimates and confidence intervals for fixed effect and random effects model, respectively, in studies with short duration as shown in the upper part of Fig. 2.8.   □

2.6 Meta-Analysis of Other Outcomes

In this section, the application of the generic inverse variance method to other outcomes will be described. All examples use the metagen function to conduct the meta-analysis. Other functions are available in R package **meta** for specific

outcomes:

- `metacor` function for meta-analysis of correlations,
- `metainc` function for meta-analysis of incidence rate ratios,
- `metaprop` function for meta-analysis of single proportions.

The first two R functions are not covered in this book and the `metaprop` function is only briefly used in Chap. 9 to calculate confidence intervals for sensitivities and specificities. The corresponding help pages of these functions give further details on these methods as well as a couple of examples.

2.6.1 Meta-Analysis with Survival Outcomes

Statistical methods for binary data are described in detail in Chap. 3. Very often not only the information that an event occurred but also when the event happened is of central interest. This type of data is called time-to-event or survival data if the event of interest is death. Time to an event is a continuous quantity, however, in contrast to the examples with continuous outcomes used so far time to an event can typically not be observed for all participants as the maximum follow-up time is limited in a study. Patients where the event of interest did not occur during the follow-up period are called *censored* observations. Censoring is a distinguishing feature of time-to-event data. Another important aspect of time-to-event data, not covered in this book, are competing events, e.g. time to either cardiovascular or non-cardiovascular death. In this situation only the time to death either due to a cardiovascular or non-cardiovascular reason can be observed. Specific statistical methods for survival data have been developed [2, 24] and should be used in the analysis.

In survival analysis the hazard function, i.e. a function describing the instantaneous risk of dying given survival up to a specific timepoint, plays a central role. To compare two treatments the hazard ratio, i.e. a ratio of hazard functions, is typically used. The interpretation of a hazard ratio is similar to a risk ratio which is introduced in Sect. 3.1.2.

A meta-analysis with survival time outcomes is typically based on the hazard ratio as measure of treatment effect [26]. Accordingly, the logarithm of the hazard ratio and its standard error are the basic quantities utilised in meta-analysis. As hazard ratio and corresponding standard error are not always reported in publications, several methods exist to derive these quantities, e.g. from published survival curves [26, 42].

The generic inverse variance method can be used straightforward with log hazard ratio $\hat{\theta}_k$ and its standard error S.E. $(\hat{\theta}_k)$, for study k, $k = 1, \ldots, K$.

Using these quantities, all methods described in Sects. 2.2 and 2.3 can be used for meta-analysis. In the following example we consider the most basic case, i.e. fixed effect and random effects meta-analysis using the DerSimonian–Laird method to estimate the between-study variance τ^2.

```
> # 1. Read in the data
> data4 <- read.csv("dataset04.csv")
> # 2. Print data
> data4
       author year  Ne   Nc    logHR selogHR
1 FCG on CLL 1996   53   52  -0.5920  0.3450
2  Leporrier 2001  341  597  -0.0791  0.0787
3        Rai 2000  195  200  -0.2370  0.1440
4      Robak 2000  133  117   0.1630  0.3120
```

Fig. 2.9 Data from meta-analysis of single-agent purine analogues for the treatment of chronic lymphocytic leukaemia [38]

Example 2.13 Steurer et al. [38] conducted a Cochrane review to evaluate the effect of single-agent purine analogues for the treatment of chronic lymphocytic leukaemia. Data for the main outcome overall survival are reported in Fig. 2.9. Columns `logHR` and `selogHR` correspond to the log hazard ratio and its standard error.

The following R command can be used to conduct a meta-analysis using the generic inverse variance method.

```
> mg1 <- metagen(logHR, selogHR,
+                studlab=paste(author, year), data=data4,
+                sm="HR")
```

Specifying argument `sm="HR"`, it is assumed that hazard ratios are entered on the log scale. If hazard ratios instead of log hazard ratios are available in a dataset, the base `log` function can be used to transform the hazard ratio, e.g. `metagen(log(HR), ...)`. Regardless of the input of hazard ratios or log hazard ratios, the `metagen` function expects that the standard error from the log hazard ratio and not the standard error of the hazard ratio is provided as input for argument `seTE`. Note, sample sizes given in columns `Ne` and `Nc` in Fig. 2.9 are not utilised in the calculations.

As usual we can print the results of the meta-analysis.

```
> print(mg1, digits=2)
                  HR        95%-CI %W(fixed)  %W(random)
FCG on CLL 1996 0.55 [0.28; 1.09]      3.68        5.85
Leporrier 2001  0.92 [0.79; 1.08]     70.70       59.76
Rai 2000        0.79 [0.59; 1.05]     21.12       27.32
Robak 2000      1.18 [0.64; 2.17]      4.50        7.08

Number of studies combined: k=4

                      HR        95%-CI      z  p-value
Fixed effect model   0.89 [0.78; 1.01]  -1.82   0.0688
Random effects model 0.87 [0.74; 1.03]  -1.58   0.1142

Quantifying heterogeneity:
tau^2 = 0.0061; H = 1.1 [1; 2.81]; I^2 = 17.2% [0%; 87.3%]
```

```
Test of heterogeneity:
    Q d.f.  p-value
 3.62    3   0.3049

Details on meta-analytical method:
– Inverse variance method
– DerSimonian–Laird estimator for tau^2
```

These results correspond to those reported in [38].   □

2.6.2 Meta-Analysis of Cross-Over Trials

Until now methods have been described to conduct a meta-analysis of trials comparing two parallel treatment groups. Cross-over trials are another popular design to compare treatments [23]. In a cross-over trial each participant serves as his/her own control. Accordingly, between-patient variation is removed from the treatment comparison resulting in a smaller number of patients to achieve the same statistical power. A typical setting for a cross-over trial is chronic but stable diseases, i.e. a patient neither gets cured nor does the condition (dramatically) worsen over time.

In a simple cross-over design, a patient is randomly assigned to treatment sequence AB or BA, i.e. either receiving treatment A first and "cross-over" to treatment B or vice versa. Typically, the first and second treatment period are separated by a so-called washout period such that the effect of the treatment effect in the first treatment period is not carried over to the second treatment period. In principle, longer sequences of two treatments A and B are possible, e.g. ABBA. Note, the first period of a cross-over trial is equivalent to a parallel group study design.

Statistical methods for meta-analysis of cross-over trials and the combination of parallel group and cross-over trials have been described in a series of papers in Statistics in Medicine [5, 6, 8]. For the meta-analysis of cross-over trials with a continuous outcome the generic inverse variance method can be used [5].

Example 2.14 Curtin et al. [5, Table 2] report the results of 12 parallel group and 21 cross-over trials to evaluate the effect of potassium supplementation on the reduction of systolic and diastolic blood pressure. Here, we only look at the 21 cross-over trials and diastolic blood pressure as outcome of interest.

Mean difference in diastolic blood pressure (column mean) and its standard error (SE) as well as the within-patient correlation (corr) are given in Fig. 2.10. Correlations are not utilised in the meta-analysis, however, the values give some indication on the gain in precision by using a cross-over design. All correlations are above zero and ranging from 0.29 to 0.88. Accordingly, using a cross-over design results in a gain in precision in all trials.

```
> # 1. Read in the data
> data5 <- read.csv("dataset05.csv")
> # 2. Print data
> data5
                  author  year  N   mean  SE corr
1        Skrabal et al. 1981a 20  -4.5 2.1 0.49
2        Skrabal et al. 1981b 20  -0.5 1.7 0.54
3      MacGregor et al.  1982 23  -4.0 1.9 0.41
4        Khaw and Thom  1982 20  -2.4 1.1 0.83
5      Richards et al.  1984 12  -1.0 3.4 0.50
6          Smith et al.  1985 20   0.0 1.9 0.50
7        Kaplan et al.  1985 16  -5.8 1.6 0.65
8        Zoccali et al.  1985 23  -3.0 3.0 0.50
9        Matlou et al.  1986 36  -3.0 1.5 0.61
10        Barden et al.  1986 44  -1.5 1.4 0.44
11   Poulter and Sever  1986 19   2.0 2.2 0.36
12        Grobbee et al. 1987 40  -0.3 1.5 0.61
13        Krishna et al. 1989 10  -8.0 2.2 0.48
14 Mullen and O'Connor 1990a 24   3.0 2.0 0.50
15 Mullen and O'Connor 1990b 24   1.4 2.0 0.50
16          Patki et al. 1990 37 -13.1 0.7 0.53
17        Valdes et al.  1991 24  -3.0 2.0 0.50
18        Barden et al.  1991 39  -0.6 0.6 0.88
19     Overlack et al.  1991 12   3.0 2.0 0.50
20          Smith et al. 1992 22  -1.7 2.5 0.29
21 Fotherby and Potter  1992 18  -6.0 2.5 0.81
```

Fig. 2.10 Data from meta-analysis of potassium supplementation for blood pressure reduction [5]

The following R code can be used for the meta-analysis of these cross-over trials

```
> mg2 <- metagen(mean, SE, studlab=paste(author, year),
+                data=data5, sm="MD")
```

which yields the results

```
> print(summary(mg2), digits=2)
Number of studies combined: k=21

                     MD         95%-CI      z  p-value
Fixed effect model   -3.71 [-4.32; -3.11] -12.03 < 0.0001
Random effects model -2.38 [-4.76; -0.01]  -1.96   0.0495

Quantifying heterogeneity:
tau^2 = 27.03; H = 3.66 [3.14; 4.25]; I^2 = 92.5% [89.9%; 94.5%]

Test of heterogeneity:
     Q d.f.  p-value
 267.24   20 < 0.0001
```

```
Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

Both fixed effect and random effects model show a statistically significant reduction in diastolic blood pressure for potassium supplementation. Due to the very large between-study heterogeneity the confidence interval for the random effects estimate is much wider than the confidence interval for the fixed effect estimate. Accordingly, the *p*-value for the random effects model is much larger.

Results for the fixed effect model have also been reported in [5, Table 3] and are almost identical.   □

2.6.3   Meta-Analysis of Adjusted Treatment Effects

Another application of the generic inverse variance method is a meta-analysis of adjusted treatment effects, e.g. adjusted log odds ratios from a logistic regression model [1] or log hazard ratios from a Cox regression model [24].

Example 2.15 Greenland and Longnecker [12] describe a method to combine trend estimates from summarised dose–response data. A meta-analysis of 16 case–control studies evaluating the impact of alcohol consumption on breast cancer risk was used as an illustrative example (see [12, Table 3]).

Data for these studies are given in Fig. 2.11. For meta-analysis the adjusted log risk ratio (column b) and its standard error (SE) are utilised. In order to report results as log risk ratios like the authors [12] we use argument backtransf=FALSE.

```
> # 1. Read in the data
> data6 <- read.csv("dataset06.csv")
> # 2. Print data
> data6
                      author year          b       SE
1         Hiatt and Bawol 1984   0.004340  0.00247
2            Hiatt et al. 1988   0.010900  0.00410
3           Willett t al. 1987   0.028400  0.00564
4        Schatzkin et al. 1987   0.118000  0.04760
5           Harvey et al. 1987   0.012100  0.00429
6        Rosenberg et al. 1982   0.087000  0.02320
7          Webster et al. 1983   0.003110  0.00373
8  Paganini-Hill and Ross 1983   0.000000  0.00940
9          Byers and Funch 1982   0.005970  0.00658
10    Rohan and McMichael 1988   0.047900  0.02050
11          Talamini et al. 1984   0.038900  0.00768
12          O'Connell et al. 1987   0.203000  0.09460
13       Harris and Wynder 1988  -0.006730  0.00419
14                 Le et al. 1984   0.011100  0.00481
15         La Vecchia et al. 1985   0.014800  0.00635
16                Begg et al. 1983  -0.000787  0.00867
```

Fig. 2.11 Data from meta-analysis evaluating impact of alcohol consumption on breast cancer risk [12]

```
> mg3 <- metagen(b, SE, studlab=paste(author, year),
+                data=data6, sm="RR", backtransf=FALSE)
```

The results for the meta-analysis are as follows.

```
> summary(mg3)
Number of studies combined: k=16

                      logRR          95%-CI      z  p-value
Fixed effect model   0.0082 [0.0056; 0.0108] 6.2409 < 0.0001
Random effects model 0.0131 [0.0062; 0.0199] 3.7298   0.0002

Quantifying heterogeneity:
tau^2 = 0.0001; H = 2.24 [1.78; 2.82]; I^2 = 80.1% [68.5%; 87.4%]

Test of heterogeneity:
    Q d.f.  p-value
 75.31   15 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

As we used argument `backtransf=FALSE`, treatment estimates are reported on the log scale (see `logRR` in the printout). Results for the fixed effect model are identical to those reported in [12].   □

2.7 Summary

In this chapter the generic inverse variance method and its application in meta-analysis has been described in detail using continuous outcomes. Both fixed effect and random effects methods have been introduced. We have shown how typical data can be used with the `metacont` and `metagen` function, respectively, and how the results of a meta-analysis can be printed and plotted.

We also discussed various methods for estimating the between-study variance τ^2 and the Hartung–Knapp adjustment has been described as an alternative method to the classic random effects method. Furthermore, we have illustrated the use of the `byvar` option, which makes subgroup analysis straightforward. More details on tests for subgroup differences are provided in Sect. 4.3.

Lastly, the generic inverse variance method has been used in very different settings (survival outcomes, cross-over trials, adjusted treatment effects) indicating the wide applicability of the method.

In the next chapter, we describe the analogue of these analyses for binary data.

References

1. Agresti, A.: Categorical Data Analysis, 2nd edn. Wiley, New York (2002)
2. Beyersmann, J., Allignol, A., Schumacher, M.: Competing Risks and Multistate Models with R. Springer, New York (2012)
3. Cooper, H., Hedges, L.V. (eds.): The Handbook of Research Synthesis. Russell Sage Foundation, New York (1994)
4. Cornell, J.E., Mulrow, C.D., Localio, R., Stack, C.B., Meibohm, A.R., Guallar, E., Goodman, S.N.: Random-effects meta-analysis of inconsistent effects: a time for change. Ann. Intern. Med. **160**(4), 267–270 (2014)
5. Curtin, F., Altman, D.G., Elbourne, D.: Meta-analysis combining parallel and cross-over clinical trials. I: Continuous outcomes. Stat. Med. **21**, 2131–2144 (2002)
6. Curtin, F., Elbourne, D., Altman, D.G.: Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes. Stat. Med. **21**, 2145–2159 (2002)
7. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Control Clin. Trials **7**, 177–188 (1986)
8. Elbourne, D.R., Altman, D.G., Higgins, J.P.T., Curtin, F., Worthington, H.V., Vail, A.: Meta-analyses involving cross-over trials: methodological issues. Int. J. Epidemiol. **31**, 140–149 (2002)
9. Fleiss, J.L.: The statistical basis of meta-analysis. Stat. Methods Med. Res. **2**, 121–145 (1993)
10. Furukawa, T.A., McGuire, H., Barbui, C.: Low dosage tricyclic antidepressants for depression. Cochrane Database Syst. Rev. (3) (2003). Art. No. CD003197. doi:10.1002/14651858.CD003197
11. Greenland, S.: Invited commentary: a critical look at some popular meta-analytic methods. Am. J. Epidemiol. **140**, 290–296 (1994)
12. Greenland, S., Longnecker, M.P.: Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am. J. Epidemiol. **135**, 1301–1309 (1992)
13. Hartung, J.: An alternative method for meta-analysis. Biom. J. **41**, 901–916 (1999)
14. Hartung, J., Knapp, G.: A refined method for the meta-analysis of controlled clinical trials with binary outcome. Stat. Med. **20**, 3875–3889 (2001)
15. Hedges, L.V.: Distribution theory for glass's estimator of effect size and related estimators. J. Educ. Behav. Stat. **6**(2), 107–128 (1981)
16. Hedges, L.V.: Estimation of effect size from a series of independent experiments. Psychol. Bull. **92**(2), 490–499 (1982)
17. Hedges, L.: A random effects model for effect sizes. Psychol. Bull. **93**(2), 388–395 (1983)
18. Hedges, L.V., Olkin, I.: Statistical Methods for Meta-Analysis. Academic, San Diego (1985)
19. Higgins, J.P., Green, S. (eds.): Cochrane Handbook for Systematic Reviews of Interventions – Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. http://www.cochrane-handbook.org (2011)
20. Higgins, J.P.T., Thompson, S.G.: Quantifying heterogeneity in a meta-analysis. Stat. Med. **21**, 1539–1558 (2002)
21. Higgins, J.P., Thompson, S.G., Spiegelhalter, D.J.: A re-evaluation of random-effects meta-analysis. J. R. Stat. Soc. **172**, 137–159 (2009)
22. Hunter, J.E., Schmidt, F.L.: Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, 2nd edn. Sage, Thousand Oaks (2004)
23. Jones, B., Kenward, M.G.: Design and Analysis of Cross-Over Trials. Chapman & Hall/CRC, Boca Raton (2003)
24. Klein, J.P., Moeschberger, M.L.: Survival Analysis. Techniques for Censored and Truncated Data. Springer, New York (2005)
25. Knapp, G., Hartung, J.: Improved tests for a random effects meta-regression with a single covariate. Stat. Med. **22**, 2693–2710 (2003)
26. Parmar, M.K.B., Torri, V., Stewart, L.: Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat. Med. **17**, 2815–2834 (1998)

27. Paule, R., Mandel, J.: Consensus values and weighting factors. J. Res. Natl. Bur. Stand. **87**(5), 377–385 (1982)
28. Pocock, S.: Editorials. Stat. Methods Med. Res. **2**, 117–119 (1993)
29. Pocock, S.J.: Safety of drug-eluting stents: demystifying network meta-analysis. Lancet **370**, 2099–2100 (2007)
30. Poole, C., Greenland, S.: Random-effects meta-analysis are not always conservative. Am. J. Epidemiol. **150**, 469–75 (1999)
31. Poole, P.J., Black, P.N.: Mucolytic agents for chronic bronchitis or chronic obstructive pulmonary disease. Cochrane Database Syst. Rev. (3) (2006). Art. No. CD001287. doi:10.1002/14651858.CD001287.pub2
32. Rücker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M.: Undue reliance on I^2 in assessing heterogeneity may mislead. BMC Med. Res. Methodol. **8**, 79 (2008). http://www.biomedcentral.com/1471-2288/8/79. doi:10.1186/1471-2288-8-79
33. Schwarzer, G.: meta: an R package for meta-analysis. R News **7**(3), 40–45 (2007). http://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf
34. Schwarzer, G.: meta: Meta-Analysis with R. R package version 4.0-2. URL http://cran.R-project.org/package=meta (2014)
35. Sidik, K., Jonkman, J.N.: Simple heterogeneity variance estimation for meta-analysis. J. R. Stat. Soc. Ser. C **54**(2), 367–384 (2005)
36. Sidik, K., Jonkman, J.N.: A comparison of heterogeneity variance estimators in combining results of studies. Stat. Med. **26**(9), 1964–1981 (2007). 10.1002/sim.2688. http://dx.doi.org/10.1002/sim.2688
37. Spooner, C., Saunders, L.D., Rowe, B.H.: Nedocromil sodium for preventing exercise-induced bronchoconstriction. Cochrane Database Syst. Rev. (1) (2002). Art. No. CD001183. doi:10.1002/14651858.CD001183
38. Steurer, M., Pall, G., Richards, S., Schwarzer, G., Bohlius, J., Greil, R.: Single-agent purine analogues for the treatment of chronic lymphocytic leukaemia: a systematic review and meta-analysis. Cancer Treat. Rev. **32**(5), 377–389 (2006)
39. Stijnen, T., Van Houwelingen, J.C.: Empirical Bayes methods in clinical trials meta-analysis. Biom. J. **32**(3), 335–346 (1990)
40. The Cochrane Collaboration: Review Manager (RevMan) [Computer program]. Version 5.3. The Nordic Cochrane Centre, Copenhagen (2014)
41. Thompson, S.G.: Controversies in meta-analysis: the case of trials of serum cholesterol reduction. Stat. Methods Med. Res. **2**, 173–192 (1993)
42. Tierney, J.F., Stewart, L.A., Ghersi, D., Burdett, S., Sydes, M.R.: Practical methods for incorporating summary time-to-event data into meta-analysis. Trials **8**, 16 (2007). doi:10.1186/1745-6215-8-16
43. Viechtbauer, W.: Bias and efficiency of meta-analytic variance estimators in the random-effects model. J. Educ. Behav. Stat. **30**, 261–293 (2005)