Guido Schwarzer
James R. Carpenter
Gerta Rücker

Meta-Analysis with R

Use R!

More information about this series at http://www.springer.com/series/6991

Guido Schwarzer • James R. Carpenter •
Gerta Rücker

Meta-Analysis with R

Guido Schwarzer
Institute for Medical Biometry
  and Statistics
Medical Center – University of Freiburg
Freiburg, Germany

Gerta Rücker
Institute for Medical Biometry and Statistics
Medical Center – University of Freiburg
Freiburg, Germany

James R. Carpenter
MRC Clinical Trials Unit, London
  and London School of Hygiene
  and Tropical Medicine
London, United Kingdom

Printed on acid-free paper

Preface

Meta-analysis plays a key role in evidence synthesis in many research disciplines, not least the social sciences, medicine and economics. The aim of this book is to equip those involved in such work (who are often not trained statisticians) to use R for meta-analysis, and thus promote both the use of R and the latest statistical methods in this area.

The attractions of R in this context (besides its free availability from http://www.r-project.org/) are its fast yet powerful and flexible graphics and its well-established algorithmic base.

The book assumes no prior knowledge of R, and takes readers through every step of the way from installing R, loading data from other packages, performing and interpreting the analyses. Parts I and II cover the essentials, while Part III considers more advanced topics, which remain the subject of active research.

Throughout, the ideas are illustrated with examples, and all the codes necessary to repeat these examples (including creating all the plots in the book) are either in the text itself or the web-appendix http://meta-analysis-with-r.org/. In selecting the code to include in the main text, we have assumed readers are relatively new to R. More experienced users can easily skip over familiar material.

<table>
<tr><td>Freiburg, Germany</td><td>Guido Schwarzer</td></tr>
<tr><td>London, UK</td><td>James R. Carpenter</td></tr>
<tr><td>Freiburg, Germany</td><td>Gerta Rücker</td></tr>
<tr><td>December, 2014</td><td></td></tr>
</table>

Acknowledgements

Contents

Part I
Getting Started

Chapter 1
An Introduction to Meta-Analysis in R

The world is awash with information. For any question, the briefest of internet searches will throw up a range of frequently contradictory answers. This underlies increasing awareness of the value of systematic evidence synthesis—both qualitative and quantitative—by researchers, policy makers and the broader public. It is reflected in the continuing development of the Cochrane Collaboration (http://www.cochrane.org/), an international collaboration devoted to undertaking, publishing and promoting systematic evidence synthesis [2].

Quantitative aspects have a key role to play in evidence synthesis. The statistical methodology for combining quantitative evidence from studies, known as meta-analysis, therefore features in almost every systematic review, and continues to undergo rapid development. A major concern of such developments is the need to detect and adjust for possible biases, as well as the synthesis of evidence from studies that have compared different combinations of interventions and measured these interventions with different outcomes.

The statistical software environment R [4] is now firmly established, and is one of the most widely used software packages for both the development of statistical methodology and day-to-day data analysis using established methods. It features excellent graphics and is readily extensible. It is freely available, both as source code and as compiled binaries for Windows, Linux, and Mac OS. In addition to the broad range of inbuilt functions, over 6000 additional packages are available on the Comprehensive R Archive Network (CRAN). R has been registered for use in clinical trials (http://www.r-project.org/doc/R-FDA.pdf) and has its own journal "The R Journal" (http://journal.r-project.org/).

We believe that R is a natural choice for meta-analysis, as it provides the greatest range of methodology for meta-analysis in any single statistical software package. Our aim in writing this book is to introduce data analysts involved in meta-analysis, whether they are existing R users or not, to this software. No previous knowledge of R is assumed, and throughout the development is illustrated with worked examples.
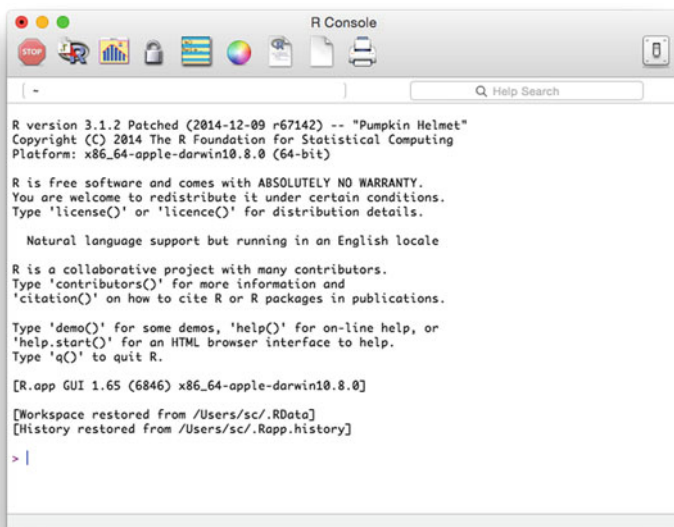
The data, and script files to repeat the examples, are downloadable from the website http://meta-analysis-with-r.org/.

Established R users can probably skip the remainder of this chapter, in which we introduce some key R concepts and illustrate them with a simple meta-analysis. The aim in what follows is to get new users (who have used other statistics packages) started with R so that they can follow the presentation in the remainder of the book. For a more comprehensive introduction we recommend [1].

The remainder of this chapter is best read alongside a computer running R. Thus, if you have not yet installed R, we would encourage you to do so: full installation details are given in Appendix A.1.

1.1   Getting Started with R

To start R either press on the R desktop icon created during installation or select the corresponding entry in the menu of installed applications. This should open the R Console window, an example of which in Mac OS X is shown in Fig. 1.1. R is not a menu driven program, so the user must either type commands at the command line, or execute a pre-written script. R is case sensitive, and most function names are in lower case. Typing the function name shows the code; to execute a function we must type parentheses. For example,

Fig. 1.1 Screenshot of R Console window under Mac OS X

- type demo (followed by <RETURN>), and R will print the code for the demo function,
- type demo() (followed by <RETURN>), and R will open a window showing some of the demonstrations available.

Towards the top of the list is the graphics demonstration. To run this, we need to give the argument graphics to the demo function, by placing it in the parentheses: demo(graphics) <RETURN>. Try this; you will need to press <RETURN> twice more to start the demonstration, and to move through the various plots. In summary: we pass information to R functions through the arguments which go between the parentheses that immediately follow function names.

1.1.1 Quitting R

To exit R at the command line type q(). You will be asked Save workspace image? [y/n/c]: . Type one of y (yes: quit and save your data); n (no: quit without saving your data) or c (continue with R session), followed by <RETURN>. An alternative way to exit R is using the menu *File → Exit* (under Windows) or *R → Quit* (under Mac OS).

1.1.2 R as a Calculator

To begin, we note the following:

- Commands in R are typed into the R Console window to the right of the *command prompt*: >
- If R detects a command is not finished (for example, parentheses or quotes are not matched) then the next line will show the *continuation prompt*: +
  Either complete the command, or interrupt it by typing Ctrl+C under Linux.
- The "#" means that the remainder of the line will not be executed: we use this repeatedly to give a commentary on our R code.

Armed with this, we can start to explore R. In the R Console try the following commands (you do not need to type lines that begin with the "#" sign):

```
> # set x to be 2
> x <- 2
> # same assignment
> x = 2
> # display x
> x
[1] 2
> # set x to be the vector 1,2,3 using R function c
> x <- c(1,2,3)
> # same assignment
> x <- 1:3
> # display x
```

```
> x
[1] 1 2 3
> # add up the elements of x:
> sum(x)
[1] 6
> # set x to be a matrix
> x <- matrix(c(1,2,3,4), nrow=2)
> # display x
> x
     [,1] [,2]
[1,]    1    3
[2,]    2    4
> # square the matrix x
> x*x
     [,1] [,2]
[1,]    1    9
[2,]    4   16
> # square x and add 3 to all entries
> x*x + 3
     [,1] [,2]
[1,]    4   12
[2,]    7   19
> # create a function, called f, which squares its argument
> f <- function(x){x*x}
> f(2)
[1] 4
> f(c(1,2,3))
[1] 1 4 9
> # square the matrix x using R function f
> f(x)
     [,1] [,2]
[1,]    1    9
[2,]    4   16
```

1.1.3 Getting Help

The easiest way to get help on a function, for example the `ls` function is to type a
"?" followed by the function name at the R prompt: `?ls`. Equivalently, we can type
`help(ls)`. Either will open a window giving details of how to use the function.
For more extensive help, the command `help.start()` opens a window or tab in
your web browser with links to all the documentation.

1.2 Loading, Saving and Restoring Data

Data are stored in R in a slightly different way to many other packages: instead
of having a single dataset available at a time, we have instead a *workspace* (also
called global environment or environment) that may be thought of as a "virtual

library" or "virtual directory", which can contain many datasets, together with other R objects such as functions. This workspace is stored as a file, with default name .RData under Mac/Linux/Unix and _RData under Windows. Not only can we have separate .RData files for different projects, we can make the contents of as many as we like available to R. This can cause some confusion if they contain different objects of the same name!

For this book, we will only need one .RData file. This we store in the current *Working directory*. To find the current working directory type getwd() at the R command line.

Within R, objects are accessed by typing their names. Typical objects include numbers, variables, vectors, matrices, data frames (approximately equivalent to datasets in other packages) and functions.

To see the objects in the attached .RData workspace, use the ls function. After doing the graphics demonstration in Sect. 1.1, you should see the following listing of R objects if the workspace has been empty before; otherwise some additional R objects will be listed.

```
> ls()
 [1] "g"      "lev" "n"    "opar"    "pie.sales" "pin"
 [7] "scale" "usr" "x"    "xadd"    "xdelta"    "xscale"
[13] "xx"     "y"   "yadd" "ydelta" "yscale"     "yy"
```

In order to delete R object g, you can use R function rm:

```
> rm(g)
> ls()
 [1] "lev" "n"    "opar"    "pie.sales" "pin"     "scale"
 [7] "usr" "x"    "xadd"    "xdelta"    "xscale" "xx"
[13] "y"   "yadd" "ydelta" "yscale"     "yy"
```

To delete all R objects in the workspace (without any further warning!) you can use the following command:

```
> rm(list=ls())
> ls()
character(0)
```

The result "character(0)" means that no R objects are found in the workspace.

Note, changes to the workspace are not permanent as long as you do not save them in the file .RData. If R stops due to an internal error (which almost never happens) or if you kill the R process, all changes to the workspace will be lost.

You can save the workspace at any time in an R session using R command save.image() which saves the file .RData in the current working directory. When you quit R (see Sect. 1.1.1) you also have the option to save the workspace in the .RData file.

To change the current working directory, type `setwd("my_directory")`, replacing `my_directory` with the path to a pre-existing directory on your computer.[1] Alternatively, the current working directory may be changed from the toolbar under Mac OS and Windows.

We will now go through a sequence of commands to load, view and save data from a meta-analysis. We will then go through a second sequence: restart R, reload the data and view it. First, check the current working directory. At the command prompt type `getwd()`. Next, outside R download the file `dataset01.csv` from the website http://meta-analysis-with-r.org/ and save it in the current working directory. These data come from a meta-analysis comparing Nedocromil sodium with placebo for preventing exercise-induced bronchoconstriction [7]. To check it is there, use the following command:

```
> list.files(pattern="dataset01")
[1] "dataset01.csv"
```

This will list all files in the current working directory containing the text "dataset01", in this case only the file "dataset01.csv".

We can now load these data into R with the commands given in Fig. 1.2. The `str` command shows that R object `data1` is a data frame, i.e. a dataset. The data consist of 17 studies; for each we have

- the `author`,
- the `year` of publication,
- `Ne`, the Number of patients (sample size) in the experimental treatment group,
- `Me`, the Mean response among the patients in the experimental group,
- `Se`, the Standard deviation of the response in the experimental group, and
- `Nc`, `Mc` and `Sc` the sample size, mean response and standard deviation in the control patients.

Finally, we simply type `save.image()` to save the data. Satisfy yourself this has worked. Quit R not saving the data: `q("no")`. Now restart R, if you are not using the default working directory change to your working directory, then type `load(".RData")` and type `data1`. All being well, the data should still be there.

Note, instead of saving the whole workspace, a subset of R objects, e.g. datasets and/or functions, can be saved in a file for access in future R sessions using the `save` function. For example, in order to save R object `data1` in the file "data1.rda" in the current working directory we can use the command `save(data1, file="data1.rda")`; the extension ".rda" is recommended for files containing R objects. The file `data1` can be loaded into R using the command `load("data1.rda")`.

---

[1]The easiest way to create a new directory is to use the *Explorer* under Windows, the *Finder* under Mac OS, or your preferred file manager under Linux.

```
> # 1. Read in the data
> data1 <- read.csv("dataset01.csv", as.is=TRUE)
> # 2. Print structure of R object data1
> str(data1)
'data.frame': 17 obs. of  8 variables:
 $ author: chr  "Boner" "Boner" "Chudry" "Comis" ...
 $ year  : chr  "1988" "1989" "1987" "1993" ...
 $ Ne    : int  13 20 12 12 17 8 13 12 12 12 ...
 $ Me    : num  13.5 15.7 21.3 14.5 14.4 ...
 $ Se    : num  13.8 13.1 13.1 12.2 11.1 ...
 $ Nc    : int  13 20 12 12 17 8 13 12 12 12 ...
 $ Mc    : num  20.8 22.7 39.7 31.3 27.4 ...
 $ Sc    : num  21.5 16.5 12.9 15.1 17.3 ...
> # 3. To view an R object, just type its name:
> data1
          author year Ne    Me    Se Nc    Mc    Sc
1          Boner 1988 13 13.54 13.85 13 20.77 21.46
2          Boner 1989 20 15.70 13.10 20 22.70 16.47
3         Chudry 1987 12 21.30 13.10 12 39.70 12.90
4          Comis 1993 12 14.50 12.20 12 31.30 15.10
5   DeBenedictis 1994a 17 14.40 11.10 17 27.40 17.30
6   DeBenedictis 1994b  8 14.80 18.60  8 31.40 20.60
7   DeBenedictis 1995 13 15.70 16.80 13 29.60 18.90
8        Debelic 1986 12 29.83 15.95 12 48.08 15.08
9      Henriksen 1988 12 17.50 13.10 12 47.20 16.47
10         Konig 1987 12 12.00 14.60 12 26.20 12.30
11        Morton 1992 16 15.83 13.43 16 38.36 18.01
12      Novembre 1994f 24 15.42  8.35 24 28.46 13.84
13      Novembre 1994s 19 11.00 12.40 19 26.10 14.90
14         Oseid 1995 20 14.10  9.50 20 28.90 18.00
15       Roberts 1985  9 18.90 17.70  9 38.90 18.90
16          Shaw 1985  8 10.27  7.02  8 34.43 10.96
17        Todaro 1993 13 10.10  8.90 13 23.50  4.00
```

Fig. 1.2 Code to read in data for the bronchoconstriction meta-analysis [7] using the `read.csv` function. The structure of the imported dataset is shown with the `str` command. The whole dataset is printed by typing its name

1.2.1   Importing Data from Other Packages

It is usually straightforward to import meta-analysis data from other packages, including the Review Manager 5 (RevMan 5) [9], the program for preparing and maintaining Cochrane Reviews, and Stata [8]. We describe how to do this in Appendix A.2, which we recommend referring to after finishing this chapter.

1.3   Select Variables from an R Dataset

We created the dataset `data1` in the last section and have seen that we can print
the whole dataset by typing its name (followed by <RETURN>). Notice that the
variable names `author`, `year` and so on are not known to R without first referring
to the name of object `data1`. Accordingly, using an R command existing only of
the variable name will result in an error.

```
> author
Error: object 'author' not found
```

Different ways are possible to print data for variable `author`. First, we can use
the dollar sign "$" to select a single variable of a dataset.[2] The following R code
selects variable `author` from dataset `data1`:

```
> data1$author
 [1] "Boner"        "Boner"        "Chudry"       "Comis"
 [5] "DeBenedictis" "DeBenedictis" "DeBenedictis" "Debelic"
 [9] "Henriksen"    "Konig"        "Morton"       "Novembre"
[13] "Novembre"     "Oseid"        "Roberts"      "Shaw"
[17] "Todaro"
```

Second, we can use square brackets to select rows (observations) and columns
(variables) of a dataset.[3] Variable `author` from dataset `data1`, i.e. a column, can
be extracted in the following way:

```
> data1[, "author"]
 [1] "Boner"        "Boner"        "Chudry"       "Comis"
*** Output truncated ***
```

Furthermore, the first four authors/rows of dataset `data1` can be printed either
using

```
> data1[1:4, "author"]
[1] "Boner"  "Boner"  "Chudry" "Comis"
```

or

```
> data1$author[1:4]
[1] "Boner"  "Boner"  "Chudry" "Comis"
```

Third, we can use the `with` function to select the list of authors.

```
> # List the first four authors in data frame data1
> with(data1, author[1:4])
[1] "Boner"  "Boner"  "Chudry" "Comis"
> with(data1[1:4,], author)
[1] "Boner"  "Boner"  "Chudry" "Comis"
```

---

[2]See `help("$")` for more details.

[3]See `help("[")` for more details.

First argument of the `with` function is a dataset, here `data1` and `data1[1:4,]`, respectively. The second argument of the `with` function is an expression that is evaluated within the specified dataset.

Fourth, we can attach a dataset, i.e. make all variables of a dataset available in the search path, using R command `attach`:

```
> # 1. Make variables of data frame data1 directly available:
> attach(data1)
> # 2. List the first four authors in data frame data1
> author[1:4]
 [1] "Boner"      "Boner"      "Chudry"      "Comis"
> # 3. Detach data frame data1
> detach(data1)
```

One disadvantage of this method is that an R object in the workspace may conceal a variable in the attached dataset `data1`. We show this in the next example R code.

```
> # 1. Create a new R object author (numeric variable)
> author <- 123
> # 2. Attaching data frame data1 results in a warning
> attach(data1)
The following object is masked _by_ .GlobalEnv:

    author

> # 3. The following command prints the numeric variable author
> author[1:4]
[1] 123  NA  NA  NA
> # 4. Search for R objects called "author"
> find("author")
[1] ".GlobalEnv" "data1"
> # 5. Detach data frame data1
> detach(data1)
> # 6. Remove R object author from global environment
> rm(author)
```

The `find` command reveals that an R object `author` is present both in the workspace (called `.GlobalEnv` in the search path) and in R object `data1`.[4] As the workspace is searched first, this R object is used in the print command.

Accordingly, in this book we use one of the first three methods to select variables from a dataset. The dollar assignment is preferred if we are interested in the information of a single variable. The square brackets can be used to select several variables from a dataset. For example, the following R code can be used to print year of publication as well as sample sizes in addition to the author name.

```
> data1[1:4, c("author", "year", "Ne", "Nc")]
  author year Ne Nc
1  Boner 1988 13 13
2  Boner 1989 20 20
```

---

[4]The complete search path can be printed using the command `search()`.

```
3 Chudry 1987 12 12
4  Comis 1993 12 12
```

On the other hand, function `with` is typically more concise in calculations based on several variables of a dataset. We demonstrate this in an example with only two variables.

```
> # 1. Calculate and display the total sample sizes:
> data1$Ne + data1$Nc
 [1] 26 40 24 24 34 16 26 24 24 24 32 48 38 40 18 16 26
> with(data1, Ne + Nc)
 [1] 26 40 24 24 34 16 26 24 24 24 32 48 38 40 18 16 26
> # 2. Calculate and display the total sample size
> #    for the Chudry study
> data1$Ne[data1$author=="Chudry"] +
+   data1$Nc[data1$author=="Chudry"]
[1] 24
> with(data1[data1$author=="Chudry",], Ne + Nc)
[1] 24
```

1.4 Running Scripts

It is very often useful to generate a file, or script, of R commands both as a record of what has been done and for subsequent execution. An R script is simply a text file of commands, which can be created and edited either inside or outside R. If it is created outside R, then commands need to be copied and pasted into the R console window. If it is created inside R, we can highlight regions of the script and execute them alone.

We now illustrate how to create and execute a simple source file. From the toolbar click on *File → New Script* (under Windows) or *File → New Document* (under Mac OS). A new script file will appear. In the script file enter the following commands:

```
getwd()
dir(pattern="example1")
data1 <- read.csv("dataset01.csv")
summary(data1)
print(summary(data1))
```

Save the script file. Although it is a text file, it will be given the suffix ".R". To execute the first line, move the cursor to this line, hold down the `Ctrl` key and press R (under Windows) or `cmd` key and <RETURN> on a Mac. To execute part of the code or the whole script, you first have to highlight the corresponding region.

Alternatively, if the file is saved in the current working directory as `example1.R` in the R console, you can enter `source("example1.R")` to execute all commands in the script file. Note that this will only display the output from commands explicitly printing R objects but that output generated by other

commands as they execute is not shown. Accordingly, the summary(data1) command will not print any output, however, the print command in the last line of the script will print the summary of dataset data1.

The analyses presented in each of the chapters in this book are available as R script files from the website http://meta-analysis-with-r.org/. These are best executed step by step.

1.5   Installing and Using Libraries of Additional Functions

The final step before we can perform a meta-analysis is to install the R library of meta-analysis functions. For this, your computer needs to be connected to the internet, and R running. At the R command prompt either type install.packages ("meta") or install.packages(). In the later case, a dialogue box will open, showing the available R packages in alphabetical order. Scroll down to **meta** and click on OK. You may be prompted to choose the nearest R repository; the package should then install without further user input. To find out about the package type help(package=meta). To make the functions in the package available in your R session type library(meta).

It is worth noting that **meta** [5, 6] is not the only R package with functions for meta-analysis. Other R packages with general methods for meta-analysis are **metafor** [10] and **rmeta** [3], for example. In particular, **metafor** is a suggested R package for **meta** as certain methods, e.g. restricted maximum-likelihood estimator for the between-study variance and meta-regression methods, are only available if **metafor** has been installed. Accordingly, it is recommended to also install R package **metafor**.[5]

See Appendix A.3 for information on R packages not explicitly used in the book.

1.6   A First Meta-Analysis with R

Spooner et al. [7] report a meta-analysis comparing Nedocromil sodium (experimental treatment) with placebo (control) for preventing exercise-induced bronchoconstriction. The response is the maximum fall in the forced expiratory volume in 1 second (FEV_1) over the course of follow-up, expressed as a percentage. For each study, mean value, standard deviation and sample size are reported for both experimental and control group. The mean difference is used as effect measure, i.e. mean value in Nedocromil sodium group minus mean value in placebo group.

---

[5]R command to install R package **metafor**: install.packages("metafor").

```
> # 1. Load add-on package meta
> library(meta)
Loading 'meta' package (version 4.0-2).
> # 2. Do meta-analysis
> m <- metacont(Ne, Me, Se, Nc, Mc, Sc,
+               studlab=paste(author, year),
+               data=data1)
> # 3. Produce forest plot
> forest(m, xlab="Maximum % fall in FEV1")
```

Fig. 1.3 Code for producing forest plot for the bronchoconstriction meta-analysis [7]. The response is the maximum fall in FEV_1 over the course of follow-up, expressed as a percentage
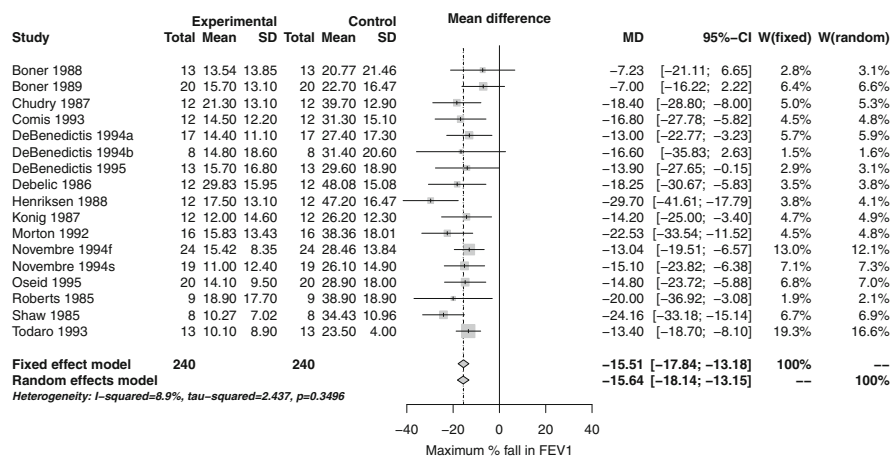
Fig. 1.4 Forest plot for the bronchoconstriction meta-analysis [7]. For details, see text

We described how to read in the data in Sect. 1.2. The code in Fig. 1.3 carries out a meta-analysis for these data with continuous response using the `metacont` function, and the `forest` command produces the plot shown in Fig. 1.4.

This figure shows, for each study, the estimated mean difference and its 95 % confidence interval in the middle of the plot. The area of the square centred on the estimated mean difference is proportional to the inverse of the variance of the study estimate resulting in a larger square for studies with more precise results, i.e. smaller variances. Note the use of the `xlab` option to label the x-axis. For each study—labelled by first author and date—we see the data which has been used in the calculations on the left side of the forest plot. On the right side of the forest plot, mean difference (MD) and its 95 % confidence interval are depicted. By default, results for both fixed effect and random effects meta-analysis are given. The columns labelled `W(fixed)` and `W(random)` reflect the percentage weight given to a study in the respective meta-analysis. Results of fixed effect or random effects model could be easily removed from the forest plot by using the arguments `comb.fixed=FALSE` or `comb.random=FALSE` in the `forest`

function, respectively. In Chap. 2 we discuss the estimation of the overall effect and its confidence interval, and explain how to control the output of the forest function.

To conclude this chapter, we make four points about the code in Fig. 1.3, which are relevant to the rest of the book:

1. We began by typing library(meta), thus making the functions in the R package **meta** available. Unless explicitly stated, this should be done before all the analyses presented in the remainder of the book.
2. Data from the meta-analysis is given to the metacont function through the *arguments* of the function which can be extracted by using the args function:

```
> args(metacont)
function (n.e, mean.e, sd.e, n.c, mean.c, sd.c, studlab,
    data = NULL, subset = NULL, sm = .settings$smcont,
    pooledvar = .settings$pooledvar, level = .settings$level,
*** Output truncated ***
```

   The key arguments of the metacont function, which are used to provide the function with the minimum data it needs, are the first six arguments (n.e, mean.e, sd.e, n.c, mean.c, sd.c) with

   - n.e, n.c
     the number of patients in the experimental treatment group and control group;
   - mean.e, sd.e
     the mean and standard deviation of the response in the experimental group, and
   - mean.c, sd.c
     the mean and standard deviation of the response in the control group.

   When calling the metacont function we are matching up the first argument n.e of the metacont function with the column Ne of the dataset data1; and similarly for the other arguments.
3. Notice that as we want to label the studies by their first author and year, we have to give this information to the metacont function. The paste function joins strings together, separated by a space. Thus studlab=paste(author, year) gives the author and year separated by a space. Other arguments of the metacont function will be explained later in the text.
4. The output of the metacont function is given to a new *R object*, called m. This R object is not simply a variable. Instead, it has a specific structure, consisting of a number of objects, together with additional information. The additional information means that the object can be used for other purposes with minimum additional input from the user. Thus the command forest(m) recognised that object m has been created using data from a meta-analysis; it

also correctly identifies the variables in the object `m` needed to produce a forest plot. Accordingly, we say that the `forest` function is a *generic function* which considers the class of an R object during execution.

This illustrates that R is an *object orientated language*. Another example is that we can use the same command, such as `summary(m)`, for any R object `m` resulting in different output depending on the *class* of the R object. R notes the class of the object we wish to summarise (e.g. meta-analysis, numeric vector, matrix) and returns an appropriate summary. For example, the following commands calculate a summary for the numbers from 1 to 10:

```
> # Class of meta-analysis object m
> class(m)
[1] "metacont" "meta"
> # Assign a numeric vector to object m
> m <- 1:10
> # Class of numeric vector m
> class(m)
[1] "integer"
> # Summary of vector with integers
> summary(m)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.25    5.50    5.50    7.75   10.00
```

Note, the command `m <- 1:10` overwrites the existing meta-analysis object `m` generated in Fig. 1.3.

1.7   Summary

In this chapter we have sought to introduce new R users to the basics of R necessary to get the most out of the remainder of this book. Hopefully both the elegance of R, illustrated by the relatively few commands needed to produce Fig. 1.4, and the quality of the output have whetted your appetite to read on.

References

1. Dalgaard, P.: Introductory Statistics with R, 2nd edn. Springer, New York (2008)
2. 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)
3. Lumley, T.: rmeta: Meta-Analysis. R package version 2.16. http://CRAN.R-project.org/package=rmeta (2012)
4. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org (2014)
5. 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

6. Schwarzer, G.: meta: Meta-Analysis with R. R package version 4.0-2 http://cran.R-project. org/package=meta (2014)
7. 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
8. StataCorp.: Stata Statistical Software: Release 13. StataCorp LP, College Station (2013)
9. The Cochrane Collaboration: Review Manager (RevMan) [Computer program]. Version 5.3. The Nordic Cochrane Centre, Copenhagen (2014)
10. Viechtbauer, W.: Conducting meta-analyses in R with the metafor package. J. Stat. Softw. **36**(3), 1–48 (2010)

Part II
Standard Methods

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]

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

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)]

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

---

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

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

with standard error $\text{S.E.}\,(\hat{g}_k) = \sqrt{\widehat{\text{Var}}\,(\hat{g}_k)}$ 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

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

and thus

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

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

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

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

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

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

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

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 Total	Mean	SD	Control Total	Mean	SD	MD	95%–CI	W(fixed)
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

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, 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

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

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

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 $\hat{\theta}_R$:

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

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

Hartung [13] showed that

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

with standard error $\mathrm{S.E.}_{\mathrm{HK}}(\hat{\theta}_R) = \sqrt{\widehat{\mathrm{Var}}_{\mathrm{HK}}(\hat{\theta}_R)}$ 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

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

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

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

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											

−40 −20   0   20   40

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)
```
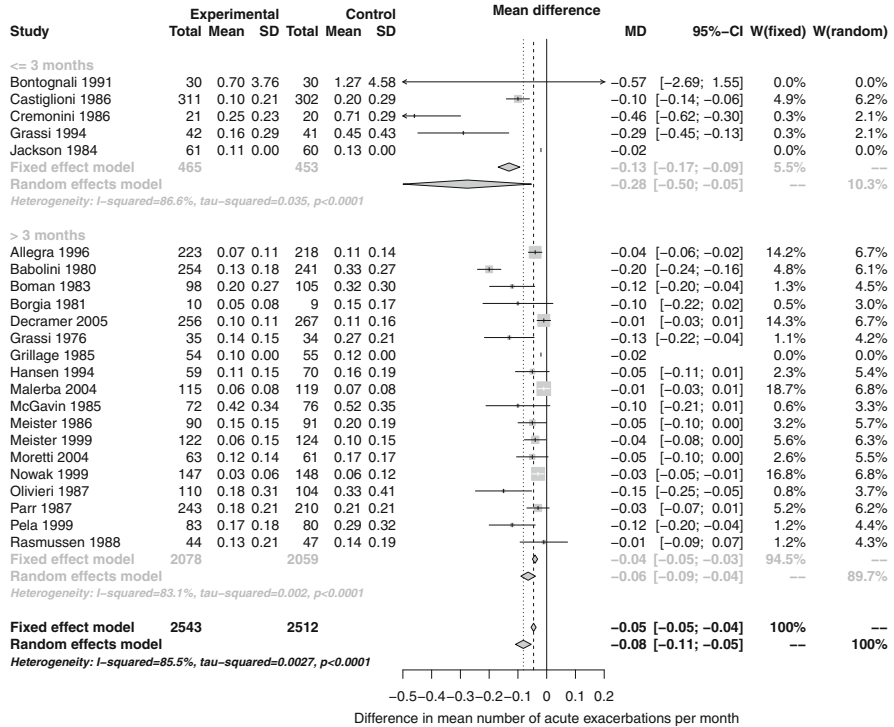
Study	Experimental Total	Mean	SD	Control Total	Mean	SD	Mean difference	MD	95%–CI	W(fixed)	W(random)
<= 3 months											
Bontognali 1991	30	0.70	3.76	30	1.27	4.58		−0.57	[−2.69; 1.55]	0.0%	0.0%
Castiglioni 1986	311	0.10	0.21	302	0.20	0.29		−0.10	[−0.14; −0.06]	4.9%	6.2%
Cremonini 1986	21	0.25	0.23	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)

Chapter 3
Meta-Analysis with Binary Outcomes

This chapter describes how to perform meta-analysis with binary data using R. We introduce the usual effect measures for binary outcomes and discuss issues raised by sparse binary data. We describe how to perform meta-analysis using the inverse variance method [17] and the DerSimonian–Laird method [12]. Furthermore, we introduce the Mantel–Haenszel method [24] and the Peto method [36] which are specific to binary outcomes. Several examples use base R commands. We also describe the `metabin` function from R package **meta** [31, 32] which provides a unified syntax for all methods in this chapter.

3.1 Effect Measures for Binary Outcomes

The most commonly used effect measures for binary outcomes are the odds ratio, risk ratio and risk difference. In addition, we describe another effect measure, the arcsine difference, which is used in tests for small-study effects in Chap. 5.

In the context of systematic reviews with binary outcomes, a discussion of the pros and cons of the various choices of effect measure, together with some recommendations, is given by Deeks [10] and Deeks and Altman [11]. In practice, the odds ratio and risk ratio are typically used. The main reason is that these relative effect measures are on average more stable across studies than the risk difference [10, 15], especially if the individual studies include patients with different follow-up times.

As before we consider a meta-analysis of K randomised controlled trials, but now each study has binary outcome data as shown in Table 3.1. Given the number of patients in the two groups, $n_{ek} = a_k + b_k$ and $n_{ck} = c_k + d_k$, we assume the number of events in each group follows a binomial distribution [1, p. 39]. Specifically, cell count $a_k \sim \text{Binomial}(n_{ek}, p_{ek})$ and cell count $c_k \sim \text{Binomial}(n_{ck}, p_{ck})$, where p_{ek} and p_{ck} denote the probability of the event in the experimental (i.e. intervention) and

Table 3.1 Summary data of study k in meta-analysis with binary responses $(k = 1, \ldots, K)$

	Event	No event	Group size
Experimental	a_k	b_k	$n_{ek} = a_k + b_k$
Control	c_k	d_k	$n_{ck} = c_k + d_k$
	$a_k + c_k$	$b_k + d_k$	n_k

Table 3.2 Variable names in R datasets for meta-analyses of binary responses; same notation used as in Table 3.1

Variable name	Notation	Description
study		Unique study label consisting of first author of study and year of publication (if necessary)
Ee	a_k	Number of events in the experimental (i.e. active) treatment arm
Ne	n_{ek}	Number of patients in the experimental treatment arm
Ec	c_k	Number of events in the control arm
Nc	n_{ck}	Number of patients in the control arm

control group, respectively. These probabilities are estimated from the observed cell counts by $\hat{p}_{ek} = a_k/(a_k + b_k) = a_k/n_{ek}$ and $\hat{p}_{ck} = c_k/(c_k + d_k) = c_k/n_{ck}$.

Example 3.1 Greb et al. [20] conducted a Cochrane review to assess the effects of high-dose chemotherapy with autologous stem cell transplantation as part of first-line treatment of adult patients with aggressive non-Hodgkin lymphoma. The primary outcome was survival time; complete response, a binary outcome, was one of several secondary outcomes. The Cochrane review used the risk ratio as measure of treatment effect for binary outcomes and the fixed effect model for pooling.

For datasets with binary data, we always use the same variable names which are described in Table 3.2. R code to import the complete response data using the read.csv function and to print the dataset is shown in Fig. 3.1.

We see that the meta-analysis contains 14 trials with sample sizes ranging between 48 (Intragumtornchai) and 370 (Gisselbrecht). Before moving on, we note that in these data the number of events Ee and Ec, which correspond to cell counts (a_k, c_k) in Table 3.1, are all greater than zero—as are the number of "non-events" Ne-Ee and Nc-Ec, which correspond to cell counts (b_k, d_k) in Table 3.1.

<div style="text-align: right">□</div>

```
> # 1. Read in the data
> data7 <- read.csv("dataset07.csv")
> # 2. Display data
> data7
                 study  Ee  Ne  Ec  Nc
1              De Souza  14  28  10  26
2                Gianni  46  48  35  50
3          Gisselbrecht 119 189 116 181
4     Intragumtornchai  10  23   9  25
5                Kaiser 110 158  97 154
6        Kluin-Nelemans  67  98  56  96
7          Martelli 1996  3  22   4  27
8          Martelli 2003 57  75  51  75
9               Milpied  74  98  56  99
10       Rodriguez 2003  39  55  30  53
11         Santini 1998  46  63  34  61
12             Santini-2 80 117  71 106
13              Verdonck  25  38  26  35
14                Vitolo  35  60  46  66
> # 3. Calculate experimental and control event probabilities
> summary(data7$Ee/data7$Ne)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1364  0.5949  0.6837  0.6370  0.7249  0.9583
> summary(data7$Ec/data7$Nc)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1481  0.5594  0.6066  0.5661  0.6775  0.7429
```

Fig. 3.1 Data from meta-analysis on high-dose chemotherapy [20]; for details of the table headers, see Table 3.2

3.1.1 Odds Ratio

The odds ratio for study k, ψ_k, is defined as the ratio of the odds of an event in the experimental arm to that in the control arm. That is

$$\psi_k = \frac{\left(\dfrac{p_{ek}}{1 - p_{ek}}\right)}{\left(\dfrac{p_{ck}}{1 - p_{ck}}\right)} = \frac{p_{ek}\,(1 - p_{ck})}{p_{ck}\,(1 - p_{ek})} \ . \tag{3.1}$$

If either of the two estimated event probabilities is zero the log odds ratio, $\log \psi_k$, is either $-\infty$ or $+\infty$. If both are zero, the log odds ratio is undefined.

Based on the data given in Table 3.1, the odds ratio from study k is estimated by

$$\hat{\psi}_k = \frac{a_k\,d_k}{b_k\,c_k} \ . \tag{3.2}$$

As this estimator has a skewed distribution in typical sample sizes, effect estimates, standard errors and confidence intervals are usually calculated using the natural logarithm of $\hat{\psi}_k$ [1, p. 71], which we write $\log(\hat{\psi}_k)$. The final results are then back-transformed to the original scale for presentation. For typical sample sizes, the

variance of the natural logarithm of the odds ratio is well approximated by

$$\widehat{\mathrm{Var}}\left(\log \hat{\psi}_k\right) = \frac{1}{a_k} + \frac{1}{b_k} + \frac{1}{c_k} + \frac{1}{d_k} \,, \tag{3.3}$$

where the approximation improves as n_{ek} and n_{ck} increase.

Using the estimate of the log odds ratio, and its estimated variance, an approximate two-sided $(1 - \alpha)$ confidence interval for the odds ratio is given by

$$\exp\left(\log \hat{\psi}_k \ \pm \ z_{1-\frac{\alpha}{2}} \ \mathrm{S.E.}\,(\log \hat{\psi}_k)\right), \tag{3.4}$$

where the standard error $\mathrm{S.E.}\,(\log \hat{\psi}_k) = \sqrt{\widehat{\mathrm{Var}}\,(\log \hat{\psi}_k)}$ and $z_{1-\frac{\alpha}{2}}$ denote the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution [1, p. 71]. Further, as illustrated in Sect. 3.3.1, the estimates (3.2) and (3.3) can be used in a meta-analysis using the inverse variance method for pooling.

Example 3.2 To illustrate the estimation of the odds ratio, we use data from the ninth trial in Fig. 3.1, i.e. the Milpied trial. The following base R code calculates the odds ratio and its approximate 95 % confidence interval for this trial.

```
> # 1. Calculate log odds ratio and its standard error for
> #    Milpied trial
> logOR <- with(data7[data7$study=="Milpied",],
+               log((Ee*(Nc-Ec)) / (Ec*(Ne-Ee))))
> selogOR <- with(data7[data7$study=="Milpied",],
+                 sqrt(1/Ee + 1/(Ne-Ee) + 1/Ec + 1/(Nc-Ec)))
> # 2. Print odds ratio and limits of 95% confidence interval
> round(exp(c(logOR,
+             logOR + c(-1,1) *
+             qnorm(1-0.05/2) * selogOR)), 4)
[1] 2.3676 1.2887 4.3495
```

In order to access the necessary data, i.e. (Ne, Ee, Nc, Ec) from the Milpied trial, it is simplest to use the with function (see Sect. 1.3). The two assignments calculate the log odds ratio (3.2) and its standard error (3.3). Then we use the exp function to back-transform the logarithmised values and the round function to print result with four decimal places. The three values returned are the odds ratio and the lower and upper limit of the 95 % confidence interval.

We obtain the same result using the metabin function with arguments sm="OR" (i.e. summary measure is the Odds Ratio) and method="I" (i.e. Inverse variance method).

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+         data=data7, subset=study=="Milpied")
    OR           95%-CI      z  p-value
 2.3676 [1.2887; 4.3495] 2.7774   0.0055

Details:
- Inverse variance method
```

The software reports the *z*-score and *p*-value for the test of an overall treatment effect in addition to the odds ratio and its 95 % confidence interval.

Note, the `metabin` function requires the number of events (`Ee`, `Ec`) and the sample sizes (`Ne`, `Nc`) as input as this information is typically reported in publications for binary outcomes. Likewise, RevMan 5 [35] requires this information as input. By contrast, the Stata function `metan` [33] requires the number of events (`Ee`, `Ec`) and non-events (`Ne-Ee`, `Nc-Ec`) as input. This subtle difference is no restriction as any number can be calculated from the other two numbers. □

3.1.2 Risk Ratio

The risk ratio ϕ_k, often called relative risk, is defined as the ratio of the two event probabilities,

$$\phi_k = \frac{p_{ek}}{p_{ck}} \ .$$
(3.5)

If either of the two event probabilities is zero the log risk ratio $\log \phi_k$, is either $-\infty$ or $+\infty$. If both probabilities are zero, the log risk ratio is undefined.

The risk ratio is estimated by

$$\hat{\phi}_k = \frac{\left(\dfrac{a_k}{a_k + b_k}\right)}{\left(\dfrac{c_k}{c_k + d_k}\right)},$$
(3.6)

whose variance is approximated by

$$\widehat{\mathrm{Var}}\,(\log \hat{\phi}_k) = \frac{1}{a_k} + \frac{1}{c_k} - \frac{1}{a_k + b_k} - \frac{1}{c_k + d_k} \ ,$$
(3.7)

where—as with the log odds ratio—the approximation works well for typical trials and improves as the number of patients, n_{ek} and n_{ck}, increases. A confidence interval for $\hat{\phi}_k$ can be constructed by replacing $\hat{\psi}_k$ with $\hat{\phi}_k$ in Eq. (3.4).

Example 3.3 Continuing with the previous example, we first use base R code for calculating the risk ratio and its confidence interval for the `Milpied` trial.

```
> # 1. Calculate log risk ratio and its standard error for
> #    Milpied study
> logRR <- with(data7[data7$study=="Milpied",],
+               log((Ee/Ne) / (Ec/Nc)))
> selogRR <- with(data7[data7$study=="Milpied",],
+               sqrt(1/Ee + 1/Ec - 1/Ne - 1/Nc))
> # 2. Print risk ratio and limits of 95% confidence interval
```

```
> round(exp(c(logRR,
+                 logRR + c(-1,1) *
+                 qnorm(1-0.05/2) * selogRR)), 4)
[1] 1.3349 1.0862 1.6406
```

As before, a simpler alternative is to use the `metabin` function with argument `sm="RR"` (Risk Ratio).

```
> metabin(Ee, Ne, Ec, Nc, sm="RR", method="I",
+           data=data7, subset=study=="Milpied")
      RR           95%-CI        z p-value
 1.3349 [1.0862; 1.6406] 2.7461   0.006

Details:
- Inverse variance method
```

Again, in addition to the risk ratio and confidence interval, the `metabin` function reports z-score and p-value for the test of an overall treatment effect. Note, the R code for the risk ratio differs from the R code for the odds ratio only by setting parameter `sm="RR"` instead of `sm="OR"`. Actually, we do not have to use argument `sm="RR"` as this is the default effect measure in the `metabin` function.

In the `Milpied` trial, the odds ratio is almost twice as large as the risk ratio. This is because the probability of an event is large in both the experimental and control arms (0.755 and 0.566, respectively), so the risk ratio is not well approximated by the odds ratio. Nevertheless, the overall conclusions about the effectiveness of high-dose chemotherapy with autologous stem cell transplantation as part of first-line treatment in adult patients with aggressive non-Hodgkin lymphoma are very similar: the p-value of the test for treatment difference is 0.0055 for the odds ratio and 0.006 for the risk ratio.   □

3.1.3 Risk Difference

The risk difference η_k is defined as the difference between the two event probabilities

$$\eta_k = p_{ek} - p_{ck} . \tag{3.8}$$

The risk difference is always defined and has finite range from -1 to 1.

The natural estimate of the risk difference is

$$\hat{\eta}_k = \frac{a_k}{a_k + b_k} - \frac{c_k}{c_k + d_k}, \tag{3.9}$$

with corresponding variance estimate

$$\widehat{\text{Var}}(\hat{\eta}_k) = \frac{a_k b_k}{(a_k + b_k)^3} + \frac{c_k d_k}{(c_k + d_k)^3} . \tag{3.10}$$

An approximate two-sided $(1 - \alpha)$ confidence interval for the risk difference is thus

$$\hat{\eta}_k \ \pm \ z_{1-\frac{\alpha}{2}} \ \text{S.E.} \, (\hat{\eta}_k) \tag{3.11}$$

with standard error $\text{S.E.} \, (\hat{\eta}_k) = \sqrt{\widehat{\text{Var}} \, (\hat{\eta}_k)}$ and $z_{1-\frac{\alpha}{2}}$ denoting the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution. Unfortunately, for when the risk difference is close to -1 or 1 this confidence interval will sometimes include values below -1 or above 1, respectively.

Example 3.4 Continuing with the `Milpied` trial, we could use base R code to calculate a confidence interval. However, it is more straightforward to use the `metabin` function with argument `sm="RD"` (Risk Difference).

```
> metabin(Ee, Ne, Ec, Nc, sm="RD", method="I",
+         data=data7, subset=study=="Milpied")
    RD          95%-CI      z  p-value
 0.1894 [0.0599; 0.319] 2.8662   0.0042

Details:
- Inverse variance method
```

Again, the only change in the R code is to give a different value, `RD`, to function argument `sm`. Using the risk difference as the effect measure, we obtain a similar *p*-value to those obtained using the odds ratio and risk ratio.   □

3.1.4   Arcsine Difference

The arcsine difference has a long history, dating back to the 1940s [3, 4, 16, 18, 25], and is often used in other contexts [2, 9, 23]. It has been considered as a measure of effectiveness in clinical trials [13, 15, 30], though it is rarely used in practice. However, it has certain advantages when assessing whether a meta-analysis may be affected by publication bias or other small-study effects. We therefore introduce it here.

The arcsine difference is defined as the difference of the arcsine-transformed event probabilities, that is

$$\Delta_k = \arcsin \sqrt{p_{ek}} - \arcsin \sqrt{p_{ck}} \ . \tag{3.12}$$

Like the risk difference, the arcsine difference is always defined, with a finite range from $-\pi/2$ to $\pi/2$. The value of the arcsine difference is similar to the risk difference if both event probabilities are close to 0.5, otherwise the values can be quite different.

The arcsine difference is estimated by

$$\hat{\Delta}_k = \arcsin\sqrt{\frac{a_k}{n_{ek}}} - \arcsin\sqrt{\frac{c_k}{n_{ck}}}, \tag{3.13}$$

with approximate variance

$$\widehat{\mathrm{Var}}\left(\hat{\Delta}_k\right) = \frac{1}{4n_{ek}} + \frac{1}{4n_{ck}} \approx \frac{1}{n_k}\,, \text{ if } n_{ek} \approx n_{ck}, \tag{3.14}$$

where the approximation improves as n_{ek} and n_{ck} increase. Notice that the approximate variance of $\hat{\Delta}_k$ only depends on the sample size in the two groups. A confidence interval for $\hat{\Delta}_k$ can be constructed by replacing $\hat{\eta}_k$ with $\hat{\Delta}_k$ in formula (3.11).

Example 3.5 Continuing with the previous example, estimation of the arcsine difference is straightforward either using base R code.

```
> # 1. Calculate arcsine difference and its standard error for
> #    Milpied study
> ASD <- with(data7[data7$study=="Milpied",],
+             asin(sqrt(Ee/Ne)) - asin(sqrt(Ec/Nc)))
> seASD <- with(data7[data7$study=="Milpied",],
+              sqrt(1/(4*Ne) + 1/(4*Nc)))
> # 2. Print arcsine difference and its 95% confidence interval
> round(c(ASD,
+         ASD + c(-1,1) *
+         qnorm(1-0.05/2) * seASD), 4)
[1] 0.2019 0.0622 0.3415
```

or using the `metabin` function argument with `sm="ASD"` (Arc Sine Difference).

```
> metabin(Ee, Ne, Ec, Nc, sm="ASD", method="I",
+         data=data7, subset=study=="Milpied")
    ASD           95%-CI       z   p-value
 0.2019 [0.0622; 0.3415] 2.8333    0.0046

Details:
- Inverse variance method
```

In the `Milpied` trial, both the arcsine difference and the *p*-value are very close to their respective values for the risk difference. In summary, the overall conclusion concerning the effectiveness of high-dose chemotherapy is insensitive to the choice of metric.   □

3.2 Estimation in Sparse Data

We now consider the issues raised when the number of events in one or both of the study arms is small. In such cases, we have already noted that estimates of the odds ratio and risk ratio may be undefined. In the literature, a two-by-two table is referred to as sparse if any of the cell counts is small [1, p. 391]. Sparse data are likely to occur if either the total sample size of a study is small, or if the sample size is large but the probability of an event is very close to zero or one. In the context of systematic reviews of randomised controlled trials, both cases are likely.

Example 3.6 Quan et al. [27] conducted a Cochrane review to evaluate whether the benefit of treating hypertension in women differed between younger and older women, as well as between white and African-American women. In the systematic review, the Peto method, i.e. a fixed effect estimate using the odds ratio as effect measure, was used for pooling (see Sect. 3.3.3). The primary outcome was the occurrence of fatal cerebrovascular events—a rare event in hypertension. Here, we only look at the subgroup of women older than 55 years. These data are shown in Fig. 3.2.

The meta-analysis contains 11 trials having large sample sizes ranging between 349 (`Shep Pilot`) and 3710 (`MRC`). However, experimental and control event probabilities are small in some trials (`STOP`, `Shep Pilot`), and in one trial

```
> # 1. Read in the data
> data8 <- read.csv("dataset08.csv")
> # 2. Print dataset
> data8
         study Ee   Ne Ec   Nc
1   Australian  0  300  1  295
2       CASTEL  7  232  8  192
3        Coope  3  297  6  314
4        EWPHE 17  287 21  299
5         HDFP 11  984 19  938
6          MRC  6 1858  5 1852
7   MRC elderly 18 1273 17 1287
8         STOP  1  510 10  509
9         Shep  8 1331  9 1359
10  Shep Pilot  2  279  1   70
11    Syst-eur 12 1618 12 1520
> # 3. Calculate experimental and control event rates
> summary(data8$Ee/data8$Ne)
    Min. 1st Qu.   Median     Mean 3rd Qu.     Max.
0.000000 0.004620 0.007417 0.013690 0.012660 0.059230
> summary(data8$Ec/data8$Nc)
    Min. 1st Qu.   Median     Mean 3rd Qu.     Max.
0.002700 0.007259 0.014290 0.019910 0.019950 0.070230
```

Fig. 3.2 Data from meta-analysis on hypertension in women [27]; see Table 3.2 for details on table headers

(`Australian`), no events are observed in the experimental group. These three
studies are typical examples of what is usually referred to as sparse binary data. □

Suppose study k has sparse data. As already noted, if either a_k or c_k is zero, both
the estimated odds ratio and estimated risk ratio are either 0 or ∞. If both cell counts
are zero, the two summary measures are undefined. In both cases, variance estimates
given in (3.3) and (3.7) are ∞ due to division by zero. For the odds ratio, the same
holds true if either cell count b_k or d_k is zero; while the risk ratio has finite estimates
in this case, the variance formula is unreliable.

In order to proceed with the meta-analysis, we either have to omit such studies,
or add a small increment to each cell of two-by-two tables with zero entries. The
latter is known as adding a *continuity correction*. Gart and Zweifel [19] showed
that, if any of the cell counts are zero (but the underlying true probabilities are not
zero or 1), adding 0.5 to the cell counts (a_k, b_k, c_k, d_k) in Table 3.1 improves the
estimators (3.2) and (3.3) by reducing their bias. We obtain

$$\hat{\psi}_k^{\text{mod}} = \frac{(a_k + 0.5)(d_k + 0.5)}{(b_k + 0.5)(c_k + 0.5)} \tag{3.15}$$

and

$$\widehat{\text{Var}}\,(\log \hat{\psi}_k^{\text{mod}}) = \frac{1}{a_k + 0.5} + \frac{1}{b_k + 0.5} + \frac{1}{c_k + 0.5} + \frac{1}{d_k + 0.5}\,. \tag{3.16}$$

While this modification can be used quite generally, in the literature it is usually
only used if any of the cell counts is zero. The estimated variance of the log odds
ratio is reduced if 0.5 is added to each cell of the two-by-two table since all four
denominators in (3.3) are increased. Obviously, a different increment than 0.5 could
be used, for example 0.01 or 0.1 to evaluate how sensitive results are to different
choices of increments.

Sweeting et al. [34] proposed an alternative continuity correction, so called
treatment arm continuity correction, based on the number of observations in the two
groups. The continuity correction in experimental (incr_e) and control group (incr_c)
are defined as

$$\text{incr}_e = \frac{n_e}{n_e + n_c}$$

and

$$\text{incr}_c = \frac{n_c}{n_e + n_c}.$$

The two increments incr_e and incr_e add up to 1. If both groups have equal sample
size the increment is 0.5 in both groups. The idea of this continuity correction is to
get less biased results if groups are severely unbalanced with respect to sample size.

Example 3.7 The odds ratio and 95 % confidence interval for the `Australian` study can be calculated using base R code by manually adding 0.5 to each count.

```
> # 1. Calculate log odds ratio and its standard error for
> #    Australian study (with continuity correction)
> logOR <- with(data8[data8$study=="Australian",],
+               log(((Ee+0.5)*(Nc-Ec+0.5)) /
+                    ((Ec+0.5)*(Ne-Ee+0.5))))
> selogOR <- with(data8[data8$study=="Australian",],
+                  sqrt(1/(Ee+0.5) + 1/(Ne-Ee+0.5) +
+                       1/(Ec+0.5) + 1/(Nc-Ec+0.5)))
> # 2. Print odds ratio and limits of 95% confidence interval
> round(exp(c(logOR,
+             logOR + c(-1,1) * qnorm(1-0.05/2) * selogOR)), 4)
[1] 0.3267 0.0133 8.0515
```

or—less tediously—using the `metabin` function.

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+         data=data8, subset=study=="Australian")
     OR           95%-CI      z  p-value
 0.3267 [0.0133; 8.0515] -0.6842   0.4938

Details:
- Inverse variance method
- Continuity correction of 0.5 in studies with zero cell
  frequencies
```

We see that `metabin` prints the information that an increment of 0.5 has been added to the data of studies with zero cells, i.e. the `Australian` study. The default increment of 0.5 can be changed by setting a different value using argument `incr` in the `metabin` function. For example, adding 0.1 to the cells results in a somewhat different estimated odds ratio.

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+         data=data8, subset=study=="Australian",
+         incr=0.1)
     OR           95%-CI      z  p-value
 0.0891 [1e-04; 57.8269] -0.7319   0.4642

Details:
- Inverse variance method
- Continuity correction of 0.1 in studies with zero cell
  frequencies
```

However, confidence intervals are very wide and overlapping for the two different choices of the increment.

We can conduct an analysis based on the treatment arm continuity correction [34] using argument `incr="TACC"`.

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+         data=data8, subset=study=="Australian",
+         incr="TACC")
     OR           95%-CI      z  p-value
```

```
 0.3303 [0.0135; 8.0698] −0.6793   0.4969

Details:
- Inverse variance method
- Treatment arm continuity correction in studies with
  zero cell frequencies
```

In the `Australian` study—with similar sample sizes in the two groups, see Fig. 3.2—results are very similar to the default method using an increment of 0.5.

□

Similarly, when calculating the risk ratio with sparse data, Pettigrew et al. [26] showed that the estimators given by Eqs. (3.6) and (3.7) behave better when a continuity correction is added

$$\hat{\phi}_k^{\text{mod}} = \frac{a_k + 0.5}{a_k + b_k + 0.5} \bigg/ \frac{c_k + 0.5}{c_k + d_k + 0.5} \tag{3.17}$$

and

$$\widehat{\text{Var}}\,(\log \hat{\phi}_k^{\text{mod}}) = \frac{1}{a_k + 0.5} + \frac{1}{c_k + 0.5} - \frac{1}{a_k + b_k + 0.5} - \frac{1}{c_k + d_k + 0.5} \,. \tag{3.18}$$

Again, this modification can be used in general, but is typically only applied to a study if a cell count is zero.

Example 3.8 The risk ratio and 95% confidence interval for the `Australian` study can be easily calculated using the `metabin` function:

```
> metabin(Ee, Ne, Ec, Nc, sm="RR", method="I",
+         data=data8, subset=study=="Australian")
    RR           95%-CI        z p-value
 0.3278 [0.0134; 8.014] −0.6839   0.494

Details:
- Inverse variance method
- Continuity correction of 0.5 in studies with zero cell
  frequencies
```

As we have already noted, by default the `metabin` function only adds an increment of 0.5 in a study with a zero cell. You can use the argument `addincr=TRUE` to apply a continuity correction to a study regardless of zero cells.   □

3.2.1 Peto Odds Ratio

An alternative method for the estimation of the odds ratio, which we term the *Peto Odds Ratio* method, was proposed by Yusuf et al. [36]. This method is sometimes referred to as the Yusuf and Peto method, or just the Peto method. The latter name

is confusing, since the Peto Odds Ratio method is different from the Peto method for meta-analysis (see Sect. 3.3.3).

The key advantage of this method is that no correction for zero cell counts is necessary. The method is based on the observed cell count a_k and the expected cell count $E(a_k | \ldots ; \psi = 1)$ where "$\cdots$" denotes the four marginal totals in Table 3.1. The Peto estimate of the odds ratio is

$$\hat{\psi}_k^* = \exp\left(\frac{a_k - E(a_k | \cdots ; \psi_k = 1)}{\text{Var}(a_k | \cdots ; \psi_k = 1)} \right), \tag{3.19}$$

where $E(a_k | \cdots ; \psi_k = 1)$ and $\text{Var}(a_k | \ldots ; \psi_k = 1)$ are the mean and variance of a_k under the hypergeometric distribution. Under this distribution, we have

$$E(a_k | \cdots ; \psi_k = 1) = \frac{(a_k + b_k)(a_k + c_k)}{n_k}$$

and

$$\text{Var}(a_k | \cdots ; \psi_k = 1) = \frac{(a_k + b_k)(c_k + d_k)(a_k + c_k)(b_k + d_k)}{n_k^2(n_k - 1)} .$$

An estimator of the variance of $\log \hat{\psi}_k^*$ is

$$\widehat{\text{Var}}(\log \hat{\psi}_k^*) = \frac{1}{\text{Var}(a_k | \ldots ; \psi_k = 1)} . \tag{3.20}$$

Greenland and Salvan [22] showed that the Peto estimator performs poorly in unbalanced designs (i.e. when n_{ek} is very different from n_{ck}) and in nearly balanced designs when the true odds ratio differs substantially from 1. Brockhaus et al. [8] showed that the Peto estimator is not consistent (i.e. does not converge to the true odds ratio for large sample sizes) if the design is unbalanced. In the context of randomised controlled trials, the use of the Peto method is reasonable in most cases, i.e. if treatment groups are of comparable size (in the case of 1:1 randomisation) and treatment effects are moderate.

Once the log odds ratio and its standard error have been calculated, the meta-analysis proceeds in the usual way as described in Sects. 3.3 and 3.4 for the fixed effect and random effects model, respectively.

Example 3.9 Returning to the high-dose chemotherapy study, the Peto odds ratio for the `Milpied` trial is calculated using the `metabin` function with argument `method="P"`.

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="P",
+         data=data7, subset=study=="Milpied")
    OR           95%-CI      z  p-value
 2.316 [1.2863; 4.1698] 2.7992   0.0051

Details:
- Peto method
```

For the `Milpied` trial, the value of the Peto odds ratio, 2.316, is similar to the usual odds ratio estimate, 2.368, calculated in Example 3.2.   □

Example 3.10 For the `Australian` study, the Peto odds ratio can be calculated similarly.

```
> metabin(Ee, Ne, Ec, Nc, sm="OR", method="P",
+         data=data8, subset=study=="Australian")
    OR           95%-CI        z  p-value
 0.1331 [0.0026; 6.7068] -1.0084   0.3132

Details:
- Peto method
```

In this case, the Peto odds ratio, 0.1331 differs substantially from the usual odds ratio, 0.3267, calculated in Example 3.7 with a continuity correction of 0.5.   □

3.3 Fixed Effect Model

For the fixed effect model there are three approaches to estimate the pooled treatment effect with binary data: inverse variance, Mantel–Haenszel and Peto method. While the inverse variance method can be used for all effect measures, the Mantel–Haenszel method is only suitable for the odds ratio, risk ratio and risk difference and the Peto method is specific for the odds ratio. For a more general discussion of the fixed effect model, see Sect. 2.2.

3.3.1 Inverse Variance Method

As already noted, the inverse variance method is generic. To use it, we require only an estimate of the treatment effect and its variance from each study. We apply the inverse variance method on the scale where the distribution of the estimators is best approximated by the normal distribution, and then back-transform the results if necessary.

The pooled fixed effect estimate is therefore

$$\hat{\theta}_F = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{\sum_{k=1}^{K} w_k}, \text{ with } \text{Var}\left(\hat{\theta}_F\right) = \left(\sum_{k=1}^{K} w_k\right)^{-1}, \tag{3.21}$$

and weights $w_k = \widehat{\text{Var}}(\hat{\theta}_k)^{-1}$. As usual, an approximate two-sided $(1 - \alpha)$ confidence interval is given by

$$\hat{\theta}_F \pm z_{1-\frac{\alpha}{2}} \text{ S.E.}(\hat{\theta}_F), \tag{3.22}$$

with standard error S.E. $(\hat{\theta}_F) = \sqrt{\widehat{\text{Var}}(\hat{\theta}_F)}$ [17].

When working on a transformed scale as we do for the odds ratio and risk ratio, $\hat{\theta}_k$ is the transformed effect measure, w_k the reciprocal of the variance of the transformed effect measure and S.E. $(\hat{\theta}_F)$ the standard error of the transformed effect measure. Then, Eqs. (3.21) and (3.22) give the fixed effect estimate and the confidence interval on the transformed scale. Usually, these values are back-transformed to the original scale to report the results.

Depending on the metric, $\hat{\theta}_k$ and $\widehat{\text{Var}}(\hat{\theta}_k)$ are given by

- risk difference: (3.9) and (3.10);
- arcsine difference: (3.13) and (3.14);
- odds ratio: natural logarithm of (3.2) and (3.3);
- odds ratio (with continuity correction for sparse data): natural logarithm of (3.15) and (3.16);
- Peto odds ratio: natural logarithm of (3.19) and (3.20);
- risk ratio: natural logarithm of (3.6) and (3.7);
- risk ratio (with continuity correction for sparse data): natural logarithm of (3.17) and (3.18).

Unfortunately, for sparse data the pooled estimate based on the inverse variance method is biased [7, 21]. Preferable methods for estimating the pooled fixed effect estimate in these situations are described in Sects. 3.3.2 and 3.3.3.

Example 3.11 For the high-dose chemotherapy data, a meta-analysis based on the inverse variance method with odds ratio as effect measure can be carried out using base R code as follows.

```
> # 1. Calculate log odds ratio, variance and weights
> logOR <- with(data7,
+               log((Ee*(Nc-Ec)) / (Ec*(Ne-Ee))))
> varlogOR <- with(data7,
+               1/Ee + 1/(Ne-Ee) + 1/Ec + 1/(Nc-Ec))
> weight <- 1/varlogOR
> # 2. Calculate the inverse variance estimator
> round(exp(weighted.mean(logOR, weight)), 4)
[1] 1.3228
> # 3. Calculate the variance
> round(1/sum(weight), 4)
[1] 0.0089
```

These calculations can be done much easier using the `metabin` function.

```
> mb1 <- metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+                data=data7, studlab=study)
```
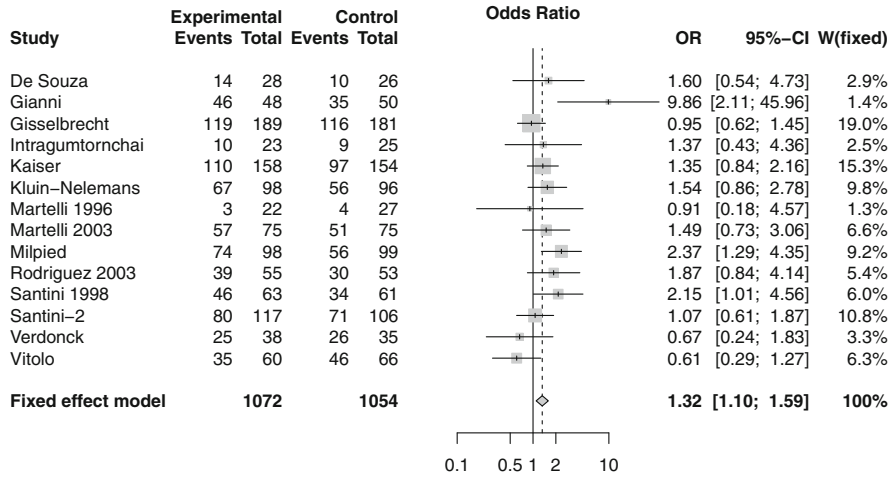
Study	Experimental Events Total		Control Events Total		Odds Ratio	OR	95%–CI	W(fixed)
De Souza	14	28	10	26		1.60	[0.54; 4.73]	2.9%
Gianni	46	48	35	50		9.86	[2.11; 45.96]	1.4%
Gisselbrecht	119	189	116	181		0.95	[0.62; 1.45]	19.0%
Intragumtornchai	10	23	9	25		1.37	[0.43; 4.36]	2.5%
Kaiser	110	158	97	154		1.35	[0.84; 2.16]	15.3%
Kluin–Nelemans	67	98	56	96		1.54	[0.86; 2.78]	9.8%
Martelli 1996	3	22	4	27		0.91	[0.18; 4.57]	1.3%
Martelli 2003	57	75	51	75		1.49	[0.73; 3.06]	6.6%
Milpied	74	98	56	99		2.37	[1.29; 4.35]	9.2%
Rodriguez 2003	39	55	30	53		1.87	[0.84; 4.14]	5.4%
Santini 1998	46	63	34	61		2.15	[1.01; 4.56]	6.0%
Santini–2	80	117	71	106		1.07	[0.61; 1.87]	10.8%
Verdonck	25	38	26	35		0.67	[0.24; 1.83]	3.3%
Vitolo	35	60	46	66		0.61	[0.29; 1.27]	6.3%
Fixed effect model		**1072**		**1054**		**1.32**	**[1.10; 1.59]**	**100%**

```
0.1   0.5 1  2    10
```

Fig. 3.3 Meta-analysis on high-dose chemotherapy [20]: forest plot using the inverse variance method using the `forest` function

```
> round(c(exp(mb1$TE.fixed), mb1$seTE.fixed^2), 4)
[1] 1.3228 0.0089
```

A summary for the meta-analysis is returned by the following R command.

```
> print(summary(mb1), digits=2)
Number of studies combined: k=14

                   OR       95%-CI    z  p-value
Fixed effect model 1.32 [1.10; 1.59] 2.97   0.003
*** Output truncated ***
```

The forest plot shown in Fig. 3.3 was produced by the R command

```
> forest(mb1, comb.random=FALSE, hetstat=FALSE).                  □
```

Example 3.12 For the hypertension in women dataset, the base R code used in the previous example to calculate the inverse variance estimate would give the Australian study zero weight in the calculation as one of the cells in the Australian study is zero.

```
> # 1. Calculate log odds ratio, variance and weights
> logOR <- with(data8,
+               log((Ee*(Nc-Ec)) / (Ec*(Ne-Ee))))
> varlogOR <- with(data8,
+               1/Ee + 1/(Ne-Ee) + 1/Ec + 1/(Nc-Ec))
> weight <- 1/varlogOR
> # 2. Weight for Australian study
> weight[data8$study=="Australian"]
[1] 0
```

□

Building on Sect. 3.2, we can use a continuity correction for meta-analysis with sparse data in one of four ways:

- add 0.5 (or any other small increment) to all two-by-two tables regardless of zero cells (referred to as "add all");
- add 0.5 (or any other small increment) to all two-by-two tables only in the case of zero cell counts in one or more studies (referred to as "add all conditional"),
- add 0.5 only to cell counts of corresponding two-by-two tables with zero cell counts (referred to as "add selective"), or
- treatment arm continuity correction based on the number of observations in the two groups.

All four approaches are available in the `metabin` function. By default, the "add selective" approach is used. The "add all" approach is applied using the argument `addincr=TRUE`; the "add all conditional" approach is applied using the argument `allincr=TRUE`. A meta-analysis based on the treatment arm continuity correction can be conducted using argument `incr="TACC"`.

Usually studies with zero events in both treatment groups are excluded from a meta-analysis with odds ratio or risk ratio as measure of treatment effect.

Example 3.13 While it would be possible to modify the base R code to allow for zero cells, it is much more convenient to conduct the hypertension in women meta-analysis using the `metabin` function.

```
> mb2 <- metabin(Ee, Ne, Ec, Nc, sm="OR", method="I",
+                data=data8, studlab=study)
> print(summary(mb2), digits=2)
Number of studies combined: k=11

                      OR       95%-CI      z  p-value
Fixed effect model  0.78 [0.58; 1.05] -1.63   0.1026
*** Output truncated ***
- Continuity correction of 0.5 in studies with zero cell
  frequencies
```

The last line tells us that a continuity correction has been applied to studies with zero cells ("add selective" approach). The continuity correction has only been applied to the `Australian` study as we can easily check.

```
> as.data.frame(mb2)[,c("studlab", "incr.e", "incr.c")]
        studlab incr.e incr.c
1    Australian    0.5    0.5
2        CASTEL    0.0    0.0
*** Output truncated ***
11     Syst-eur    0.0    0.0
```

This R command converts an object of class `meta` into a data frame. From this newly created data frame, we only print information on the study label (variable `studlab`) and the increments added to cells in the experimental and control group (variables `incr.e` and `incr.c`). □

3.3.2   Mantel–Haenszel Method

Mantel and Haenszel [24] proposed an estimator for the common odds ratio in a stratified case–control study, and this method can also be used in a meta-analysis of randomised controlled trials. The Mantel–Haenszel method was extended to the risk ratio and risk difference as measure of treatment effect by Greenland and Robins [21]. The Mantel–Haenszel method has been recommended as the "method of choice" for estimating a pooled odds ratio in most situations [14]. Thus, for binary outcome data, the Mantel–Haenszel method is the default procedure in the metabin function (argument method="MH") and in RevMan 5 [35].

For the Mantel–Haenszel method, there is no need to add 0.5 to each cell of two-by-two tables with zero cell counts. Nevertheless, this modification is utilised in commonly used software, e.g. in RevMan 5 [35] and in the Stata procedure metan [33]. Accordingly, this modification is also the default in the metabin function; the exact Mantel–Haenszel method is available by specifying the argument MH.exact=TRUE.

We now describe the Mantel–Haenszel method for pooling for the odds ratio, risk ratio and risk difference.

Odds Ratio

The pooled odds ratio is estimated by combining the individual odds ratios $\hat{\psi}_k$ on the natural scale

$$\hat{\psi}_{\mathrm{MH}} = \frac{\sum\limits_{k=1}^{K} w_k \hat{\psi}_k}{\sum\limits_{k=1}^{K} w_k} \tag{3.23}$$

with weights $w_k = \dfrac{b_k c_k}{n_k}$.

An estimator of the variance of the logarithm of $\hat{\psi}_{\mathrm{MH}}$ that is robust both in sparse data and large strata models was introduced by Robins et al. [29], see also [28],

$$\widehat{\mathrm{Var}}\left(\log \hat{\psi}_{\mathrm{MH}}\right) = \frac{\sum\limits_{k=1}^{K} P_k R_k}{2\left(\sum\limits_{k=1}^{K} R_k\right)^2} + \frac{\sum\limits_{k=1}^{K} (P_k S_k + Q_k R_k)}{2 \sum\limits_{k=1}^{K} R_k \sum\limits_{k=1}^{K} S_k} + \frac{\sum\limits_{k=1}^{K} Q_k S_k}{2\left(\sum\limits_{k=1}^{K} S_k\right)^2} \tag{3.24}$$

with $P_k = \dfrac{a_k + d_k}{n_k}$, $Q_k = \dfrac{b_k + c_k}{n_k}$, $R_k = \dfrac{a_k d_k}{n_k}$, and $S_k = \dfrac{b_k c_k}{n_k}$.

Risk Ratio

The pooled risk ratio $\hat{\phi}_{\text{MH}}$ is calculated by combining individual risk ratios $\hat{\phi}_k$ on the natural scale

$$\hat{\phi}_{\text{MH}} = \frac{\sum\limits_{k=1}^{K} w_k \hat{\phi}_k}{\sum\limits_{k=1}^{K} w_k} \tag{3.25}$$

using weights $w_k = \dfrac{(a_k + b_k)c_k}{n_k}$.

A robust estimator of the variance of the logarithm of $\hat{\phi}_{\text{MH}}$ is given by Greenland and Robins [21]

$$\widehat{\text{Var}} \left(\log \hat{\phi}_{\text{MH}} \right) = \frac{\sum\limits_{k=1}^{K} \dfrac{(a_k + b_k)(c_k + d_k)(a_k + c_k) - a_k c_k n_k}{n_k^2}}{\sum\limits_{k=1}^{K} \dfrac{a_k(c_k + d_k)}{n_k} \sum\limits_{k=1}^{K} \dfrac{c_k(a_k + b_k)}{n_k}} . \tag{3.26}$$

Risk Difference

The pooled risk difference $\hat{\eta}_{\text{MH}}$ is calculated by combining risk differences $\hat{\eta}_k$

$$\hat{\eta}_{\text{MH}} = \frac{\sum\limits_{k=1}^{K} w_k \hat{\eta}_k}{\sum\limits_{k=1}^{K} w_k} \tag{3.27}$$

with weights $w_k = \dfrac{(a_k + b_k)(c_k + d_k)}{n_k}$.

A robust estimator of the variance of $\hat{\eta}_{\text{MH}}$ [21] is given by

$$\widehat{\text{Var}} \left(\hat{\eta}_{\text{MH}} \right) = \frac{\sum\limits_{k=1}^{K} \dfrac{(a_k b_k n_c)^3 + (c_k d_k n_e)^3}{(n_e n_c (n_e + n_c))^2}}{\left(\sum\limits_{k=1}^{K} \dfrac{(a_k + b_k)(c_k + d_k)}{n_k} \right)^2} . \tag{3.28}$$

A $(1 - \alpha)$ confidence interval for $\log \hat{\psi}_{\mathrm{MH}}$ or $\log \hat{\phi}_{\mathrm{MH}}$ is calculated using (3.22) with $\hat{\theta}_F$ equal to $\log \hat{\psi}_{\mathrm{MH}}$ or $\log \hat{\phi}_{\mathrm{MH}}$ and S.E. $(\hat{\theta}_F)$ the square root of (3.24) and (3.26), respectively. Typically, this interval is back-transformed to the original scale.

Example 3.14 The following R command performs a meta-analysis for the high-dose chemotherapy dataset using the odds ratio as measure of treatment effect (argument `sm="OR"`) based on the Mantel–Haenszel method (the argument `method="MH"` is the default).

```
> mb1.mh <- metabin(Ee, Ne, Ec, Nc, sm="OR",
+                   data=data7, studlab=study)
> print(summary(mb1.mh), digits=2)
Number of studies combined: k=14

                        OR        95%-CI      z    p-value
Fixed effect model    1.35 [1.12; 1.61] 3.21    0.0013
*** Output truncated ***
```

The Mantel–Haenszel estimate is slightly different from the fixed effect estimate obtained using the inverse variance method. Thus measures which depend on the value of the fixed effect estimate are also slightly different, i.e. τ^2, H^2, I^2, and Q.

A forest plot, following analysis using the Mantel–Haenszel method is shown in Fig. 3.4. This was produced using the following R command.

```
> forest(mb1.mh, comb.random=FALSE, hetstat=FALSE,
+        text.fixed="MH estimate")
```
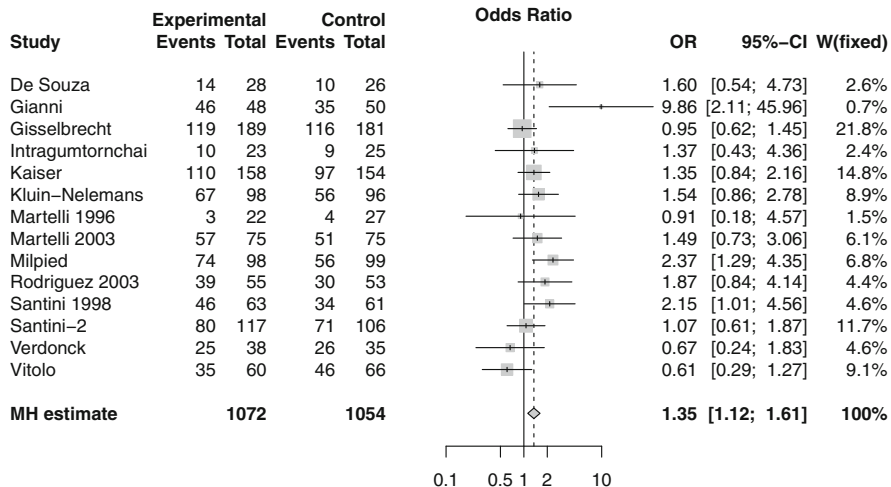

Fig. 3.4 Meta-analysis on high-dose chemotherapy [20]: forest plot based on Mantel–Haenszel method using the `forest` function

The Mantel Haenzel method is only used for pooling the data (i.e. calculating the weights); thus for individual studies estimated odds ratio and 95 % confidence interval are identical to those in Fig. 3.3.   □

Example 3.15 The following R command can be used to calculate the Mantel–Haenszel estimate in the hypertension in women dataset.

```
> mb2.mh <- metabin(Ee, Ne, Ec, Nc, sm="OR", method="MH",
+                   data=data8, studlab=study)
> print(summary(mb2.mh), digits=2)
Number of studies combined: k=11

                        OR       95%-CI     z  p-value
Fixed effect model   0.75 [0.56; 1.00] -1.98   0.0479
*** Output truncated ***
```

The Mantel–Haenszel odds ratio and its standard error are slightly smaller than results for the inverse variance method (see Example 3.13). This leads to a statistically significant result.   □

3.3.3 Peto Method

The Peto method for pooling is a variant of the inverse variance method for meta-analysis described in Sect. 3.3.1. The Peto odds ratio which is introduced in Sect. 3.2 and its variance are calculated for each study. These quantities are used in the inverse variance method to calculate an overall estimate.

$$\hat{\psi}_{\text{Peto}} = \exp\left(\frac{\sum\limits_{k=1}^{K} w_k \log \hat{\psi}_k^*}{\sum\limits_{k=1}^{K} w_k} \right) \tag{3.29}$$

with $w_k = 1/\widehat{\text{Var}}(\log \hat{\psi}_k^*)$; $\hat{\psi}_k^*$ and $\widehat{\text{Var}}(\log \hat{\psi}_k^*)$ as defined in (3.19) and (3.20), respectively.

The variance of the log odds ratio is given by

$$\widehat{\text{Var}}(\log \hat{\psi}_{\text{Peto}}) = \frac{1}{1/\sum\limits_{k=1}^{K} \widehat{\text{Var}}(\log \hat{\psi}_k^*)}. \tag{3.30}$$

As usual, a $(1-\alpha)$ confidence interval for $\log \hat{\psi}_{\text{Peto}}$ can be calculated using (3.22) with $\hat{\theta}_F = \log \hat{\psi}_{\text{Peto}}$, and S.E. $(\log \hat{\psi}_{IV})$ given by the square root of (3.30). Typically, this is then back-transformed to the original scale.

For the Peto method to combine individual log odds ratios, no correction for zero cell counts is necessary. The method has been found to perform well in meta-analysis with very sparse data [6].

Example 3.16 A meta-analysis of the high-dose chemotherapy dataset based on the Peto method can be performed with the `metabin` function.

```
> mb1.peto <- metabin(Ee, Ne, Ec, Nc, sm="OR", method="P",
+                     data=data7, studlab=study)
> print(summary(mb1.peto), digits=2)
Number of studies combined: k=14

                      OR       95%-CI      z  p-value
Fixed effect model  1.35 [1.12; 1.61] 3.22   0.0013
*** Output truncated ***
```

The result labeled as "fixed effect model" is actually the Peto odds ratio which—in this example—is almost identical to the result of the Mantel–Haenszel method (see Example 3.14).   □

Example 3.17 The following R code performs a meta-analysis based on the Peto method for the hypertension in women dataset.

```
> mb2.peto <- metabin(Ee, Ne, Ec, Nc, sm="OR", method="Peto",
+                     data=data8, studlab=study)
> print(summary(mb2.peto), digits=2)
Number of studies combined: k=11

                      OR     95%-CI      z  p-value
Fixed effect model  0.75 [0.56; 1] -1.99   0.0465
*** Output truncated ***
```

As in the previous example, the result for the Peto odds ratio (fixed effect estimate) is almost identical to the result for the Mantel–Haenszel method (see Example 3.15).   □

3.4 Random Effects Model

In a random effects model, the assumption of a constant treatment effect across studies is relaxed by allowing the treatment effect from each study to have a probability distribution about the pooled treatment effect. Usually a normal distribution is used, so that

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

where the u's and ϵ's are independent. The between-study variance τ^2 describes the extent of heterogeneity between individual study results. For the odds ratio and risk

ratio, the random effects model is estimated using the log odds ratio or log risk ratio, and the corresponding variance.

See Sect. 2.3 for a more general introduction of the random effects model.

3.4.1   DerSimonian–Laird Method

The most widely used method to estimate the between-study variance τ^2 is due to DerSimonian and Laird [12] which is the only method available in RevMan 5 [35]. Accordingly, we describe this method in the context of a meta-analysis with a binary outcome.

The DerSimonian and Laird [12] estimate is given by

$$\hat{\tau}^2 = \frac{Q - (K - 1)}{\sum_{k=1}^{K} w_k - \dfrac{\sum_{k=1}^{K} w_k^2}{\sum_{k=1}^{K} w_k}},$$

where Q, the heterogeneity statistic, is given by $Q = \sum_{k=1}^{K} w_k (\hat{\theta}_k - \hat{\theta}_F)^2$ and $w_k = \widehat{\mathrm{Var}}\,(\hat{\theta}_k)^{-1}$. The estimator $\hat{\tau}^2$ is set to zero if $Q < K - 1$.

Following RevMan 5 [35], $\hat{\theta}_F = \log \hat{\psi}_{\mathrm{MH}}$ as defined in (3.23) is used in the `metabin` function to calculate the heterogeneity statistic Q for the Mantel–Haenszel method. For the Peto method, $\hat{\theta}_k$ is the natural logarithm of (3.19), w_k the reciprocal of (3.20) and $\hat{\theta}_F$ the natural logarithm of (3.29) yielding a different value for the heterogeneity statistic Q. Accordingly, the DerSimonian–Laird estimates $\hat{\tau}^2$ will also differ for the inverse variance, Mantel–Haenszel, and Peto method.

Using the DerSimonian–Laird method to estimate τ^2, the pooled log odds ratio in the random effects model is calculated by

$$\hat{\theta}_{\mathrm{DL}} = \frac{\sum_{k=1}^{K} w_k^* \hat{\theta}_k}{\sum_{k=1}^{K} w_k^*} \tag{3.31}$$

with weights $w_k^* = (\widehat{\mathrm{Var}}\,(\hat{\theta}_k) + \hat{\tau}^2)^{-1}$.

The variance of $\hat{\theta}_{\mathrm{DL}}$ is estimated by $(\sum_{k=1}^{K} w_k^*)^{-1}$, and a $(1 - \alpha)$ confidence interval is given by

$$\hat{\theta}_{\mathrm{DL}} \ \pm \ z_{1-\frac{\alpha}{2}} \ \sqrt{\mathrm{Var}\,(\hat{\theta}_{\mathrm{DL}})}.$$

If $\tau^2 = 0$, the DerSimonian–Laird estimator corresponds to the inverse variance estimator given in Eq. (3.21).

Note as the DerSimonian–Laird estimates $\hat{\tau}^2$ typically differ for the inverse variance, Mantel–Haenszel, and Peto method the random effects estimate $\hat{\theta}_{DL}$ will also differ for these methods.

Example 3.18 The results of the random effects model are automatically returned after any call to the `metabin` function.

```
> print(summary(mb1), digits=2)
Number of studies combined: k=14

                        OR       95%-CI     z  p-value
Fixed effect model   1.32 [1.10; 1.59] 2.97   0.003
Random effects model 1.37 [1.06; 1.77] 2.39   0.0166
*** Output truncated ***
```

Table 3.3 brings together our various analyses of the chemotherapy data. For inverse variance and Mantel–Haenszel method, the estimates of τ^2 and hence the random effects estimates are very similar; the result of the random effects model based on Peto's log odds ratio and its standard error is slightly different. In general, as is usually the case, the confidence interval for the random effects model is larger as compared to the fixed effect model resulting in a larger *p*-value. Nevertheless, both fixed effect and random effects models are consistent with increased chance of remission for patients with autologous stem cell transplantation in aggressive non-Hodgkin lymphoma who receive high-dose chemotherapy.   □

Example 3.19 Results for both fixed effect and random effects meta-analysis in the hypertension in women dataset are summarised in Table 3.4. For all three methods, the estimate $\hat{\tau}^2$ is zero indicating no between-study heterogeneity. Thus, fixed effect and random effects model estimates are identical for the inverse variance and the Peto method, respectively. The fixed effect and random effects estimates are slightly different for the Mantel–Haenszel method, because the inverse-variance weights are used in the random effects model.   □

Table 3.3 Summary of results for the chemotherapy dataset [20] given in Fig. 3.1

Method	Fixed effect model	Random effects model	$\hat{\tau}^2$
Inverse variance	1.32 [1.10; 1.59]	1.37 [1.06; 1.77]	0.0897
Mantel–Haenszel	1.35 [1.12; 1.61]	1.37 [1.06; 1.77]	0.0900
Peto	1.35 [1.12; 1.61]	1.39 [1.08; 1.80]	0.0992

Table 3.4 Summary of results for the hypertension in women dataset [27] given in Fig. 3.2

Method	Fixed effect model	Random effects model	$\hat{\tau}^2$
Inverse variance	0.78 [0.58; 1.05]	0.78 [0.58; 1.05]	0.00
Mantel–Haenszel	0.75 [0.56; 1.00]	0.78 [0.58; 1.05]	0.00
Peto	0.75 [0.56; 1.00]	0.75 [0.56; 1.00]	0.00

3.5 Heterogeneity and Subgroup Analyses

Subgroup analyses are easily performed with the `metabin` function and associated measures of heterogeneity are calculated automatically. More technical details on subgroup analyses are provided in Sect. 4.3.

Example 3.20 Bassler et al. [5] conducted a Cochrane Review to evaluate the effects of Ketotifen alone or in combination with other co-interventions in children with asthma and/or wheezing. The primary outcome was the reduction in use of rescue bronchodilators. Secondary clinical endpoints included physicians judgement on the overall efficacy of Ketotifen. For this outcome, a prespecified subgroup analysis was conducted to evaluate whether the treatment effect is different in trials with adequate blinding compared to trials with inadequate/unclear blinding. The data for this outcome is shown in Fig. 3.5.

In the systematic review, the random effects model with the risk ratio as measure of treatment effect is used throughout. The meta-analysis of the clinical judgement data contains ten trials having rather small sample sizes ranging between 20 (`Chay 1992`) and 133 (`Rackham 1989`). Despite these small sample sizes, no studies have zero cells. In three trials the method of blinding was judged adequate whereas in the remaining seven trials the method of blinding was unclear, i.e. these seven trials are likely to be a mixture of trials with and without adequate blinding.

```
> # 1. Read in the data
> data9 <- read.csv("dataset09.csv")
> # 2. Print dataset
> data9
              study Ee Ne Ec Nc              blind
1          Chay 1992  1 10  6 10 Adequate blinding
2       Rackham 1989 31 68 38 65 Adequate blinding
3   Van Asperen 1992 16 52 19 51 Adequate blinding
4         Croce 1995 19 39 17 36    Method unclear
5  de Benedictis 1990  7 34 35 41    Method unclear
6         Longo 1986 10 18 15 18    Method unclear
7       Montoya 1988  6 20 14 20    Method unclear
8       Mulhern 1982  6 16  8 15    Method unclear
9           Salmon  8 28 16 34    Method unclear
10        Spicak 1983  9 25 20 25    Method unclear
> # 3. Calculate experimental and control event rates
> summary(data9$Ee/data9$Ne)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1000  0.2893  0.3338  0.3433  0.4357  0.5556
> summary(data9$Ec/data9$Nc)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3725  0.4875  0.5923  0.6220  0.7750  0.8537
```

Fig. 3.5 Data from meta-analysis on Ketotifen in children with asthma and/or wheezing [5]; for details on table headers, see Table 3.2

The following R code fits the fixed effect and random effects model.

```
> mb3 <- metabin(Ee, Ne, Ec, Nc, sm="RR", method="I",
+                data=data9, studlab=study)
> print(summary(mb3), digits=2)
Number of studies combined: k=10

                         RR       95%-CI      z  p-value
Fixed effect model   0.65 [0.55; 0.78] −4.81 < 0.0001
Random effects model 0.60 [0.46; 0.79] −3.64   0.0003

Quantifying heterogeneity:
tau^2 = 0.0915; H = 1.45 [1.01; 2.07]; I^2 = 52.3% [2.2%; 76.7%]

Test of heterogeneity:
    Q d.f.  p-value
 18.87    9   0.0263

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

Both fixed effect and random effects estimates show a statistically significant benefit of Ketotifen, according to the clinical judgement of the physicians. However, both heterogeneity measures Q and I^2 indicate a large amount of heterogeneity in this meta-analysis.

A subgroup analysis comparing trials with adequate blinding to trials with inadequate/unclear blinding can be readily conducted as follows.

```
> mb3s <- update(mb3, byvar=blind, print.byvar=FALSE)
> print(summary(mb3s), digits=2)
Number of studies combined: k=10

                         RR       95%-CI      z  p-value
Fixed effect model   0.65 [0.55; 0.78] −4.81 < 0.0001
Random effects model 0.60 [0.46; 0.79] −3.64   0.0003

Quantifying heterogeneity:
tau^2 = 0.0915; H = 1.45 [1.01; 2.07]; I^2 = 52.3% [2.2%; 76.7%]

Test of heterogeneity:
    Q d.f.  p-value
 18.87    9   0.0263

Results for subgroups (fixed effect model):
                  k    RR       95%-CI      Q    tau^2    I^2
Adequate blinding 3 0.77 [0.58; 1.01]  2.49   0.0237 19.7%
Method unclear    7 0.59 [0.47; 0.74] 14.29   0.1282   58%

Test for subgroup differences (fixed effect model):
                Q d.f.  p-value
Between groups 2.09    1   0.1483
Within groups 16.79    8   0.0324
```

```
Results for subgroups (random effects model):
                    k   RR        95%-CI     Q     tau^2   I^2
Adequate blinding   3 0.75 [0.53; 1.08]   2.49    0.0237 19.7%
Method unclear      7 0.56 [0.39; 0.79]  14.29    0.1282   58%

Test for subgroup differences (random effects model):
                Q d.f.   p-value
Between groups  1.4    1   0.2367

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

The subgroup analysis is conducted by using the argument `byvar` in the `update.meta` function. The subgroup analysis shows that method of blinding does not explain the statistical heterogeneity between trials neither in the fixed effect model (Q_{between} = 2.09, p-value = 0.1483) nor in the random effects model (Q_{between} = 1.40, p-value = 0.2367).   □

3.6 Summary

In this chapter, the most commonly used effect measures for binary outcomes, i.e. odds ratio, risk ratio, risk difference, together with the arcsine difference, have been described. We have illustrated the use of both base R code and the `metabin` function for estimating these effect measures, their standard errors and confidence intervals, and performing fixed effect and random effects meta-analysis. In addition, we have described the common approaches for tackling the issues raised by sparse binary data, especially situations where one or more of the constituent studies in a meta-analysis have zero counts.

The most important meta-analytical methods for binary outcome data, i.e. inverse variance, Mantel–Haenszel, Peto method, and the DerSimonian–Laird random effects model have been introduced and their estimation in R has been illustrated.

We concluded with a detailed example of using R, and specifically the `metabin` function, for subgroup analyses.

References

1. Agresti, A.: Categorical Data Analysis, 2nd edn. Wiley, New York (2002)
2. A'hern, R.P.: Widening eligibility to phase II trials: constant arcsine difference phase II trials. Control. Clin. Trials **25**, 251–264 (2004)
3. Anscombe, F.: The transformation of Poisson, binomial and negative-binomial data. Biometrika **35**(3/4), 246–254 (1948)
4. Armitage, P., Berry, G., Matthews, J.N.S.: Statistical Methods in Medical Research. Blackwell Science, Oxford (1987)

5. Bassler, D., Mitra, A., Ducharme, F., Forster, J., Schwarzer, G.: Ketotifen alone or as additional medication for long-term control of asthma and wheeze in children. Cochrane Database Syst. Rev. **1**, CD001,384 (2004). doi:10.1002/14651858.CD001384.pub2

6. Bradburn, M.J., Deeks, J.J., Berlin, J.A., Localio, A.R.: Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Stat. Med. **26**, 53–77 (2007)

7. Breslow, N.E.: Odds ratio estimators when the data are sparse. Biometrika **68**, 73–84 (1981)

8. Brockhaus, A.C., Bender, R., Skipka, G.: The Peto odds ratio viewed as a new effect measure. Stat. Med. **33**(28), 4861–4874 (2014). doi:10.1002/sim.6301

9. Choudhury, J.B.: Non-parametric confidence interval estimation for competing risks analysis: application to contraceptive data. Stat. Med. **21**, 1129–1144 (2002)

10. Deeks, J.J.: Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat. Med. **21**, 1575–1600 (2002)

11. Deeks, J.J., Altman, D.G.: Effect measures for meta-analysis of trials with binary outcome. In: Egger, M., Smith, G.D., Altman, D.G. (eds.) Systematic Reviews in Health Care: Meta-Analysis in Context, 2 edn., pp. 313–335. BMJ Publishing Group, London (2001)

12. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Control. Clin. Trials **7**, 177–188 (1986)

13. Duncan, B., Olkin, I.: Bias of estimates of the number needed to treat. Stat. Med. **24**, 1837–1848 (2005)

14. Emerson, J.D.: Combining estimates of the odds ratio: the state of the art. Stat. Methods Med. Res. **3**, 157–178 (1994)

15. Engels, E.A., Schmid, C.H., Terrin, N., Olkin, I., Lau, J.: Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat. Med. **19**, 1707–1728 (2000)

16. Fisher, R., Yates, F.: Statistical Tables for Biological, Agricultural and Medical Research, 3rd edn. Oliver and Boyd, Edinburgh (1948)

17. Fleiss, J.L.: The statistical basis of meta-analysis. Stat. Methods Med. Res. **2**, 121–145 (1993)

18. Freeman, M.F., Tukey, J.W.: Transformations related to the angular and the square root. Ann. Math. Stat. **21**, 607–611 (1950)

19. Gart, J.J., Zweifel, J.R.: On the bias of various estimators of the logit and its variance with application to quantal bioassay. Biometrika **54**, 181–187 (1967)

20. Greb, A., Bohlius, J., Schiefer, D., Schwarzer, G., Schulz, H., Engert, A.: High-dose chemotherapy with autologous stem cell transplantation in the first line treatment of aggressive non-Hodgkin Lymphoma (NHL) in adults. Cochrane Database Syst. Rev. **1**, CD004,024 (2008). doi:10.1002/14651858.CD004024.pub2

21. Greenland, S., Robins, J.M.: Estimation of a common effect parameter from sparse follow-up data. Biometrics **41**, 55–68 (1985). (C/R: V45 p1323–1324)

22. Greenland, S., Salvan, A.: Bias in the one-step method for pooling study results. Stat. Med. **9**, 247–252 (1990)

23. Hoover, D.R., Blackwelder, W.C.: Allocation of subjects to test null relative risks smaller than one. Stat. Med. **20**, 3071–3082 (2001)

24. Mantel, N., Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. J. Natl. Cancer Inst. **22**, 719–748 (1959)

25. McCullagh, P., Nelder, J.: Generalized Linear Models. Chapman and Hall, London (1989)

26. Pettigrew, H.M., Gart, J.J., Thomas, D.G.: The bias and higher cumulants of the logarithm of a binomial variate. Biometrika **73**, 425–435 (1986)

27. Quan, A., Kerlikowske, K., Gueyffier, F., Boissel, J.P., for the INDANA Investigators: Pharmacotherapy for hypertension in women of different races. Cochrane Database Syst. Rev. **2**, CD002,146 (2000). doi:10.1002/14651858.CD002146

28. Robins, J., Greenland, S., Breslow, N.E.: A general estimator for the variance of the Mantel–Haenszel odds ratio. Am. J. Epidemiol. **124**, 719–723 (1986)

29. Robins, J., Breslow, N.E., Greenland, S.: Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. Biometrics **42**, 311–323 (1986)

30. Rücker, G., Schwarzer, G., Carpenter, J.R.: Arcsine test for publication bias in meta-analyses with binary outcomes. Stat. Med. **27**(5), 746–763 (2008)
31. Schwarzer, G.: meta: An R package for meta-analysis. R News **7**(3), 40–45 (2007). http://www.cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf
32. Schwarzer, G.: Meta: meta-analysis with R (2014). http://www.cran.R-project.org/package= meta. R package version 4.0-2
33. StataCorp.: Stata Statistical Software: Release 13. StataCorp LP, College Station, TX (2013)
34. Sweeting, M.J., Sutton, A.J., Lambert, P.C.: What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat. Med. **23**, 1351–1375 (2004)
35. The Cochrane Collaboration: Review Manager (RevMan) [Computer program]. Version 5.3. The Nordic Cochrane Centre, Copenhagen (2014)
36. Yusuf, S., Peto, R., Lewis, J., Collins, R., Sleight, P.: Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog. Cardiovasc. Dis. **27**, 335–371 (1985)

Chapter 4
Heterogeneity and Meta-Regression

Various aspects of statistical heterogeneity have been introduced briefly in previous chapters. We now review and develop these ideas and describe the connection between subgroup analysis and meta-regression.

4.1 Sources of Heterogeneity

In meta-analysis, three principal sources of heterogeneity can be distinguished [13]:

- *Clinical baseline heterogeneity* between patients from different studies, measured, e.g., in patient baseline characteristics and not necessarily reflected on the outcome measurement scale,
- *Statistical heterogeneity*, quantified on the outcome measurement scale, that may or may not be clinically relevant and may or may not be statistically significant,
- *Heterogeneity from other sources*, e.g. design-related heterogeneity.

In this chapter, we only deal with statistical heterogeneity. There is a substantial literature on statistical heterogeneity in meta-analysis, for example [6, 8, 9, 11–15, 18, 19].

4.2 Measures of Heterogeneity

Let K denote the number of studies in a meta-analysis. Further, let $\hat{\theta}_k$ be the treatment effect estimate (e.g., a log odds ratio), $\hat{\sigma}_k^2$ its estimated variance, and $w_k = 1/\hat{\sigma}_k^2$ the corresponding weight from study k, $k = 1, \ldots, K$. Several measures

of statistical heterogeneity are widely used:

1. Cochran's Q statistic, which under the null hypothesis of no heterogeneity follows a χ^2-distribution with $K - 1$ degrees of freedom [3]. Q is given by

$$Q = \sum_{k=1}^{K} w_k \left(\hat{\theta}_k - \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{\sum_{k=1}^{K} w_k} \right)^2 ;$$

2. Higgins' and Thompson's I^2 [9], derived from Cochran's Q by defining

$$I^2 = \max\left\{ 0, \frac{Q - (K - 1)}{Q} \right\} ;$$

3. the between-study variance, τ^2, as estimated in a random effects meta-analysis. There are several proposals for estimating τ^2 in a meta-analysis (see Sect. 2.3.1). Nevertheless, most reviewers use the moment-based estimate of τ^2 [5], implemented in RevMan 5 [16] and calculated as

$$\hat{\tau}^2 = \max\left\{ 0, \frac{Q - (K - 1)}{\sum_{k=1}^{K} w_k - \frac{\sum_{k=1}^{K} w_k^2}{\sum_{k=1}^{K} w_k}} \right\} ;$$

4. H^2, derived from Cochran's Q by defining [9]

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

   and

5. R^2, similar to H^2 and calculated from τ^2 and a so-called "typical" within-study variance σ^2 (which must be estimated), and defined as:

$$R^2 = \frac{\tau^2 + \sigma^2}{\sigma^2}.$$

Some measures are directly related [9, 13], and others approximately related. Table 4.1 shows key properties of the various measures; more details are given in Higgins and Thompson [9]. In summary:

1. Q, which follows a χ^2-distribution with $K - 1$ degrees of freedom under the null hypothesis of no heterogeneity, is the weighted sum of squared differences

Table 4.1 Properties of measures of heterogeneity

Measure	Measured on		Increasing with	
			Number of studies in	Precision (size of
	Scale	Range	meta-analysis	studies)
Q	Absolute	$[0, \infty)$	Yes	Yes
I^2	Percent	$[0, 100\,\%]$	No	Yes
τ	Outcome	$[0, \infty)$	No	No
H^2	Absolute	$[1, \infty)$	No	Yes
R^2	Absolute	$[1, \infty)$	No	Yes

between the study means and the fixed effect estimate. It always increases with the number of studies, K, in the meta-analysis.

2. In contrast to Q, the statistic I^2 was introduced by Higgins and Thompson [9] as a measure independent of K, the number of studies in the meta-analysis. I^2 is interpreted as the percentage of variability in the treatment estimates which is attributable to heterogeneity between studies rather than to sampling error.

3. τ^2 describes the underlying between-study variability. Its square root, τ, is measured in the same units as the outcome. Its estimate does not systematically increase with either the number of studies or the sample size.

4. H^2 is a test statistic. It describes the relative difference between the observed Q and its expected value in the absence of heterogeneity. Thus it does not systematically increase with the number of studies [9]. H corresponds to the residual standard deviation in a radial plot [7]. $H = 1$ indicates perfect homogeneity.

5. R^2 is the square of a statistic R which describes the inflation of the random effects confidence interval compared to that from the fixed effect model. It does not increase with K. $R^2 = 1$ indicates perfect homogeneity [9].

Notice that, in contrast to τ^2, the measures Q, I^2, H and R all depend on the precision, which is proportional to study size. Thus, given an underlying model, if the study sizes are enlarged, the confidence intervals become smaller and the heterogeneity, measured for example using I^2, increases. This is reflected in the interpretation: I^2 is the percentage of variability that is due to between-study heterogeneity and $1 - I^2$ is the percentage of variability that is due to sampling error. When the studies become very large, the sampling error tends to 0 and I^2 tends to 1. Such heterogeneity may not be clinically relevant.

Example 4.1 The standard printout of a meta-analysis object contains the most commonly used measures of heterogeneity, i.e. Q, I^2, and τ^2, see Fig. 2.2.[1] As these measures of heterogeneity are part of meta-analysis object `mc1` created in Example 2.3 we can print them directly.

---

[1]The test statistic H is also printed in Fig. 2.2, however, it is rarely used in meta-analysis.

```
> round(mc1$Q, 2) # Cochran's Q statistic
[1] 17.57
> round(100*c(mc1$I2, mc1$lower.I2, mc1$upper.I2), 1) # I-squared
[1]  8.9  0.0 45.3
> round(mc1$tau^2, 4) # Between-study variance tau-squared
[1] 2.4374
```

These are the same values as given in Fig. 2.2.   □

4.3 Test for Subgroup Differences

Often we are interested in comparing subgroups of studies in a meta-analysis. For example, if studies differ in the eligibility criteria for patients, we might ask whether the treatment is more effective in some studies than in others. In this case, the factor which defines the subgroups is said to be an effect moderator. To address this question we need to test for a treatment-subgroup interaction, i.e. whether the treatment effect is modified, or moderated, by subgroup membership.

There are a number of issues we need to be aware of when considering using subgroup analyses or its more general form of meta-regression. The first is that if—as in many medical meta-analyses—there are only a few studies, we do not have enough information to look for treatment-subgroup interactions. Specifically, it does not make sense to look for a treatment-subgroup interaction if there are only three studies, or to investigate two or more factors if there are less than ten studies in a meta-analysis [10].

The second issue is that while meta-regression and subgroup analysis are particularly useful when we have individual participant data, in meta-analysis of aggregate data their use is limited [17]. This is because we typically only have a few study level variables to include in such analyses. We should not include the aggregate value of individual patient characteristics because of the risk of the so-called ecological bias [1, 17]. For example, if age is suspected of being an effect moderator, it is tempting to use mean age (or the proportion of patients older than some threshold) as a covariate in meta-regression. However, the fact that, say, studies having older patients *on average* show a smaller treatment effect does not support the conclusion that the treatment is less effective for the elderly at an individual level.

We have already shown how to perform tests for subgroup differences in Sects. 2.5 and 3.5. In this section we give additional technical details about these statistical tests. Those readers who are not interested in these details may skip to the next section.

Following Borenstein et al. [2, Chap. 19] we distinguish three different models which we may use to conduct tests for subgroup differences:

- fixed effect models,
- random effects models with separate estimates for between-study variance τ^2 across the subgroups, and
- random effects models with a common estimate for τ^2.

Borenstein et al. [2] also describe three methods to test for subgroup differences which are mathematically equivalent. We restrict our attention to their Method 3, i.e. a Q-test for heterogeneity.

Assume that the studies can be divided into G subgroups. Then, estimates of the pooled treatment effect and its corresponding standard errors can be calculated for each of G subgroups. Let $\hat{\theta}_{gk}$ denote the intervention effect estimate from study k in subgroup g with $k = 1, \ldots, K_g$, $g = 1, \ldots, G$, and θ_g denote the intervention effect in subgroup g, which we wish to estimate. Furthermore, denote by $\hat{\sigma}_{gk}^2$ the sample estimate of $\mathrm{Var}\,(\hat{\theta}_{gk})$.

4.3.1 Fixed Effect Model

The model with G fixed subgroup effects is

$$\hat{\theta}_{gk} = \theta_g + \hat{\sigma}_{gk}\epsilon_{gk}, \quad \epsilon_{gk} \overset{\text{i.i.d.}}{\sim} N(0,1) \tag{4.1}$$

for $g = 1, \ldots, G$.

We now consider the fixed effect estimate of θ_g in subgroup g, denoted by $\hat{\theta}_{Fg}$. Given $(\hat{\theta}_{gk}, \hat{\sigma}_{gk})$, $k = 1, \ldots, K_g$, the maximum likelihood estimate under model (4.1) is

$$\hat{\theta}_{Fg} = \frac{\sum_{k=1}^{K_g} \hat{\theta}_{gk}/\hat{\sigma}_{gk}^2}{\sum_{k=1}^{K_g} 1/\hat{\sigma}_{gk}^2} = \frac{\sum_{k=1}^{K_g} w_{gk}\hat{\theta}_{gk}}{\sum_{k=1}^{K_g} w_{gk}}.$$

Accordingly, $\hat{\theta}_{Fg}$ is a weighted average of the individual effect estimates $\hat{\theta}_{gk}$ with weights $w_{gk} = 1/\hat{\sigma}_{gk}^2$.

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

$$\widehat{\mathrm{Var}}\,(\hat{\theta}_{Fg}) = \hat{\sigma}_g^2 = \frac{1}{\sum_{k=1}^{K_g} w_{gk}} = \frac{1}{w_g}.$$

The estimated treatment effects $\hat{\theta}_{Fg}$ and their variances $\hat{\sigma}_g^2, g = 1, \ldots, G$, can be used as inputs for the generic inverse variance method introduced in Sect. 2.2. Cochran's Q statistic

$$Q_F = \sum_{g=1}^{G} w_g \left(\hat{\theta}_{Fg} - \frac{\sum_{g=1}^{G} w_g \hat{\theta}_{Fg}}{\sum_{g=1}^{G} w_g} \right)^2 \tag{4.2}$$

can be utilised to test for differences between the G subgroups. The statistic Q_F follows a χ^2-distribution with $G-1$ degrees of freedom [3] under the null hypothesis of no heterogeneity between subgroups, i.e. $\theta_1 = \cdots = \theta_G = \theta$.

Example 4.2 In Example 2.12 a subgroup analysis was conducted in a meta-analysis on the effects of mucolytic agents and placebo in patients with chronic bronchitis/obstructive pulmonary disease. A test of subgroup differences for the outcome of interest (mean number of acute exacerbations per month) was calculated using the `metacont` function from R package **meta**. Here, we redo these calculations step by step.

It is assumed that R dataset `data3` is still available in the R session (see Fig. 2.7). The dataset consists of 23 studies which are divided in two subgroups. As two studies have zero standard errors, effectively subgroups with 4 and 17 studies are utilised in analyses.

In a first step, a meta-analysis is conducted for each subgroup respectively using the `subset` argument in the `metacont` function.

```
> # Conduct meta-analysis for first subgroup:
> mc3s1 <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                   studlab=paste(author, year),
+                   subset=duration=="<= 3 months")
> # Conduct meta-analysis for second subgroup:
> mc3s2 <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                   studlab=paste(author, year),
+                   subset=duration=="> 3 months")
```

To conduct the test for subgroup differences estimated treatment effects as well as corresponding standard errors are needed.

```
> # Subgroup treatment effects (fixed effect model)
> TE.duration <- c(mc3s1$TE.fixed, mc3s2$TE.fixed)
> # Corresponding standard errors (fixed effect model)
> seTE.duration <- c(mc3s1$seTE.fixed, mc3s2$seTE.fixed)
```

A meta-analysis of subgroup effects using the generic inverse variance method is conducted using the following command.

```
> mh1 <- metagen(TE.duration, seTE.duration,
+                sm="MD",
+                studlab=c("<= 3 months", " > 3 months"),
```

```
+                    comb.random=FALSE)
> print(mh1, digits=2)
                MD       95%-CI %W(fixed)
<= 3 months  -0.13 [-0.17; -0.09]      5.5
 > 3 months  -0.04 [-0.05; -0.03]     94.5
*** Output truncated ***
Test of heterogeneity:
    Q d.f.  p-value
 20.73    1 < 0.0001
*** Output truncated ***
```

This meta-analysis of subgroup mean values yields identical results to those given in Example 2.12 in the printout for R object `mc3s` which considers subgroups using argument `byvar`.

```
> print(summary(mc3s), digits=2)
Number of studies combined: k=21
*** Output truncated ***
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
Between groups  20.73    1 < 0.0001
Within groups  117.35   19 < 0.0001
*** Output truncated ***
```

The "Test of heterogeneity" in the meta-analysis of subgroup means corresponds to the "Test for subgroup differences (fixed effect model)—Between groups" in the meta-analysis of all studies considering subgroups.

We can also extract the corresponding test statistics directly from the two R objects.

```
> mh1$Q
[1] 20.72833
> mc3s$Q.b.fixed
[1] 20.72833
```

Both methods report the same value, i.e. a highly statistical significant difference between the two subgroups.   □

4.3.2 Random Effects Model with Separate Estimates of τ^2

The model with G random subgroup effects and separate between-study variances τ_g^2 is

$$\hat{\theta}_{gk} = \theta_g + u_{gk} + \sigma_{gk}\epsilon_{gk}, \quad \epsilon_{gk} \overset{\text{i.i.d.}}{\sim} N(0,1); \; u_{gk} \overset{\text{i.i.d.}}{\sim} N(0, \tau_g^2), \tag{4.3}$$

with $k = 1, \ldots, K_g$, $g = 1, \ldots, G$ and independent error terms u and ϵ. The fixed effect model is a special case of the random effects model when the between-study variances τ_g^2 are equal to 0. Accordingly, in the random effects model with separate between-study variances an additional parameter has to be estimated for each subgroup.

Using appropriate estimates $\hat{\tau}_g^2$, see Eq. (4.6) below, the random effects estimate $\hat{\theta}_{Rg}$ and its variance can be calculated as

$$\hat{\theta}_{Rg} = \frac{\sum_{k=1}^{K_g} w_{gk}^* \hat{\theta}_{gk}}{\sum_{k=1}^{K_g} w_{gk}^*} \tag{4.4}$$

$$\mathrm{Var}\,(\hat{\theta}_{Rg}) = \frac{1}{\sum_{k=1}^{K_g} w_{gk}^*} = \frac{1}{w_g^*} \tag{4.5}$$

with weights $w_{gk}^* = 1/(\hat{\sigma}_{gk}^2 + \hat{\tau}_g^2)$.

Again, estimated treatment effects $\hat{\theta}_{Rg}$ and variances $\hat{\sigma}_g^2 + \hat{\tau}_g^2$, $g = 1, \ldots, G$, can be used as inputs for the generic inverse variance method, see Example 4.2. Cochran's Q statistic

$$Q_R = \sum_{g=1}^{G} w_g^* \left(\hat{\theta}_{Rg} - \frac{\sum_{g=1}^{G} w_g^* \hat{\theta}_{Rg}}{\sum_{g=1}^{G} w_g^*} \right)^2$$

can be utilised to test for differences in the G subgroups. The statistic Q_R follows a χ^2-distribution with $G - 1$ degrees of freedom [3] under the null hypothesis of no heterogeneity between subgroups.

Estimate Separate Between-Study Variances (DerSimonian–Laird Method)

Defining a scaling factor $C = \sum_{k=1}^{K} w_k - \sum_{k=1}^{K} w_k^2 / \sum_{k=1}^{K} w_k$, the DerSimonian–Laird estimate of the between-study variance in a meta-analysis without subgroups described in Sect. 4.2 can be written as

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

Accordingly, for each subgroup, $g = 1, \ldots, G$, the corresponding quantities Q_g, K_g, and C_g, can be calculated and separate estimates of the between-study variances $\hat{\tau}_g^2$ are given by

$$\hat{\tau}_g^2 = \frac{Q_g - (K_g - 1)}{C_g}. \tag{4.6}$$

As always, estimates of between-study variances are set to zero for negative values: $\hat{\tau}_g^2 = \max\left\{0, \hat{\tau}_g^2\right\}, g = 1, \ldots, G$.

Example 4.3 R objects `mc3s1` and `mc3s2` created in Example 4.2 can be used to calculate the test for subgroup differences in the random effects model with separate between-study variances.

First, we print the estimated between-study variances $\hat{\tau}_g^2, g = 1, 2,$:

```
> data.frame(duration=c("<= 3 months", " > 3 months"),
+             tau2=round(c(mc3s1$tau^2, mc3s2$tau^2), 4))
    duration  tau2
1 <= 3 months 0.035
2  > 3 months 0.002
```

These values have already been given in Example 2.12 in the printout of R object `mc3s`. We can also check that using the quantities Q_1, $K_1 - 1$, and C_1 for the first subgroup yields the same estimate of τ_1^2:

```
> round((mc3s1$Q - (mc3s1$k-1))/mc3s1$C, 4)
[1] 0.035
```

Next we conduct a meta-analysis of the random effects treatment estimates in the two subgroups.

```
> # Subgroup treatment effects (random effects model)
> TE.duration.r <- c(mc3s1$TE.random, mc3s2$TE.random)
> # Corresponding standard errors (random effects model)
> seTE.duration.r <- c(mc3s1$seTE.random, mc3s2$seTE.random)
> # Do meta-analysis of subgroup estimates
> mh1.r <- metagen(TE.duration.r, seTE.duration.r,
+                  sm="MD",
+                  studlab=c("<= 3 months", " > 3 months"),
+                  comb.random=FALSE)
> print(mh1.r, digits=2)
                MD         95%-CI %W(fixed)
<= 3 months -0.28 [-0.50; -0.05]      1.28
 > 3 months -0.06 [-0.09; -0.04]     98.72
*** Output truncated ***
Test of heterogeneity:
    Q d.f.  p-value
 3.41    1   0.0647
*** Output truncated ***
```

The "Test of heterogeneity" corresponds to the "Test for subgroup differences (random effects model)—Between groups" in the printout of R object `mc3s` on

Page 44; both methods report a value of 3.41 for the Q-statistic with 1 degree of freedom and corresponding p-value of 0.0647.   □

4.3.3   Random Effects Model with Common Estimate of τ^2

The model with G random subgroup effects and assuming a common between-study variance $\tau_\star^2$ across subgroups is[2]

$$\hat{\theta}_{gk} = \theta_g + u_{gk} + \sigma_{gk}\epsilon_{gk}, \quad \epsilon_{gk} \overset{i.i.d.}{\sim} N(0, 1); \; u_{gk} \overset{i.i.d.}{\sim} N(0, \tau_\star^2), \tag{4.7}$$

with $k = 1, \ldots, K_g$, $g = 1, \ldots, G$ and independent error terms u and ϵ. The fixed effect model is a special case of the random effects model when the between-study variance $\tau_\star^2$ is equal to 0. In comparison to the random effects model with separate between-study variances a single additional parameter has to be estimated.

Using an appropriate estimate $\hat{\tau}_\star^2$, see Eq. (4.10) below, the random effects estimate $\hat{\theta}_{Rg}^\star$ and its variance can be calculated as

$$\hat{\theta}_{Rg}^\star = \frac{\displaystyle\sum_{k=1}^{K_g} w_{gk}^\star \hat{\theta}_{gk}}{\displaystyle\sum_{k=1}^{K_g} w_{gk}^\star} \tag{4.8}$$

$$\mathrm{Var}\left(\hat{\theta}_{Rg}\right) = \frac{1}{\displaystyle\sum_{k=1}^{K_g} w_{gk}^\star} = \frac{1}{w_g^\star}. \tag{4.9}$$

with weights $w_{gk}^\star = 1/(\hat{\sigma}_{gk}^2 + \hat{\tau}_\star^2)$.

Again, estimated treatment effects $\hat{\theta}_{Rg}^\star$ and its variance $\hat{\sigma}_g^2 + \hat{\tau}_\star^2$, $g = 1, \ldots, G$, can be used as inputs for the generic inverse variance method, see Example 4.2. Cochran's Q statistic

$$Q_R^\star = \sum_{g=1}^{G} w_g^\star \left(\hat{\theta}_{Rg}^\star - \frac{\displaystyle\sum_{g=1}^{G} w_g^\star \hat{\theta}_{Rg}^\star}{\displaystyle\sum_{g=1}^{G} w_g^\star} \right)^2$$

---

[2]Super- and sub-script "⋆" denote calculations based on the assumption of a common between-study variance across subgroups.

can be utilised to test for differences in the G subgroups. The statistic $Q_R^\star$ follows a χ^2-distribution with $G - 1$ degrees of freedom [3] under the null hypothesis of no heterogeneity between subgroups.

Estimate Common Between-Study Variance (DerSimonian–Laird Method)

Based on the quantities Q_g, K_g, and C_g, $g = 1, \ldots, G$ which have been defined in Sect. 4.3.2, a common between-study variance [2] is given by

$$\hat{\tau}_\star^2 = \frac{\sum_{g=1}^{G} Q_g - \sum_{g=1}^{G} (K_g - 1)}{\sum_{g=1}^{G} C_g} \tag{4.10}$$

with $K_g - 1$ denoting the degrees of freedom.

The estimate of the between-study variance is set to zero for negative values: $\hat{\tau}_\star^2 = \max\{0, \hat{\tau}_\star^2\}$.

Example 4.4 In Example 4.2, results of subgroup analyses have been conducted and saved in R objects mc3s1 and mc3s2. These R objects can be used to calculate the common between-study variance and thus to test for subgroup differences in the random effects model with a common between-study variance. The necessary quantities Q_g, K_g, and C_g, $g = 1, \ldots, G$ to estimate the common between-study variance are readily available.[3]

```
> # Q-statistic within subgroups
> Q.g <- c(mc3s1$Q, mc3s2$Q)
> # Degrees of freedom within subgroups
> df.Q.g <- c(mc3s1$k-1, mc3s2$k-1)
> # Scaling factor within subgroups
> C.g <- c(mc3s1$C, mc3s2$C)
```

Next we estimate the pooled between-study variance.

```
> # Calculate common estimate of tau-squared
> tau2.common <- (sum(Q.g) - sum(df.Q.g)) / sum(C.g)
> # Set negative value of tau.common to zero
> tau2.common <- ifelse(tau2.common < 0, 0, tau2.common)
> # Print common between-study variance
> round(tau2.common, 4)
[1] 0.0024
```

---

[3] We could have used list element mc3s1$df.Q instead of mc3s1$k-1.

This estimated between-study variance can be used in the `metacont` function with argument `tau.preset`.[4]

```
> mc3s.p <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                    studlab=paste(author, year),
+                    byvar=duration, print.byvar=FALSE,
+                    tau.preset=sqrt(tau2.common))
> print(summary(mc3s.p), digits=2)
Number of studies combined: k=21
*** Output truncated ***
Results for subgroups (random effects model):
             k    MD         95%-CI     Q    tau^2   I^2
<= 3 months  4 -0.21 [-0.29; -0.12] 22.43  0.0024 86.6%
> 3 months  17 -0.07 [-0.09; -0.04] 94.92  0.0024 83.1%

Test for subgroup differences (random effects model):
                 Q d.f.  p-value
Between groups 10.21   1   0.0014

Details on meta-analytical method:
- Inverse variance method
- Preset between-study variance: tau^2 = 0.0024
```

A more convenient way to conduct the test for subgroup differences is to use argument `tau.common` in the `metacont` function which calculates the pooled between-study variance directly.

```
> mc3s.c <- metacont(Ne, Me, Se, Nc, Mc, Sc, data=data3,
+                    studlab=paste(author, year),
+                    byvar=duration, print.byvar=FALSE,
+                    tau.common=TRUE, comb.fixed=FALSE)
> print(summary(mc3s.c), digits=2)
Number of studies combined: k=21
*** Output truncated ***
Results for subgroups (random effects model):
             k    MD         95%-CI     Q    tau^2   I^2
<= 3 months  4 -0.21 [-0.29; -0.12] 22.43  0.0024 86.6%
> 3 months  17 -0.07 [-0.09; -0.04] 94.92  0.0024 83.1%

Test for subgroup differences (random effects model):
                  Q d.f.  p-value
Between groups 10.21    1   0.0014
Within groups 117.35   19 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2 (assuming
  common tau^2 in subgroups)
```

---

[4]Alternatively, the command `update(mc3s, tau.preset=sqrt(tau2.common))` could be used—see Sect. 2.5.

This command produces identical results for the subgroup analysis as the call of the `metacont` function with argument `tau.preset`. Notice the slight difference under "Details on meta-analytical method". □

4.4 Meta-Regression

In the last section, tests for subgroup differences have been described which are based on a single covariate with a limited number of values, i.e. a binary or categorical covariate. Meta-regression is an extension for either more than one binary/categorical covariate or a continuous covariate. In the later case, a subgroup analysis could result in subgroups of size one, i.e. each covariate value generates a subgroup.

We consider the following meta-regression model

$$\hat{\theta}_k = \theta + \beta_1 x_{1k} + \cdots + \beta_P x_{Pk} + u_k + \sigma_k \epsilon_k, \quad \epsilon_k \overset{i.i.d.}{\sim} N(0, 1); \ u_k \overset{i.i.d.}{\sim} N(0, \tau^2), \quad (4.11)$$

with $k = 1, \ldots, K$ and independent error terms u and ϵ. As this model has both fixed effect (βs) and random effects terms (u_k with variance τ^2) this meta-regression model is also called a mixed effects model [20]. The fixed effect meta-regression is a special case of the mixed effects model when the between-study variance $\tau^2 = 0$.

4.4.1 Meta-Regression with a Categorical Covariate

A subgroup analysis with G subgroups and a common between-study variance can be written as a special case of our meta-regression model (4.11) with $G - 1$ dummy variables

$$\hat{\theta}_k = \theta + \beta_1 x_{1k} + \cdots + \beta_{G-1} x_{G-1k} + u_k + \sigma_k \epsilon_k \quad (4.12)$$

with $x_{gk} = 1$ if study k belongs to subgroup $g, g = 1, \ldots, G - 1$ and $x_{gk} = 0$ otherwise. Parameter θ corresponds to the treatment effect in subgroup G (baseline group) whereas parameters β_g describe the difference of the treatment effect in subgroup g from the baseline group. Accordingly, treatment effect in subgroup g is equal to $\theta + \beta_g, g = 1, \ldots, G - 1$.

Example 4.5 The `metareg` function in R package **meta** can be used to conduct a meta-regression. Actually, the `metareg` function is a wrapper function that calls the `rma.uni` function from R package **metafor** (`rma.uni` stands for <u>r</u>andom effects <u>m</u>eta-<u>a</u>nalysis with a <u>uni</u>variate outcome).

The first argument of the `metareg` function, i.e. an object of class `meta`, is mandatory. The second argument is a formula specifying the right side of Eq. (4.11). This argument is mandatory if the meta object was created without argument `byvar` otherwise it can be omitted.

We can use R object `mc3s.c` to conduct a meta-regression for study duration.

```
> mc3s.mr <- metareg(mc3s.c, duration)
Warning message:
In metafor::rma.uni(yi = x$TE, sei = x$seTE, data = dataset,  :
  Studies with NAs omitted from model fitting.
```

A warning is printed from the `rma.uni` function of R package **metafor** as two studies have zero standard errors. Note, as R object `mc3s.c` was created using argument `byvar` we could have omitted the second argument. The `metareg` command creates an R object of class `rma.uni`. Accordingly, the `print.rma.uni` function from R package **metafor** is used to print this object.

```
> print(mc3s.mr, digits=2)
Mixed-Effects Model (k = 21; tau^2 estimator: DL)

tau^2 (estimated amount of residual heterogeneity):    0.00
tau (square root of estimated tau^2 value):            0.05
I^2 (residual heterogeneity / unaccounted variability): 83.81%
H^2 (unaccounted variability / sampling variability):  6.18
R^2 (amount of heterogeneity accounted for):           11.92%

Test for Residual Heterogeneity:
QE(df = 19) = 117.35, p-val < .01

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 10.21, p-val < .01

Model Results:

                  estimate    se    zval   pval   ci.lb   ci.ub
intrcpt              -0.21   0.04   -4.94   <.01   -0.29   -0.12    ***
duration> 3 months    0.14   0.04    3.20   <.01    0.05    0.23    **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First, information on heterogeneity ($\hat{\tau}^2$, I^2 etc.) is printed. As we only wanted to print two decimal places, the estimated between-study variance is reported as zero. However, the square root of $\hat{\tau}^2$ is 0.05 which corresponds to a value of 0.0025 for the between-study variance which is very close to the common between-study variance estimate.[5]

Next comes a test for residual heterogeneity as well as a test of moderators. The later test corresponds to the test for subgroup differences in the random effects

---
[5]Actually, using command `round(mc3s.mr$tau2, 4)` would print the value 0.0024.

model with a common between-study variance. Reported *Q*-statistics as well as *p*-value are identical to the values calculated in Example 4.4: $Q = 10.21, p = 0.0014$.

Last, estimated treatment effects are reported under "Model Results". The estimated treatment effect in studies with study duration less than 3 months is reported in the line starting with "intrcpt". The estimated treatment effect (column "estimate") as well as lower and upper 95 % confidence limits (columns "ci.lb" and "ci.ub") are identical to values reported in Example 4.4. The estimated treatment effect in the subgroup with longer study duration can easily be calculated: $-0.21 + 0.14 = -0.07$. Again, this treatment estimate has been reported in Example 4.4.

In order to calculate 95 % confidence intervals for studies with longer study duration based on information from the meta-regression model, the variance–covariance matrix of the estimated coefficients has to be taken into account.

```
> # Variance-covariance matrix
> varcov <- vcov(mc3s.mr)
> # Estimated treatment effect in studies with longer duration
> TE.s2 <- sum(coef(mc3s.mr))
> # Standard error of treatment effect
> seTE.s2 <- sqrt(sum(diag(varcov)) + 2*varcov[1,2])
```

The quantities `TE.s2` and `seTE.s2` can be used as input to the `metagen` function to estimate a 95 % confidence interval.

```
> print(metagen(TE.s2, seTE.s2, sm="MD"), digits=2)
    MD         95%-CI      z  p-value
 -0.07 [-0.09; -0.04] -4.68 < 0.0001
```

The same values for lower and upper 95 % confidence interval have been reported in Example 4.4. □

Example 4.6 A subgroup analysis comparing trials with adequate blinding to trials with inadequate/unclear blinding was conducted for the Ketotifen meta-analysis with a binary outcome in Example 3.20. The subgroup analysis was based on separate estimates of the between-study variance. Here we redo the subgroup analysis using a common between-study variance accordingly results are slightly different from those reported in Example 3.20. It is assumed that R dataset `data9` is still available in the R session; otherwise we refer to Fig. 3.5 for R code to generate the R object.

```
> mb3s.c <- metabin(Ee, Ne, Ec, Nc, sm="RR", method="I",
+                   data=data9, studlab=study,
+                   byvar=blind, print.byvar=FALSE,
+                   tau.common=TRUE)
> print(summary(mb3s.c), digits=2)
Number of studies combined: k=10
*** Output truncated ***

Results for subgroups (random effects model):
                  k   RR       95%-CI     Q    tau^2   I^2
Adequate blinding 3 0.72 [0.43; 1.21]  2.49  0.1028 19.7%
Method unclear    7 0.56 [0.40; 0.78] 14.29  0.1028   58%
```

```
Test for subgroup differences (random effects model):
                    Q d.f.  p-value
Between groups  0.64    1   0.4245
Within groups  16.79    8   0.0324
*** Output truncated ***
```

Before having a closer look at this printout we conduct a meta-regression:

```
> mb3s.mr <- metareg(mb3s.c)
> print(mb3s.mr, digits=2)
Mixed-Effects Model (k = 10; tau^2 estimator: DL)
*** Output truncated ***
Test for Residual Heterogeneity:
QE(df = 8) = 16.79, p-val = 0.03

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 0.64, p-val = 0.42

Model Results:

                      estimate    se   zval  pval  ci.lb  ci.ub
intrcpt                  -0.33  0.27  -1.23  0.22  -0.85   0.19
.byvarMethod unclear     -0.25  0.32  -0.80  0.42  -0.87   0.37
*** Output truncated ***
```

The "Test of Moderators" in the meta-regression corresponds to the "Test for subgroup differences (Between groups)": $Q =0.64$, 1 degree of freedom. Furthermore, the "Test for Residual Heterogeneity" is identical to the "Within groups" results in the subgroup analysis: $Q =16.79$, 1 degree of freedom. Otherwise, results of the two analyses look at first glance differently. However, the subgroup analysis reports risk ratios whereas the meta-regression reports log risk ratios. Using the exponential function we get the same treatment estimates for studies with adequate blinding as well as studies with unclear method of blinding.

```
> # Treatment effect in studies with adequate blinding
> round(exp(coef(mb3s.mr)["intrcpt"]), 2)
intrcpt
   0.72
> # Treatment effect in studies with unclear method of blinding
> round(exp(sum(coef(mb3s.mr))), 2)
[1] 0.56
```

□

4.4.2 Meta-Regression with a Continuous Covariate

In this subsection we introduce meta-regression with a continuous covariate using a famous vaccination meta-analysis. The importance of centring a continuous covariate in a meta-regression is well illustrated by this example.

Example 4.7 Colditz et al. [4] evaluated the overall effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis. In addition, covariates that may potentially influence the effect of vaccination were examined. We will use one covariate, absolute geographical latitude (i.e. distance from the equator), as an example.

The BCH data with 13 studies is part of the R package **metafor** and can be easily loaded in R.

```
> data(dat.colditz1994, package="metafor")
> data10 <- dat.colditz1994
```

This R command creates an R object `dat.colditz1994` in the workspace which we assign to R object `data10`. A meta-analysis of the BCG data shows that substantial heterogeneity exists.

```
> mh2 <- metabin(tpos, tpos+tneg, cpos, cpos+cneg,
+                data=data10, studlab=paste(author, year))
> summary(mh2)
Number of studies combined: k=13
*** Output truncated ***
Test of heterogeneity:
      Q d.f.  p-value
 152.57   12 < 0.0001
*** Output truncated ***
```

Accordingly, the next step is to examine whether the absolute geographical latitude can explain the very large between-study heterogeneity at least to some extent. This covariate has nine different values, ranging from 13 to 55:

```
> table(data10$ablat)
13 18 19 27 33 42 44 52 55
 2  1  1  1  2  2  2  1  1
```

Given this, it is natural to conduct a meta-regression analysis with the continuous covariate `ablat`.

```
> mh2.mr <- metareg(mh2, ablat)
> print(mh2.mr, digits=2)
Mixed-Effects Model (k = 13; tau^2 estimator: DL)
*** Output truncated ***
Test for Residual Heterogeneity:
QE(df = 11) = 30.73, p-val < .01

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 18.85, p-val < .01

Model Results:

         estimate    se   zval  pval  ci.lb  ci.ub
intrcpt      0.26  0.23   1.12  0.26  -0.20   0.71
ablat       -0.03  0.01  -4.34  <.01  -0.04  -0.02  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The "Test for Residual Heterogeneity" and the "Test for Moderators" clearly show the large influence of the absolute geographical latitude on the effectiveness of the BCG vaccine. The residual heterogeneity decreases from $Q =152.57$ (with 12 degrees of freedom) to $Q =30.73$ (with 11 degrees of freedom). The effect of absolute geographical latitude is highly statistical significant: $Q =18.85$ (with 1 degree of freedom).

The estimated effect of BCG vaccination at the equator, i.e. at an absolute geographical latitude of 0, is reported in the line starting with "intrcpt". The influence of a 1° change in absolute geographical latitude on the efficacy of the BCG vaccine is given in the line starting with "ablat". These results are reported on the log scale. As we can see the effect of absolute geographical latitude on the efficacy of the BCG vaccine is negative, -0.03, which translates in a stronger reduction of positive tuberculosis cases with increasing distance from the equator.

A bubble plot can be generated using the `bubble.metareg` function.

```
> bubble(mh2.mr)
```

The resulting bubble plot is shown in Fig. 4.1. It is clear from this figure that the intercept, i.e. effect of BCG vaccination at the equator, involves extrapolation way beyond the range of the available data.

In situations like this, it is typically more sensible to centre the continuous covariate around its mean value and to conduct a meta-regression using this centered covariate. This we now do:

```
> mean(data10$ablat)
```
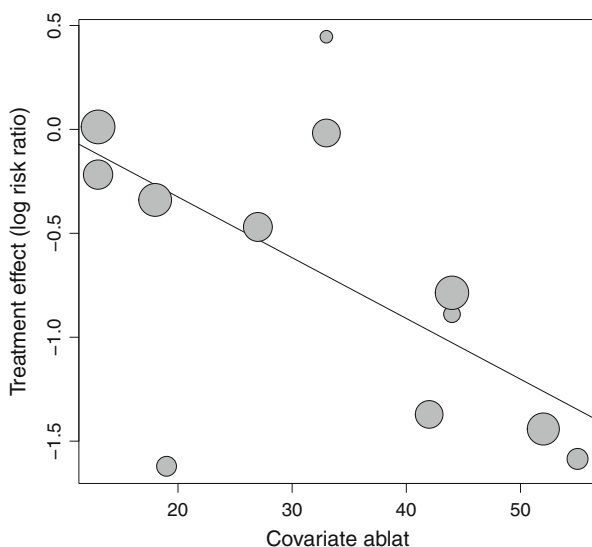

Fig. 4.1 *Bubble* plot for meta-regression of BCG data [4] with absolute geographical latitude (covariate `ablat`) as continuous covariate

```
[1] 33.46154
> ablat.c <- with(data10, ablat - mean(ablat))
> mh2.mr.c <- metareg(mh2, ablat.c)
> print(mh2.mr.c, digits=2)
Mixed-Effects Model (k = 13; tau^2 estimator: DL)
*** Output truncated ***
Test for Residual Heterogeneity:
QE(df = 11) = 30.73, p-val < .01

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 18.85, p-val < .01

Model Results:

         estimate     se   zval  pval  ci.lb  ci.ub
intrcpt     -0.72   0.10  -7.09  <.01  -0.92  -0.52  ***
ablat.c     -0.03   0.01  -4.34  <.01  -0.04  -0.02  ***
*** Output truncated ***
```

This meta-regression based on a centered continuous covariate has exactly the same results for the test of heterogeneity, test of moderators, effect of the absolute graphical latitude, and fitted values. However, the estimate of the intercept and its interpretation are different. Here, the intercept corresponds to the effect of BCG vaccination at an absolute geographical latitude of 33.5°. As this is a log risk ratio, we will backtransform this to a risk ratio.

```
> round(exp(coef(mh2.mr.c)["intrcpt"]), 2)
intrcpt
   0.49
```

We see that BCG vaccination at an absolute geographical latitude of 33.5° results in a reduction of the tuberculosis positive cases by about 50 %.

We can also use the `metagen` function to print the risk ratio and corresponding confidence interval.

```
> TE.33.5 <- coef(mh2.mr.c)["intrcpt"]
> seTE.33.5 <- sqrt(vcov(mh2.mr.c)["intrcpt", "intrcpt"])
> print(metagen(TE.33.5, seTE.33.5, sm="RR"), digits=2)
   RR      95%-CI      z  p-value
 0.49 [0.4; 0.59] -7.09 < 0.0001
*** Output truncated ***
```

These results are identical to the meta-regression results, e.g. the *z*-value is −7.09.   □

4.5 Summary

In this chapter various measures for heterogeneity used in meta-analysis have been introduced. Furthermore, statistical tests for subgroup differences based respectively on fixed effect and random effects models have been described. Last,

meta-regression—an extension of subgroup analysis—has been introduced and its connection to a subgroup analysis assuming a common between-study variance has been illustrated through examples. Our final example also illustrated the importance of centring a continuous covariate in a meta-regression.

References

1. Berlin, J.A., Santanna, J., Schmid, C.H., Szczech, L.A., Feldman, H.I.: Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat. Med. **21**, 371–387 (2002)
2. Borenstein, M., Hedges, L., Higgins, J., Rothstein, H.: Introduction to Meta Analysis. Wiley, Chichester (2009)
3. Cochran, W.G.: The combination of estimates from different experiments. Biometrics **10**, 101– 129 (1954)
4. Colditz, G.A., Brewer, T., Berkey, C., Wilson, M., Burdick, E., Fineberg, H., Mosteller, F.: Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. JAMA **271**(9), 698–702 (1994)
5. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Control. Clin. Trials **7**, 177–188 (1986)
6. Engels, E.A., Schmid, C.H., Terrin, N., Olkin, I., Lau, J.: Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat. Med. **19**, 1707– 1728 (2000)
7. Galbraith, R.F.: A note on graphical presentation of estimated odds ratios from several clinical trials. Stat. Med. **7**, 889–894 (1988)
8. Hardy, R.J., Thompson, S.G.: Detecting and describing heterogeneity in meta-analysis. Stat. Med. **17**, 841–856 (1998)
9. Higgins, J.P.T., Thompson, S.G.: Quantifying heterogeneity in a meta-analysis. Stat. Med. **21**, 1539–1558 (2002)
10. Higgins, J.P.T., Thompson, S.G.: Controlling the risk of spurious findings from meta- regression. Stat. Med. **23**, 1663–1682 (2004)
11. Knapp, G., Biggerstaff, B.J., Hartung, J.: Assessing the amount of heterogeneity in random- effects meta-analysis. Biom. J. **48**, 271–285 (2006)
12. Mittlböck, M., Heinzl, H.: A simulation study comparing properties of heterogeneity measures in meta-analyses. Stat. Med. **25**, 4321–4333 (2006)
13. 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). doi:10.1186/1471-2288- 8-79. http://www.biomedcentral.com/1471-2288/8/79
14. Senn, S.: The many modes of meta. Drug Inf. J. **34**, 535–549 (2000)
15. Sidik, K., Jonkman, J.N.: Simple heterogeneity variance estimation for meta-analysis. J. R. Stat. Soc. Ser. C **54**(2), 367–384 (2005)
16. The Cochrane Collaboration: Review Manager (RevMan) [Computer program]. Version 5.3. The Nordic Cochrane Centre, Copenhagen (2014)
17. Thompson, S.G., Higgins, J.P.T.: How should meta-regression analyses be undertaken and interpreted? Stat. Med. **21**, 1559–1573 (2002)
18. Thompson, S.G., Sharp, S.J.: Explaining heterogeneity in meta-analysis: a comparison of methods. Stat. Med. **18**, 2693–2708 (1999)
19. Viechtbauer, W.: Confidence intervals for the amount of heterogeneity in meta-analysis. Stat. Med. **26**, 37–52 (2007)
20. Viechtbauer, W.: Conducting meta-analyses in R with the metafor package. J. Stat. Softw. **36**(3), 1–48 (2010)

Part III
Advanced Topics

Chapter 5
Small-Study Effects in Meta-Analysis

This chapter describes small-study effects in meta-analysis and how the issues they raise may be addressed. "Small-study effects" is a generic term for the phenomenon that smaller studies sometimes show different, often larger, treatment effects than large ones. This notion was coined by Sterne et al. [55]. One possible, probably the most well-known, reason is publication bias. This is said to occur when the chance of a smaller study being published is increased if it shows a stronger effect [3, 41, 52]. This can happen for a number of reasons, for example authors may be more likely to submit studies with "significant" results for publication or journals may be more likely to publish smaller studies if they have "significant" results. If this occurs, it in turn biases the results of meta-analyses and systematic reviews. There are a number of other possible reasons for small-study effects. One is selective reporting of the most favourable outcomes, known as outcome selection bias or outcome reporting bias [8, 9, 18, 61]. Another possible cause of small-study effects is clinical heterogeneity between patients in large and small studies; e.g., patients in smaller studies may have been selected so that a favourable outcome of the experimental treatment may be expected. In the case of a binary outcome, also a mathematical artefact arises from the fact that for the odds ratio or the risk ratio, the variance of the treatment effect estimate is not independent of the estimate itself [47]. This problem will be discussed in Sect. 5.2.2. Lastly, it can never be ruled out that small-study effects result from mere coincidence [42]. Empirical studies have established evidence for these and other kinds of bias [19, 42, 53]. There is a vast range of tests for small-study effects [4, 20, 24, 38, 43, 48], most of them based on a funnel plot which will be introduced in Sect. 5.1.1.

5.1   Graphical Illustration of Small-Study Effects

Typically, the first step in exploring possible small-study effects in meta-analysis is to look at a graphical presentation of the data. The most common approach is the funnel plot, which we introduce using the following example.

Example 5.1 Moore et al. [34] conducted a systematic review of 37 randomized placebo-controlled trials on the effectiveness and safety of topical non-steroidal anti-inflammatory drugs (NSAIDS) in acute pain. The main outcome was treatment success, defined as a reduction in pain of at least 50 %. R code to read in the data and print some information is shown in Fig. 5.1.

A forest plot of the NSAIDS studies using the odds ratio as treatment effect measure and the Mantel–Haenszel method for pooling is given in Fig. 5.2. This figure was created using the following R commands.

```
> ms1 <- metabin(Ee, Ne, Ec, Nc, data=data11, sm="OR")
> forest(ms1,
+        label.left="NSAIDS worse", label.right="NSAIDS better",
+        ff.lr="bold")
```

Figure 5.2 shows that both fixed effect and random effects estimates are highly significant. However, the estimated treatment effects of these two models are rather different. The estimated odds ratio is smaller than 1 (i.e. showing a detrimental effect of NSAIDS) in only a single study. In some studies estimated odds ratios are rather extreme with values up to 115.

The large difference in estimates of the fixed effect and random effects model acts as a warning of possible small-study effects. This is because the random effects model assigns more weight to smaller studies than the fixed effect model. Accordingly, the estimated treatment effect of the random effects model is shifted in direction of the treatment effects of smaller studies if a small-study effect is present.   □

```
> # 1. Read the data
> data11 <- read.csv("dataset11.csv")
> # 2. Print structure of R object data11
> str(data11)
'data.frame': 37 obs. of  5 variables:
 $ study: int  1 2 3 4 5 6 7 8 9 10 ...
 $ Ee   : int  23 69 79 35 10 23 20 35 56 28 ...
 $ Ne   : int  30 123 102 49 15 50 32 50 84 40 ...
 $ Ec   : int  10 54 45 13 2 13 9 40 33 16 ...
 $ Nc   : int  30 116 102 42 15 50 24 50 84 40 ...
> # 3. Calculate experimental and control event probabilities
> summary(data11$Ee/data11$Ne)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4200  0.6250  0.7203  0.7089  0.8276  1.0000
> summary(data11$Ec/data11$Nc)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0000  0.2750  0.4000  0.3824  0.4655  0.8000
```

Fig. 5.1 Data from meta-analysis on NSAIDS in acute pain [34]; for details on variable names, see Table 3.2

Study	Experimental Events Total		Control Events Total		Odds Ratio	OR	95%–CI	W(fixed)	W(random)
1	23	30	10	30		6.57	[2.11; 20.48]	1.0%	2.7%
2	69	123	54	116		1.47	[0.88; 2.44]	11.0%	4.0%
3	79	102	45	102		4.35	[2.37; 7.98]	4.6%	3.8%
4	35	49	13	42		5.58	[2.26; 13.73]	1.8%	3.2%
5	10	15	2	15		13.00	[2.07; 81.48]	0.3%	1.6%
6	23	50	13	50		2.42	[1.04; 5.63]	3.2%	3.3%
7	20	32	9	24		2.78	[0.93; 8.29]	1.7%	2.7%
8	35	50	40	50		0.58	[0.23; 1.46]	5.4%	3.1%
9	56	84	33	84		3.09	[1.65; 5.81]	4.9%	3.7%
10	28	40	16	40		3.50	[1.39; 8.84]	2.2%	3.1%
11	10	23	3	24		5.38	[1.25; 23.28]	0.7%	2.1%
12	44	92	40	89		1.12	[0.63; 2.01]	9.5%	3.8%
13	21	50	19	50		1.18	[0.53; 2.63]	5.0%	3.4%
14	17	20	3	20		32.11	[5.66; 182.18]	0.2%	1.7%
15	34	42	11	40		11.20	[3.97; 31.60]	1.0%	2.9%
16	25	30	13	30		6.54	[1.97; 21.74]	1.0%	2.5%
17	38	60	27	60		2.11	[1.02; 4.39]	4.4%	3.5%
18	85	118	46	113		3.75	[2.16; 6.50]	5.9%	3.9%
19	53	64	52	66		1.30	[0.54; 3.12]	4.0%	3.2%
20	16	20	6	20		9.33	[2.18; 39.96]	0.5%	2.1%
21	18	30	5	30		7.50	[2.24; 25.06]	0.9%	2.5%
22	6	12	0	12		25.00	[1.21; 516.69]	0.1%	0.7%
23	26	32	13	28		5.00	[1.57; 15.91]	1.2%	2.6%
24	41	84	33	84		1.47	[0.80; 2.72]	7.6%	3.8%
25	33	46	13	42		5.66	[2.26; 14.16]	1.7%	3.1%
26	30	38	22	36		2.39	[0.85; 6.67]	2.1%	2.9%
27	18	20	13	22		6.23	[1.15; 33.77]	0.6%	1.8%
28	22	26	14	26		4.71	[1.27; 17.56]	1.0%	2.3%
29	18	21	13	22		4.15	[0.94; 18.41]	0.8%	2.0%
30	39	51	9	47		13.72	[5.19; 36.30]	1.0%	3.0%
31	30	30	19	30		35.97	[2.00; 645.93]	0.1%	0.8%
32	11	13	0	12		115.00	[4.98; 2657.48]	0.0%	0.7%
33	44	90	40	89		1.17	[0.65; 2.11]	9.2%	3.8%
34	67	83	38	86		5.29	[2.65; 10.56]	3.2%	3.6%
35	24	29	14	27		4.46	[1.31; 15.16]	1.1%	2.5%
36	7	10	1	10		21.00	[1.78; 248.10]	0.1%	1.0%
37	21	25	12	25		5.69	[1.51; 21.42]	0.9%	2.3%
Fixed effect model	**1734**		**1693**			2.98	[2.59; 3.44]	100%	––
Random effects model						3.73	[2.80; 4.97]	––	100%

Heterogeneity: I–squared=68.3%, tau–squared=0.467, p<0.0001

0   0.1   1   10   1000
NSAIDS worse   NSAIDS better

Fig. 5.2 Forest plot for meta-analysis on NSAIDS in acute pain [34]

5.1.1 Funnel Plot

A funnel plot [31] shows the estimated treatment effects on a suitable scale (usually on the *x*-axis) against a measure of their precision, usually the standard error, on the *y*-axis with standard error at top, i.e. an inverted axis [54]. For the odds ratio (3.2) the coordinates x_k and y_k of the funnel plot are defined as

$$x_k = \log \hat{\psi}_k$$

$$y_k = \text{S.E.} (\log \hat{\psi}_k).$$

Less frequently, the *y*-axis shows the inverse standard error, the inverse variance or a function of sample size instead of the standard error.

In the absence of small-study effects, treatment effects of large and small studies would scatter around a common average treatment effect. If no excessive
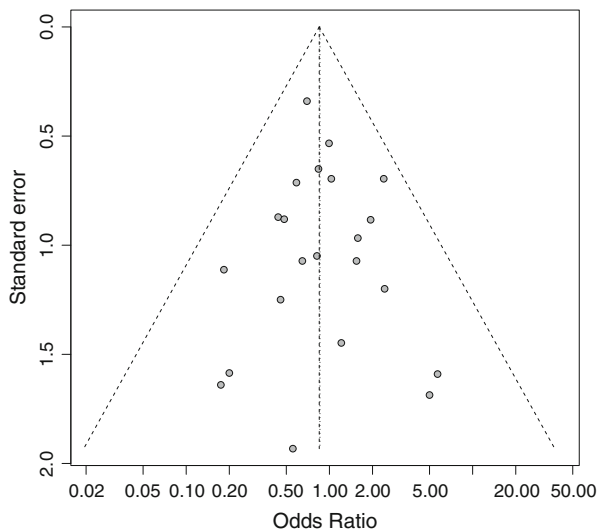
Fig. 5.3 Funnel plot for fictional meta-analysis of 25 studies without small-study effect

between-study heterogeneity exists, smaller studies (with larger standard errors) would scatter more than larger studies. That is, the funnel plot would show the form of a triangle symmetric with respect to the average treatment effect, with broad variability for small imprecise studies (at the bottom of the plot) and small dispersion for large, precise studies (at the top). Figure 5.3 shows such an example with 25 fictional studies generated from a fixed effect model with an odds ratio of 0.8.[1]

If the funnel plot appears asymmetric, this may be due to small-study effects. In the presence of selection we expect small studies which have "wrong" (that is, unfavourable) treatment effects are more likely to be missing. This results in a biased estimate if the treatment effects are pooled in a meta-analysis.

Example 5.2 The necessary information to produce a funnel plot for the NSAIDS meta-analysis is available in R object ms1, i.e. treatment estimates ms1$TE and standard errors ms1$seTE of individual studies. The following R code could be used to produce crude funnel plots (figures not shown).

```
> # Funnel plot with log odds ratio values on x-axis
> plot(ms1$TE, ms1$seTE, ylim=c(max(ms1$seTE), 0))
> # Funnel plot with odds ratio values on x-axis and
> # descriptive axis labels
> plot(exp(ms1$TE), ms1$seTE, ylim=c(max(ms1$seTE), 0),
+      log="x", xlab="Odds Ratio", ylab="Standard error")
```

Note, the argument ylim is necessary in the R commands to invert the *y*-axis.

---

[1] R code to generate the funnel plot is given in the web-appendix.
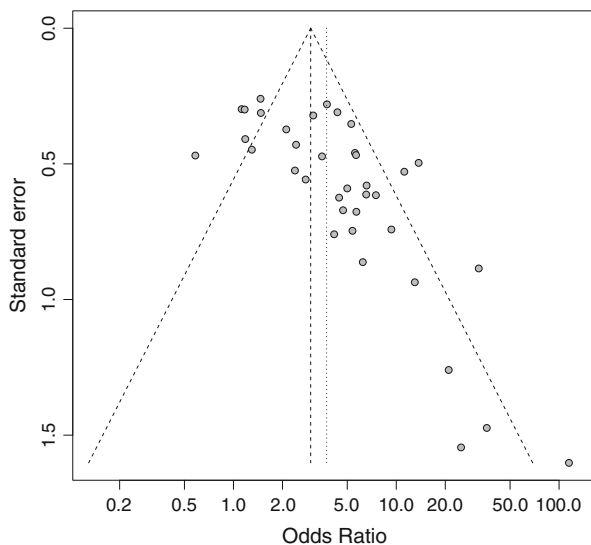
Fig. 5.4 Funnel plot for meta-analysis on NSAIDS in acute pain [34]

A more convenient way to generate a funnel plot is the use of the `funnel` function in R package **meta**.

```
> funnel(ms1)
```

The resulting funnel plot is shown in Fig. 5.4. Filled circles represent estimated treatment effect (odds ratio) and its precision (standard error) for each individual study. In addition to individual study results, the fixed effect estimate (vertical dashed line) with 95 % confidence interval limits (diagonal dashed lines) and the random effects estimate (vertical dotted line) are shown in the figure. Since the width of the confidence interval is proportional to the standard error, the choice of the standard error as *y*-axis is preferable [54]. This leads to the confidence interval limits becoming straight lines making the funnel plot ideally look like a triangle, i.e. if no small-study effects and no excessive between-study heterogeneity exists.

In Fig. 5.4 we observe a large gap in the bottom left corner of the plot, indicating that we may be missing smaller trials with odds ratios around one, corresponding to indecisive or even unfavourable results. This example of funnel plot asymmetry will be examined in detail in the following.   ☐

Details on `funnel.meta` Function

The `funnel` function is a generic function. Accordingly, for objects of class `meta` the `funnel.meta` function is utilised. This function has numerous arguments,

allowing fine control of the resulting plot. They can be printed using the `args` function.

```
> args(funnel.meta)
function (x, xlim = NULL, ylim = NULL, xlab = NULL,
    ylab = NULL, comb.fixed = x$comb.fixed,
*** Output truncated ***
    col = "black", bg = "darkgray", col.fixed = "black",
    col.random = "black", log = "", yaxis = "se",
    contour.levels = NULL, col.contour,
*** Output truncated ***
```

The key argument of the `funnel.meta` function is `x` which is an object of class `meta`, such as R object `ms1` from the NSAIDS example. The other arguments give fine control of the graphical output. For example, `xlab` and `ylab` allows the user to specify the axis labels, `xlim` and `ylim` the axis ranges, and `col`, `col.fixed`, `col.random` and `col.contour` for colours of various parts of the plot. Alternative choices for the *y*-axis are possible using `yaxis="invvar"` (inverse of the variance), `yaxis="invse"` (inverse of the standard error), and `yaxis="size"` (study size).

As usual, a detailed explanation of the various options can be found by using the help command `help(funnel.meta)` or `?funnel.meta`.

Contour-Enhanced Funnel Plot

Contour-enhanced funnel plots have been proposed to help differentiate between asymmetry due to publication bias and that due to other reasons [40]. Contour lines representing well established levels of statistical significance are added to a funnel plot to indicate regions where a test of treatment effect is significant.

A contour-enhanced funnel plot can be generated with the `funnel.meta` function by specifying the contour levels (argument `contour.levels`. By default (argument `col.contour` missing), suitable grey levels will be used to distinguish the levels of statistical significance.

Example 5.3 For the NSAIDS meta-analysis a contour-enhanced funnel plot is shown in Fig. 5.5 which was generated using the following commands.

```
> funnel(ms1, comb.random=FALSE, pch=16,
+        contour=c(0.9, 0.95, 0.99),
+        col.contour=c("darkgray", "gray","lightgray"))
> legend(0.25, 1.25,
+        c("0.1 > p > 0.05", "0.05 > p > 0.01", "< 0.01"),
+        fill=c("darkgray", "gray","lightgray"), bty="n")
```

Figure 5.5 clearly shows that almost all smaller studies have a statistically significant result, favouring NSAIDS, either at the 5 % level (medium grey region) or even the 1 % level (light grey region). On the other hand, some of the larger studies (i.e. standard error smaller than 0.5) show non-significant results (white region). In summary, the NSAIDS meta-analysis is a clear example for asymmetry, probably due to publication bias.   □
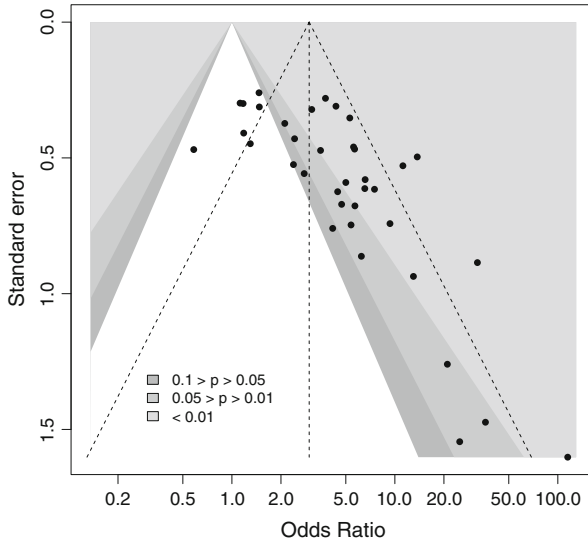
Fig. 5.5 Contour-enhanced funnel plot for meta-analysis on NSAIDS in acute pain [34]

5.1.2   Radial Plot

Galbraith [22] introduced a graphical method in order to display point estimates with different standard errors—so-called Galbraith plot or radial plot. An additional paper [21] is focused on medical applications and the use of the log odds ratio as measure of the treatment effect. If the number of estimates is large, a radial plot has certain advantages as compared to other displays like a forest plot which is commonly used for summarising the result of a meta-analysis.

For the odds ratio, a scatter plot of

$$x_k = 1/\text{S.E.} \, (\log \hat{\psi}_k)$$

and

$$y_k = \log \hat{\psi}_k / \text{S.E.} \, (\log \hat{\psi}_k)$$

is called a radial plot and has the following properties under a fixed effect model [22]:

(a) $\text{Var} \, (y_k)$ is equal to 1,
(b) for each study k, the estimated log odds ratio, $\log \hat{\psi}_k$, is equal to the slope of a line through the origin (0,0) and the point (x_k, y_k),

(c) points are close to zero on the x-axis for large S.E. ($\log \hat{\psi}_k$),

(d) logarithm of pooled odds ratio using the inverse variance method, $\hat{\theta}_F$, see (3.21), equals the slope parameter β_1 in a unweighted linear regression going through the origin: $y_k = \beta_1 x_k$.

Due to properties (b) and (d) a radial scale is sometimes shown on the right-hand side of a radial plot to indicate the treatment effects. The values y_k are often called z-scores because they correspond to the statistic of a test that $\log \psi_k$ is different from zero.

If there are no small-study effects, individual study results are expected to scatter randomly around the regression line through the origin corresponding to the fixed effect estimate (see property (d) above). Details of these considerations are discussed in the literature [11, 21, 47]. Egger's test [20], introduced in Sect. 5.2.1, is based on the radial plot.

Example 5.4 Base R code could be used to produce a crude radial plot (figure not shown).

```
> plot(1/ms1$seTE, ms1$TE/ms1$seTE)
```

The preferred way to generate a radial plot is the use of the `radial` function in R package **meta**.

```
> radial(ms1)
```

The resulting radial plot is shown in Fig. 5.6. It is obvious that results for smaller studies (inverse of standard error less than 2) do not scatter randomly around the regression line. Again, this is a clear indication of small-study effects. □
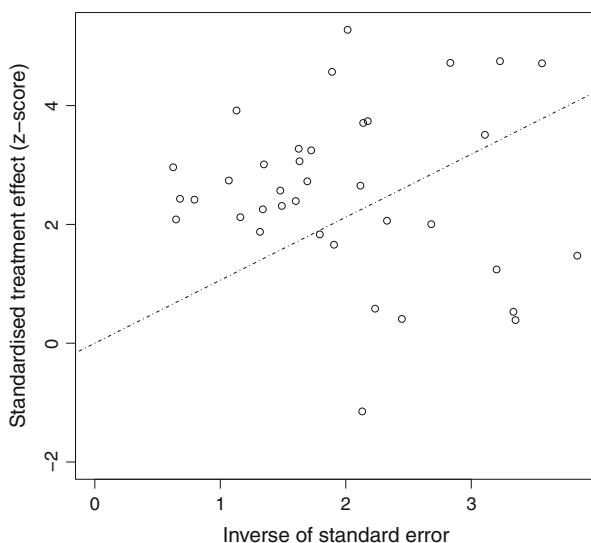
Fig. 5.6 Radial plot for meta-analysis on NSAIDS in acute pain [34]

5.2   Statistical Tests for Small-Study Effects

A large range of tests has been designed to examine whether asymmetry exists in a funnel plot. These tests are often called *tests for publication bias*. However, tests for asymmetry are not specific to publication bias and therefore it is preferred to call them tests for small-study effects or tests for funnel plot asymmetry. Regardless of publication bias, it is useful to quantify the evidence for funnel plot asymmetry, though this may be due to other sources than publication bias [27, 30, 56, 58].

Tests for small-study effects can be categorised into:

1. non-parametric tests using rank-correlation methods, introduced and influenced by Begg and Mazumdar [4, 48],
2. regression tests, represented by the so-called Egger's test [20] and its modifications [24, 32, 38, 43, 59].

A further class of methods, not be considered in detail, are models of estimating the number of missing studies, such as Rosenthal's "fail-safe *N*" [23, 37, 41]. Likewise, we do not consider selection models based on *p*-values [60] or tests based on an excess of significant findings [26].

5.2.1   Classical Tests by Begg and Egger

These tests assume that under the null hypothesis of no small-study effects, among studies included in a meta-analysis there is no association between effect size and precision.

Begg and Mazumdar Test: Rank Correlation Test

An adjusted rank correlation test was proposed by Begg and Mazumdar [4] to test for publication bias in meta-analysis; the power of the test was evaluated via simulations assuming a normal distribution for the estimated treatment effect. The test is based on the correlation between a standardised treatment effect and the within-trial variance; the inverse variance method is used in the construction of the test. For the odds ratio as measure of treatment effect, the corresponding quantities are the variance of the log odds ratio $\widehat{\mathrm{Var}}\,(\hat{\theta}_k)$ and a standardised treatment effect

$$s(\hat{\theta}_k) = \left(\hat{\theta}_k - \hat{\theta}_F\right) \Big/ \sqrt{\widehat{\mathrm{Var}}\,(\hat{\theta}_k) - \widehat{\mathrm{Var}}\,(\hat{\theta}_F)}\ .$$

Begg and Mazumdar [4] utilised Kendall's tau as correlation measure, however, other rank correlation coefficients like Spearman's rho which is probably better known might be used alternatively.

Let x denote the number of pairs of studies with standardised effects and variances ranked in the same order, i.e.

$$\left(s(\hat{\theta}_k) > s(\hat{\theta}_{k'}) \text{ and } \widehat{\text{Var}}\,(\log \hat{\theta}_k) > \widehat{\text{Var}}\,(\log \hat{\theta}_{k'}) \right)$$

or

$$\left(s(\hat{\theta}_k) < s(\hat{\theta}_{k'}) \text{ and } \widehat{\text{Var}}\,(\log \hat{\theta}_k) < \widehat{\text{Var}}\,(\log \hat{\theta}_{k'}) \right),$$

for $k \neq k'$. The number of pairs ranked in the opposite order are denoted by y. The normalised test statistic for the case that no ties neither within $s(\hat{\theta}_k)$ nor $\widehat{\text{Var}}\,(\log \hat{\theta}_k)$ are present is

$$z = \frac{x - y}{\sqrt{K(K-1)(2K+5)/18}}, \tag{5.1}$$

where K is the number of studies involved in the meta-analysis.

In the case of ties, i.e. $s(\hat{\theta}_k) = s(\hat{\theta}_{k'})$ or $\widehat{\text{Var}}\,(\log \hat{\theta}_k) = \widehat{\text{Var}}\,(\log \hat{\theta}_{k'})$, the standard error given in the denominator in Eq. (5.1) requires some modification; a modified version for tied observations can be found in Kendall and Gibbons [29, p. 66].

The test statistic z is asymptotically distributed according to a standard normal distribution under the null hypothesis of no bias in meta-analysis. The null hypothesis of no bias in meta-analysis is rejected at the significance level α if

$$|z| > z_{1-\alpha/2},$$

with $z_{1-\alpha/2}$ denoting the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Alternatively to conducting a statistical test, an estimate of Kendall's tau with $(1 - \alpha)$ confidence interval could be reported in a systematic review. Simulations showed that the power of the Begg and Mazumdar test is poor.

Example 5.5 Sufficient information to conduct the Begg and Mazumdar test is available in the R object `ms1`: treatment estimate `ms1$TE` and its standard error `ms1$seTE` (i.e. square-root of variance) as well as fixed effect estimate `ms1$TE.fixed` and its standard error `ms1$seTE.fixed`. We do not provide base R code to conduct the Begg and Mazumdar test, but use the `metabias` function of the R package **meta** instead.

```
> metabias(ms1, method="rank")

Rank correlation test of funnel plot asymmetry

data:  ms1
z = 3.5836, p-value = 0.0003389
alternative hypothesis: asymmetry in funnel plot
sample estimates:
       ks       se.ks
274.00000   76.45914
```

At first a test statistic is given ($z = 3.5836$ for the Begg and Mazumdar test), followed by the *p*-value. For the NSAIDS meta-analysis, the Begg and Mazumdar test shows a significant *p*-value ($p = 0.0003$) leading to rejection of the null hypothesis of symmetry in the funnel plot and decision in favour of the alternative hypothesis, indicating marked asymmetry of the funnel plot. At last the numerator `ks` and denominator `se.ks` of the test statistic `z` are provided.   □

Egger's Test: Linear Regression Test

The test proposed by Egger et al. [20] for the detection of publication bias in meta-analyses is strongly connected to a radial plot. The test is based on a simple linear regression including a parameter for the intercept β_0:

$$y_k = \beta_0 + \beta_1 x_k,$$

with

$$x_k = 1/\text{S.E.}\,(\hat{\theta}_k)$$

and

$$y_k = \hat{\theta}_k/\text{S.E.}\,(\hat{\theta}_k).$$

In contrast to the radial plot, the regression line is not constrained to run through the origin. In fact, the test is constructed by testing for a non-zero intercept β_0 which is asymptotically distributed according to Student's *t*-distribution with $K - 2$ degrees of freedom under the null hypothesis of no bias in meta-analysis.

The approach is justified by the intuitive argument that, in the presence of publication bias, small studies with non-significant or negative results are less likely to get published. Thus, points close to zero on the x-axis do not scatter randomly around the overall effect resulting in a non-zero intercept which is a departure from property (d) of a radial plot described in Sect. 5.1.2.

The null hypothesis of no bias in meta-analysis is rejected at the significance level α if

$$\left| \hat{\beta}_0/\text{S.E.}\,(\hat{\beta}_0) \right| = |t| > t_{K-2;1-\alpha/2}$$

with $t_{K-2;1-\alpha/2}$ denoting the $(1-\alpha/2)$ quantile of Student's *t*-distribution with $K-2$ degrees of freedom. The test procedure is implicitly based on the assumption that linearity still holds in the presence of bias.

Example 5.6 The `metabias` function can be used to conduct Egger's test.

```
> metabias(ms1, method="linreg")

Linear regression test of funnel plot asymmetry

data:  ms1
t = 4.7147, df = 35, p-value = 3.786e-05
alternative hypothesis: asymmetry in funnel plot
sample estimates:
     bias    se.bias       slope
 2.7652744  0.5865197  -0.1122134
```

At first the test statistic is given ($t = 4.7147$ with $K - 2 = 35$ degrees of freedom, `df=35`), followed by the p-value. For the NSAIDS meta-analysis, Egger's test shows a significant p-value ($p < 0.0001$) leading to rejection of the null hypothesis of symmetry in the funnel plot. At last the numerator `bias` and denominator `se.bias` of test statistic `t` are provided which correspond to the intercept of a regression line not restricted to go through the origin. In addition, the slope of the regression line is printed. These values can be compared to the result of a corresponding linear regression model using base R commands.

```
> reg <- lm(I(ms1$TE/ms1$seTE) ~ I(1/ms1$seTE))
> summary(reg)
*** Output truncated ***
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.7653     0.5865    4.715 3.79e-05 ***
I(1/ms1$seTE) -0.1122     0.2707   -0.415    0.681
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
*** Output truncated ***
```

Here, the generic `lm` function is used to build a linear model with dependent variable `ms1$TE/ms1$seTE` and independent variable `1/ms1$seTE`. These variables were framed using the `I` function in order to insure that the "/" operator is used as an arithmetical operator. Intercept and its standard error are identical to those given in the output of `metabias` function. In addition, slope estimate and its standard error are reported which are of no importance for the test of funnel plot asymmetry.

The regression line with an intercept can be printed to a radial plot in two different ways.

```
> radial(ms1)
> abline(reg)
```

Or using the `plotit` argument of the `metabias` function.

```
> metabias(ms1, method = "linreg", plotit=TRUE)
```

The resulting plot using the `radial` function and `abline` function is shown in Fig. 5.7. In the `metabias` function with `plotit` argument the regression line for the fixed effect model (dashed line) is not plotted.
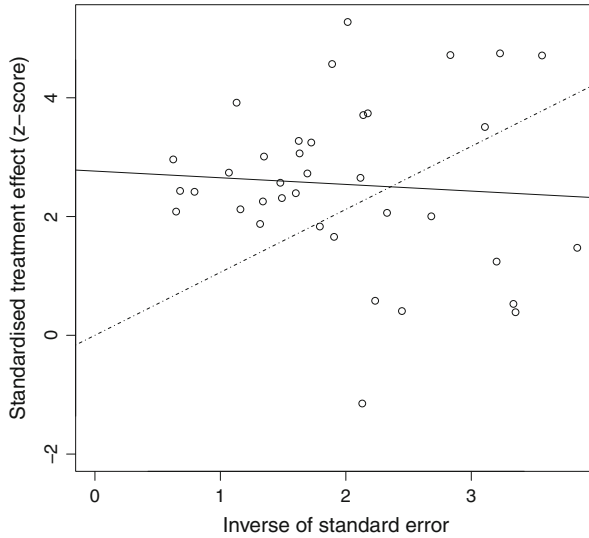
Fig. 5.7 Radial plot for meta-analysis on NSAIDS in acute pain [34] including regression line (*solid line*) for Egger's test [20]

Note that the regression line (solid line) differs markedly from the line through the origin (dashed line) in Fig. 5.7. This is due to the asymmetry in the funnel plot.   □

Test by Thompson and Sharp

A variant of Eggers's test allowing for between-study heterogeneity was proposed by Thompson and Sharp [59, method (3a)]. This test statistic is based on a weighted linear regression of the treatment effect on its standard error using the method of moments estimator for the additive between-study variance component. The test statistic follows a *t*-distribution with $K - 2$ degrees of freedom.

Example 5.7 Again, the `metabias` function can be used to conduct the Test by Thompson and Sharp.

```
> metabias(ms1, method = "mm")

Linear regression test of funnel plot asymmetry (methods
of moment)

data:  ms1
t = 5.1726, df = 35, p-value = 9.518e-06
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias        slope
 2.72761281  0.52732226  -0.09076304
```

This test also shows a significant p-value ($p < 0.0001$), leading to rejection of the null hypothesis of symmetry in the funnel plot.   □

5.2.2 Modified Versions of Classical Tests for Binary Outcomes

The assumption of funnel plot symmetry in the absence of small-study effects is plausible when the outcome is continuous, as under the assumption of normality the sample mean is statistically independent of the sample variance. However, this is not generally true for discrete data, a fact mentioned and discussed by several authors [5, 24, 32, 38, 47, 48, 59]. Specifically, suppose the outcome is binary and the effect is summarised by the log risk ratio or log odds ratio. Then the variance estimators of both the log risk ratio and log odds ratio are statistically dependent on the estimated log risk ratio and log odds ratio. Even in absence of small-study effects, this dependence induces some asymmetry in the funnel plot. Accordingly, this mathematical artefact is another source of asymmetry in funnel plots for binary outcomes. When tests designed for quantitative outcomes are applied in meta-analyses with binary outcomes, they are anti-conservative, potentially reporting significant p-values for small-study effects more often than they should [32, 47]. This observation has motivated proposals to modify existing tests for binary outcomes [24, 38, 43, 48].

Harbord's Test: Score-Based Test

For binary data with the odds ratio as effect measure, Harbord et al. [24] proposed a modification. They replaced the log odds ratio of study k, $k = 1, \ldots, K$, and its variance with the score-based effect measure Z_k/V_k (which is an approximation to the log odds ratio, if this is not too far from zero) and its variance $1/V_k$, where

$$Z_k = a_k - (a_k + b_k)(a_k + c_k)/n_k$$

and

$$V_k = (a_k + b_k)(a_k + c_k)(b_k + d_k)(c_k + d_k)/[n_k^2(n_k - 1)]$$

(notation see Table 3.1). Z_k/V_k (with variance $1/V_k$) is seldom used as a measure of outcome for a single study, but well-known as Peto odds ratio in meta-analysis, see Sects. 3.2.1 and 3.3.3.

Following Egger et al. [20], the treatment effect Z_k/V_k can be regressed on the standard error $1/\sqrt{V_k}$ with study weights V_k, testing the null hypothesis of a zero slope. The equivalent on the radial plot is to regress $Z_k/\sqrt{V_k}$ on $\sqrt{V_k}$ without weighting, and to test for a zero intercept.

The advantage of this test is that the variance estimate depends only on the marginal totals in Table 3.1. The superiority of this test for binary data compared to Egger's test was confirmed in simulations [24, 43].

Example 5.8 The `metabias` function can be used to conduct Harbord's test.

```
> metabias(ms1, method = "score")

Linear regression test of funnel plot asymmetry
(efficient score)

data:  ms1
t = 3.7618, df = 35, p-value = 0.0006181
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias        slope
 2.87991968   0.76556515  -0.07549869
```

Harbord's test shows a significant *p*-value ($p = 0.0006$), leading to rejection of the null hypothesis of symmetry in the funnel plot.   □

Macaskill's Test and Peters' Test

Peters et al. [38] proposed a modification of Macaskill's test [32] with the sample size serving as measure of precision. For both tests, the usual estimate of the log odds ratio, $\log \hat{\psi}_k = \log((a_k d_k)/(b_k c_k))$ is used. Macaskill et al. [32] proposed to regress this on the total study size n_k, weighting the study with $[1/(a_k+c_k)+1/(b_k+d_k)]^{-1}$. Peters et al. [38] proposed to regress $\log \hat{\psi}_k$ on the inverse of the total study size, $1/n_k$, using the same weights. Simulations showed superiority of the Peters' test compared to Egger's test and Macaskill's test [38].

Example 5.9 The `metabias` function can be used to conduct Peters' test.

```
> metabias(ms1, method = "peters")

Linear regression test of funnel plot asymmetry (based
on sample size)

data:  ms1
t = 4.7796, df = 35, p-value = 3.116e-05
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias        slope
74.9937870  15.6903262   0.4096863
```

Peters' test also shows a significant *p*-value ($p < 0.0001$), leading to rejection of the null hypothesis of symmetry in the funnel plot.

The test by Macaskill is not implemented in the `metabias` function. However, it could be conducted using the generic `lm` function.   □

Schwarzer's Test

Schwarzer et al. [48] described a rank test of the correlation between a standardised form of $(a_k - \mathrm{E}[A_k])$ and variance $\mathrm{Var}\,(A_k)$, where expectation $\mathrm{E}[A_k]$ and variance $\mathrm{Var}\,(A_k)$ of the random variable A_k denoting the number of events in the treatment arm (where a_k is observed) are estimated under the non-central hypergeometric distribution for each 2×2 table, given the overall fixed effect estimate of treatment on the log odds ratio scale.

Example 5.10 Again, the `metabias` function can be used to conduct Schwarzer's test.

```
> metabias(ms1, method = "count")

Rank correlation test of funnel plot asymmetry (based on
counts)

data:  ms1
z = 3.0866, p-value = 0.002024
alternative hypothesis: asymmetry in funnel plot
sample estimates:
       ks      se.ks
236.00000  76.45914
```

This test also shows a significant *p*-value ($p = 0.002$), leading to rejection of the null hypothesis of symmetry in the funnel plot.   □

Rücker's Tests: Tests Based on Arcsine Difference

Rücker et al. [43] proposed three versions of a test based on the arcsine difference as effect measure. The function $x \mapsto \arcsin \sqrt{x}$ is the variance-stabilising transform for the binomial distribution [15, 33], see also [1, 2, 28, 44]. The arcsine regression test is defined by choosing the arcsine difference Δ_k (3.12) as effect measure. Following Egger et al. [20], the observed treatment effect $\hat{\Delta}_k$ (3.13) is regressed on its standard error, which is estimated by $\sqrt{0.25/n_{ek} + 0.25/n_{ck}}$ with study weights $(0.25/n_{ek} + 0.25/n_{ck})^{-1}$, and the null hypothesis of zero slope is tested. Thus the variance estimate depends only on the row totals, that is, on the sample sizes. It does not depend on the treatment effect estimate.

Instead of Egger's test, other tests can be applied to the arcsine difference, such as the Begg and Mazumdar test. In a simulation study comparing all these tests it was shown that Harbord's test, Peters' test and the Thompson and Sharp version of Rücker's test maintain the type I error better than Egger's test [43, 59].

Example 5.11 If one of the arcsine tests is used, the arcsine difference must be specified as effect measure. To this aim, we generate a new meta-analysis
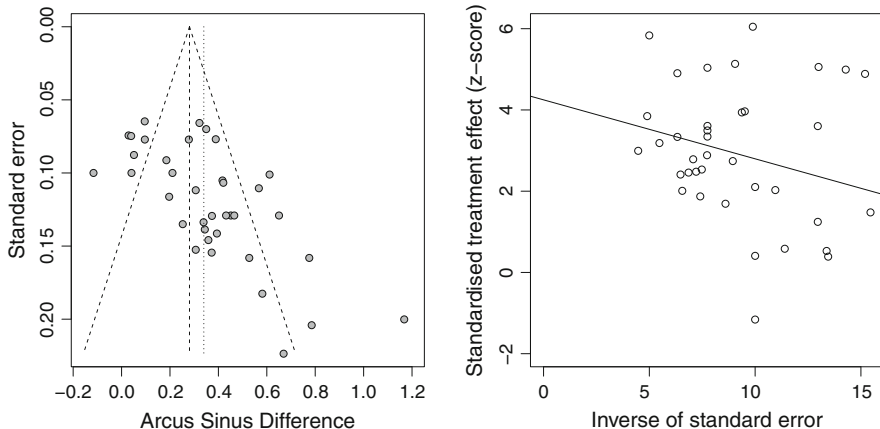
Fig. 5.8 Funnel plot and radial plot for meta-analysis on NSAIDS in acute pain [34] using arcsine transformation

`ms1.asd` using the `update.meta` function (setting argument `sm="ASD"`).[2] The Egger and Begg and Mazumdar versions of the arcsine test are called using the `metabias` function with arguments `method = "linreg"` or `method = "rank"`, respectively. The Thompson and Sharp version of the arcsine test can be carried out using `method = "mm"` in the `metabias` function. Furthermore, results can be illustrated in funnel plot and radial plot, see Fig. 5.8, here realised by setting `plotit = TRUE` in the `metabias` function.

```
> ms1.asd <- update(ms1, sm="ASD")
> summary(ms1.asd)
Number of studies combined: k=37

                         ASD         95%-CI       z  p-value
Fixed effect model   0.2806 [0.2471; 0.3141] 16.4200 < 0.0001
Random effects model 0.3398 [0.2686; 0.4111]  9.3448 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.0346; H = 2.04 [1.74; 2.39]; I^2 = 76% [67.1%; 82.5%]

Test of heterogeneity:
     Q d.f.  p-value
 149.79   36 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
> funnel(ms1.asd)
```

---

[2]Alternatively,   the   command   `ms1.asd <- metabin(Ee, Ne, Ec, Nc, data=data11, sm="ASD")` could have been used.

```
> metabias(ms1.asd, method = "mm", plotit=TRUE)

Linear regression test of funnel plot asymmetry (methods
of moment)

data:  ms1.asd
t = 5.0863, df = 35, p-value = 1.236e-05
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias    se.bias       slope
 4.2484242  0.8352690 -0.1451581
```

The Thompson and Sharp version of the arcsine test shows a significant *p*-value
($p < 0.0001$), leading to rejection of the null hypothesis of symmetry in the funnel
plot.   □

Table 5.1 gives an overview on regression tests. We note that Harbord's test and
the arcsine test follow the original principle of Egger's test to regress the treatment
effect on an estimate of its standard error, using inverse variance weights. In the
right part of the table, the alternative version of these tests is given, represented
by the radial plot, where the regression variables are defined such that weights are
unnecessary. Macaskill's test [32] and Peters' test [38] depart from this principle. In
this respect, they are ad hoc methods, nevertheless working well [32, 35, 38, 43].

Readers more deeply interested in the pros and cons of all these funnel plot-
based tests are advised to look at recommendations made by a group of authors led
by Sterne [56].

5.3   Adjusting for Small-Study Effects

All significance tests have in common that they at best give evidence of small-study
effects. They are, however, not designed to correct the treatment effect estimate
appropriately. In this section we discuss existing approaches providing adjusted
treatment effect estimates: the trim-and-fill method, the Copas selection model, and
regression approaches.

5.3.1   Trim-and-Fill Method

The trim-and-fill method is a nonparametric method to assess selection
bias/publication bias in meta-analysis [16, 17]. The method provides an estimate of
(1) the number of missing studies and (2) the treatment effect adjusted for selection
bias. The basic idea of the trim-and-fill method is to add studies to the funnel plot
until it becomes symmetric. The trim-and-fill method consists of the following five
steps:

Table 5.1 Overview: Regression tests for small-study effects in meta-analysis for binary outcomes

Test	Funnel plot version (weighted)			Radial plot version (unweighted)	
	Measure	Regressor	Weights	Measure	Regressor
Egger's test	$\hat{\theta}_k$	S.E. $(\hat{\theta}_k)$	$1/\widehat{\mathrm{Var}}(\hat{\theta}_k)$	$\hat{\theta}_k/$S.E. $(\hat{\theta}_k)$	$1/$S.E. $(\hat{\theta}_k)$
(generic, including test by Thompson and Sharp)					
Harbord's test	Z_k/V_k	$1/\sqrt{V_k}$	V_k	$Z_k/\sqrt{V_k}$	$\sqrt{V_k}$
Arcsine test	$\hat{\Delta}_k$	$\sqrt{\frac{1}{4n_{ek}}+\frac{1}{4n_{ck}}}$	$1/\left(\frac{1}{4n_{ek}}+\frac{1}{4n_{ck}}\right)$	$\hat{\Delta}_k\Big/\sqrt{\frac{1}{4n_{ek}}+\frac{1}{4n_{ck}}}$	$1\Big/\sqrt{\frac{1}{4n_{ek}}+\frac{1}{4n_{ck}}}$
(Egger version and Thompson version)					
Macaskill's test	$\hat{\theta}_k$	n_k	$1/\left(\frac{1}{a_k+c_k}+\frac{1}{b_k+d_k}\right)$	—	—
Peters' test	$\hat{\theta}_k$	$1/n_k$	$1/\left(\frac{1}{a_k+c_k}+\frac{1}{b_k+d_k}\right)$	—	—

1. Estimate the number of studies in the outlying part of the funnel plot using rank-based methods;
2. remove (trim) these studies and do meta-analysis on the remaining studies;
3. consider the estimate from the "trimmed" meta-analysis as the true center of the funnel;
4. for each "trimmed" study, create ("fill") an additional study as the mirror image about the center of funnel plot;
5. do meta-analysis on original and filled studies.

Three different methods have been proposed to estimate the number of missing studies [16, 17]. Two of these methods (L- and R-estimators) have been shown to perform better in simulations, and we use these here. In steps 1–4 and 5, respectively, either a fixed effect or random effects model can be used. Simulation results [39] indicate that the fixed–random model, that is, using a fixed effect model for steps 1–4 and a random effects model for step 5 (1) performs better than the fixed–fixed model and (2) performs no worse than and marginally better in certain situations than the random–random model, and we adopt this approach here.

As the Cochrane Handbook for Systematic Reviews of Interventions [25] notes, the trim-and-fill method assumes that the small-study effect is caused by selection, but requires no assumptions about the mechanism leading to small-study effects. However, it is built on the strong assumption of a symmetric funnel plot.

The performance of the trim-and-fill method was compared to that of the Copas selection model—described in Sect. 5.3.2—using a large set of meta-analyses [49] and to regression approaches [39, 46]—described in Sect. 5.3.3. The method is known to perform poorly in the presence of substantial between-study heterogeneity [39, 57]. Additionally, estimation and inferences are based on a dataset containing imputed intervention effect estimates, potentially resulting in too narrow CIs for the overall treatment effect.

Example 5.12 We now apply the trim-and-fill method to the NSAIDS example using the `trimfill` function of R package **meta**. The commands

```
> tf1 <- trimfill(ms1)
> class(tf1)
[1] "metagen"  "meta"     "trimfill"
> funnel(tf1)
```

yield a default trim-and-fill analysis. As we can see R object `tf1` has several classes. As the `metagen` function is called internally, R object `tf1` is of classes `metagen` and `meta`; class `trimfill` is added by the `trimfill` function. The resulting funnel plot is shown in Fig. 5.9. The filled-in study results are printed as open circles.

The following command prints R object `tf1`.

```
> print(tf1, digits=2, comb.fixed=TRUE)
                OR          95%-CI %W(fixed)  %W(random)
1             6.57 [2.11;  20.48]     1.49        2.13
*** Output truncated ***
37            5.69 [1.51;  21.42]     1.10        1.91
Filled: 37    0.95 [0.25;   3.56]     1.10        1.91
```
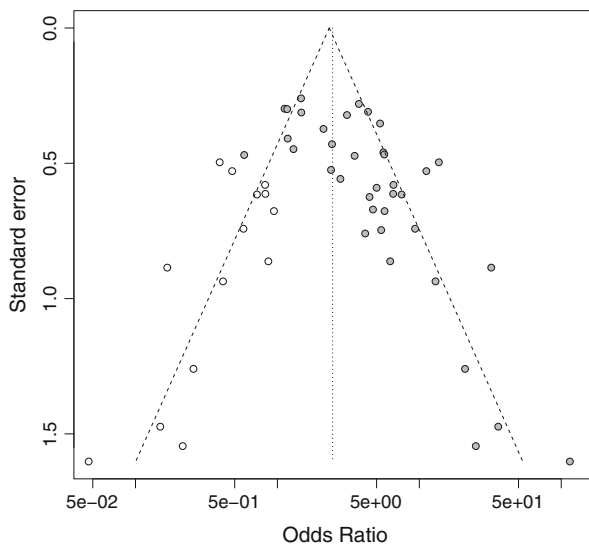
Fig. 5.9 Funnel plot for meta-analysis on NSAIDS in acute pain [34] after applying the trim-and-fill method

```
Filled: 27   0.86 [0.16;   4.68]      0.67      1.53
Filled: 16   0.82 [0.25;   2.73]      1.34      2.05
Filled: 1    0.82 [0.26;   2.55]      1.49      2.13
Filled: 21   0.72 [0.21;   2.40]      1.32      2.05
Filled: 20   0.58 [0.13;   2.47]      0.91      1.77
Filled: 15   0.48 [0.17;   1.35]      1.79      2.25
Filled: 5    0.41 [0.07;   2.59]      0.57      1.40
Filled: 30   0.39 [0.15;   1.04]      2.04      2.33
Filled: 36   0.26 [0.02;   3.03]      0.32      0.97
Filled: 22   0.22 [0.01;   4.45]      0.21      0.72
Filled: 14   0.17 [0.03;   0.95]      0.64      1.49
Filled: 31   0.15 [0.01;   2.68]      0.23      0.77
Filled: 32   0.05 [0.00;   1.08]      0.20      0.68

Number of studies combined: k=51 (with 14 added studies)

                     OR      95%-CI     z  p-value
Fixed effect model   2.32 [2.02; 2.66] 11.87 < 0.0001
Random effects model 2.45 [1.83; 3.28]  6.00 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.7113; H = 1.93 [1.68; 2.22]; I^2 = 73.2% [64.7%; 79.7%]

Test of heterogeneity:
     Q d.f.  p-value
 186.73   50 < 0.0001

Details on meta-analytical method:
- Inverse variance method
```

```
- DerSimonian-Laird estimator for tau^2
- Trim-and-fill method to adjust for funnel plot asymmetry
```

First, all 37 NSAIDS studies are printed followed by 14 filled-in studies that are mirror images of the most outlying studies (numbers given in the output, e.g. No. 37, 27, . . .). Note, the fixed effect estimate is not printed by default as the corresponding confidence interval is too narrow. Accordingly, argument `comb.fixed=TRUE` is used in the above command. Estimates from fixed effect and random effects model are much more similar after applying the trim-and-fill method.   □

5.3.2 Copas Selection Model

In contrast to the trim-and-fill method, the selection model by Copas [10, 12–14] explicitly models publication bias.

Briefly, the Copas selection model has two components:

(a) a model for the treatment effect,
(b) a model giving the probability that study k is selected for publication.

A correlation parameter ρ between these two components indicates the extent of publication bias; the stronger the correlation, the greater the chance that only the more extreme treatment effects are selected for publication and observed by others.

In more detail, let (ϵ_k, δ_k) follow a bivariate normal distribution with mean 0 and covariance matrix

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Given parameters θ, τ^2, and σ_k^2, the observed treatment effect $\hat{\theta}_k$ (e.g., log odds ratio) of study k is modelled by the random effects model

$$\hat{\theta}_k = \theta + \sqrt{\sigma_k^2 + \tau^2}\, \epsilon_k \tag{5.2}$$

(approximating a normal distribution). This form of the random effects model comprises both components of the variance into a single term. The first component, σ_k^2, represents the random error within study k. The second component, τ^2 represents the variance between studies (that constitutes the "random" effect). Thus $\sigma_k^2 + \tau^2$ is the total variance coming from both sources of variation, such that $\sqrt{\sigma_k^2 + \tau^2}$ is the standard deviation. Accordingly, the error terms are merged to one term ϵ_k. An extended version of this model will be treated in Sect. 5.3.3.

We say study k is observed if $Z_k > 0$, where

$$Z_k = \gamma_0 + \gamma_1/s_k + \delta_k \tag{5.3}$$

with fixed γ_0 and γ_1 (for their choice, see below) and standard error s_k. From (5.3), the marginal probability that study k is observed is

$$\Pr(Z_k > 0) = \Pr(\delta_k > -\gamma_0 - \gamma_1/s_k) = \Phi(\gamma_0 + \gamma_1/s_k), \tag{5.4}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal. Thus, $\Phi(\gamma_0)$ can be interpreted as the marginal probability of publishing a study with infinite standard error, and γ_1 is associated with the change in publication probability with increasing precision. Note that the appearance of s_k in (5.3) means that the probability of publication reflects the sampling variance from the data simulated for study k.

Copas and Shi [10, 13] use standard properties of the normal distribution to show that the probability of observing study k is

$$\Phi \left\{ \frac{\gamma_0 + \gamma_1/s_k + \rho\sigma_k \frac{\hat{\theta}_k - \theta}{\sigma_k^2 + \tau^2}}{\sqrt{1 - \rho^2\sigma_k^2/(\sigma_k^2 + \tau^2)}} \right\}. \tag{5.5}$$

Thus, if $\rho = 0$, (5.2) and (5.3) are unrelated, so a meta-analysis of observed studies will give an approximately unbiased estimate of θ. Conversely, if $-1 \leq \rho < 0$ then the probability of observing study k is increased the smaller $\hat{\theta}_k$, because this corresponds to smaller ϵ_k and hence larger δ_k. In this situation, a meta-analysis of observed studies will give a downwardly biased estimate of θ.

The Copas selection model has two parameters γ_0 and γ_1 describing the extent of selection. If the model is fitted, these parameters are not estimated, but fixed. Later, in a sensitivity analysis they are varied and the sensitivity of the adjusted effect to variation of γ_0 and γ_1 is examined. To interpret γ_0 and γ_1, take the marginal probability of observing study k, given by (5.4). If we look at the standard error of smallest and largest study, s_{small} and s_{large}, we can assume that the smallest study is published with probability p_{small} and the largest study with probability p_{large}, for example, take $p_{\text{small}} = 0.1$ and $p_{\text{large}} = 0.9$. Then we find γ_0 and γ_1 by solving

$$p_{\text{small}} = \Phi(\gamma_0 + \gamma_1/s_{\text{small}})$$
$$p_{\text{large}} = \Phi(\gamma_0 + \gamma_1/s_{\text{large}}).$$

For fitting the Copas selection model and carrying out the sensitivity analysis, a special R package **copas** was written [6] which has been replaced by R package **metasens** [50]. The implementation of the Copas selection model is described in detail in Carpenter et al. [7]. The Copas selection model has also been compared to the trim-and-fill method [49].

Example 5.13 The following set of commands (i) loads the R package **metasens**, (ii) applies the `copas` function to the meta-analysis object `ms1` defined above, thus
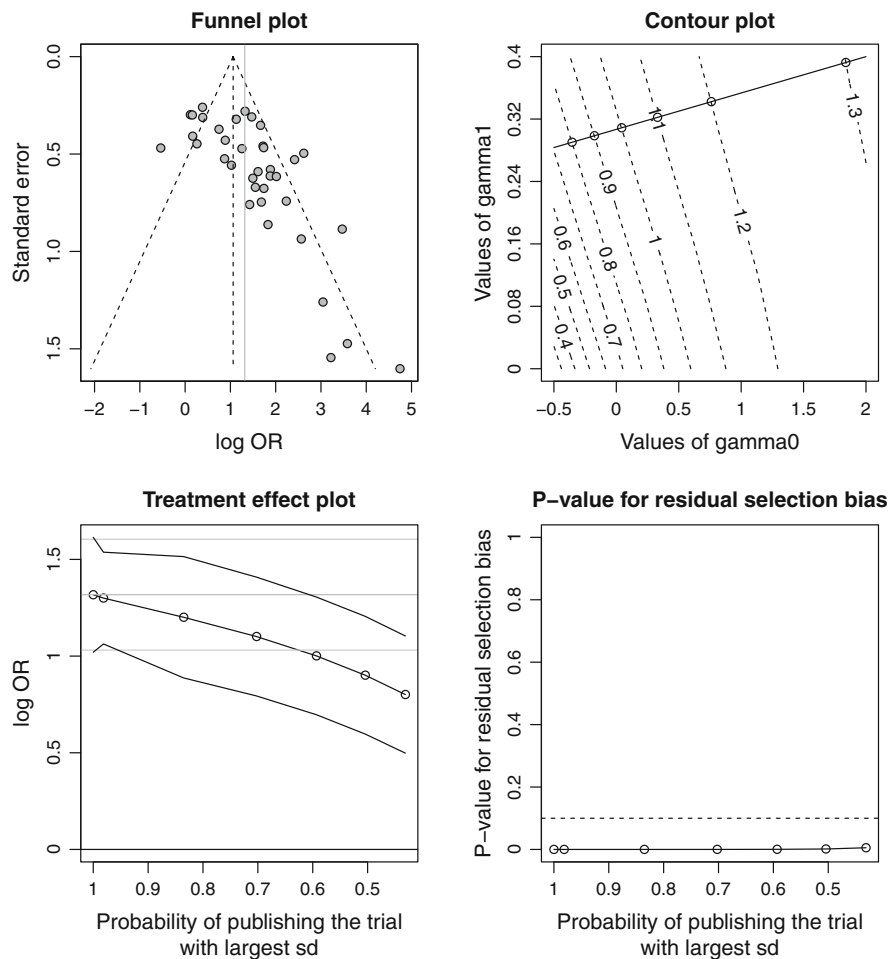
Fig. 5.10 Graphical output from Copas selection model for meta-analysis on NSAIDS in acute pain [34]

producing an object `c1` of class `copas`, and (iii) produces a graphical summary of the result of the Copas selection model.

```
> library(metasens)
Loading required package: meta
Loading 'meta' package (version 4.0-2).
Loading 'metasens' package (version 0.2-0).
> c1 <- copas(ms1)
> plot(c1)
```

Figure 5.10 shows the four plots generated by the plot command.

***Funnel plot* (Top Left)**

We know this plot already from Fig. 5.4. Note, in this funnel plot values on the *x*-axis are printed as log odds ratios instead of odds ratios.

***Contour plot* (Top Right)**

This plot shows the estimated treatment effects, measured as log odds ratios, for a range of pairs (γ_0, γ_1), with contours delineated for selected log odds ratios (dashed lines). With decreasing probability of publishing the study with the largest standard error (that is, increasing selection), we are stepping downwards along the line of steepest descent (the solid line approximately orthogonal to the contours) from the top right corner of the plot to the left. This line of steepest descent is crossing six contour lines in Fig. 5.10 with corresponding treatment estimates decreasing from a log odds ratio of about 1.3–0.8.

***Treatment effect plot* (Bottom Left)**

The six treatment estimates highlighted in the contour plot are also shown in the third plot, along with their 95 % confidence intervals. In this plot a seventh treatment estimate is added on the left side corresponding to a Copas selection model without adjustment for selection. This additional treatment estimate is very similar to the estimate from the usual random effects model, however, uses a Maximum Likelihood estimator for the between-study variance instead of the DerSimonian–Laird method. These seven treatment estimates are part of R object `c1`.

```
> c1$TE.slope
[1] 1.3170832 1.3000441 1.2007513 1.1010823 1.0010173 0.9011074
[7] 0.8011047
```

Values on the *x*-axis are calculated according to Eq. (5.4) using the largest standard error of all studies in the given meta-analysis (the "smallest study") and values for (γ_0, γ_1) from the contour plot where the line of steepest descent crosses the contour lines. The largest standard error s_{max} can be easily calculated from R object `c1`.

```
> max(c1$seTE)
[1] 1.602172
```

The corresponding values for (γ_0, γ_1) are not part of R object `c1` but can be calculated.

```
> gamma0 <- min(c1$gamma0.range) + c1$x.slope*diff(c1$gamma0.range)
> gamma1 <- min(c1$gamma1.range) + c1$y.slope*diff(c1$gamma1.range)
```

Accordingly, the probability of publishing the study with the largest standard error can be calculated.

```
> print(pnorm(gamma0 + gamma1/max(c1$seTE)), digits=2)
[1] 0.98 0.84 0.70 0.59 0.50 0.43
```

This plot clearly shows that the estimated treatment effect is decreasing if the degree of selection is increasing (i.e. by decreasing the probability of publishing the study with the largest standard error).

P-value plot (**Bottom Right**)

The last plot shows the *p*-value of the test for residual selection, which is significant (here defined as being less than 0.1—see argument `sign.rsb` in `plot.copas` function) throughout. If this would cross the $p = 0.1$ line, the adjusted treatment effect would be read off at this point of selection. This cannot be done here since we do not observe the point of crossing anymore. The reason is the large extent of selection in the example.

A summary of the two bottom plots in Fig. 5.10 is provided by the `summary.copas` function.

```
> print(summary(c1), digits=2)
Summary of Copas selection model analysis:

          publprob   OR        95%-CI pval.treat pval.rsb N.unpubl
              1.00 3.73 [2.77; 5.02]   < 0.0001 < 0.0001        0
              0.98 3.67 [2.89; 4.65]   < 0.0001 < 0.0001        0
              0.84 3.32 [2.43; 4.55]   < 0.0001 < 0.0001        3
              0.70 3.01 [2.21; 4.09]   < 0.0001 < 0.0001        8
              0.59 2.72 [2.01; 3.69]   < 0.0001   0.0004       14
              0.50 2.46 [1.82; 3.34]   < 0.0001   0.0014       20
              0.43 2.23 [1.65; 3.02]   < 0.0001   0.0054       28

Copas model (adj)
RE model           3.73  [2.80; 4.97]  < 0.0001

 Significance level for test of residual selection bias: 0.1

 Legend:
 publprob   - Probability of publishing study with largest
              standard error
 pval.treat - P-value for hypothesis of overall treatment effect
 pval.rsb   - P-value for hypothesis that no selection remains
              unexplained
 N.unpubl   - Approximate number of unpublished studies suggested
              by model
```

The column `publprob` lists the probabilities of publishing the study with the largest standard error already used in the two bottom plots of Fig. 5.10. Selection increases from top to bottom. The second column, `OR` lists the odds ratios estimated by the model if selection takes place as given by the publication probability in the first column, together with a 95 % confidence interval in the next column `95%-CI` (left bottom plot shows log odds ratios). Column `pval.treat` shows the *p*-value

for the null hypothesis that the treatment effect is equal in both groups, which for the example is highly significant throughout. This is likewise true here for column `pval.rsb` which shows the *p*-value for the hypothesis that no further selection remains unexplained already used in the bottom right plot of Fig. 5.10. In other words, selection remains unexplained here even if the probability of publishing the "smallest study" is as low as 0.43. The last column, `N.unpubl` tells us how many studies are assumed to be missing based on the Copas selection model; for details see [13]. Unfortunately, the result of the Copas selection model which should be shown after the text `Copas model (adj)` is missing here. The reason is that selection seems to be even greater than assumed by the default range of potential selection probabilities, here from 0.43 to 1.00.

As a remedy, we can extend the range of the parameters γ_0 and γ_1 by explicit specifying a greater range. To this aim, we first look at the full output using the `print.copas` function.

```
> c1
Copas selection model analysis

                   min max
 range of gamma0:  -0.5 2.0
 range of gamma1:   0.0 0.4
*** Output truncated ***
```

followed by more information on the Copas selection model analysis—including the summary given above. In the NSAIDS meta-analysis the default range is $[-0.5, 2]$ and $[0, 0.4]$ for parameter γ_0 and γ_1, respectively. We therefore decide to extend the ranges to $[-1, 2]$ and $[0, 1]$.

```
> c2 <- copas(ms1, gamma0.range=c(-1,2), gamma1.range=c(0,1))
> plot(c2)
```

The resulting plot is shown in Fig. 5.11. We find that the range of γ_0 and γ_1 in the contour plot is now sufficiently large that the *p*-value curve crosses the $p = 0.1$ line in the bottom right plot, which takes place at a selection probability of about 0.3.

From the summary output we see that no evidence of residual selection is present for a selection probability of 0.26.

```
> print(summary(c2), digits=2)
Summary of Copas selection model analysis:

        publprob   OR       95%-CI pval.treat pval.rsb N.unpubl
            1.00 3.73 [2.77; 5.02]  < 0.0001 < 0.0001        0
            0.75 3.33 [2.45; 4.52]  < 0.0001 < 0.0001        4
            0.49 2.73 [2.05; 3.64]  < 0.0001   0.0018       15
            0.35 2.24 [1.72; 2.92]  < 0.0001   0.0356       30
            0.26 1.82 [1.46; 2.26]  < 0.0001   0.3273       48

Copas model (adj) 1.82 [1.46; 2.26]  < 0.0001   0.3273       48
RE model          3.73 [2.80; 4.97]  < 0.0001
*** Output truncated ***
```
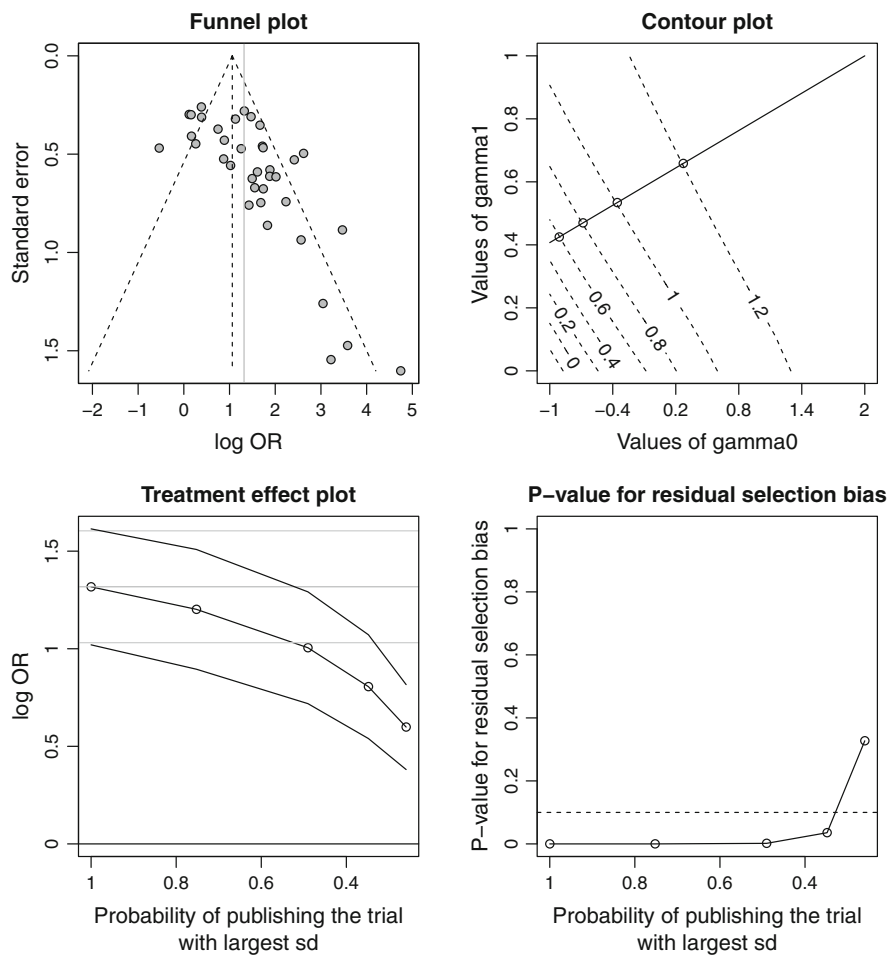
Fig. 5.11 Graphical output from Copas selection model for meta-analysis on NSAIDS in acute pain [34] after changing ranges for sensitivity parameters *gamma*$_0$ and *gamma*$_1$

The adjusted treatment effect estimate based on the Copas selection model is therefore read off at the corresponding place in the treatment effect plot (and printed in the last row of the table in the summary) as 1.82 [1.46; 2.26]. This result is also given as "Copas model (adj)" in the printed output. We conclude that though there is much selection, the adjusted estimate from the Copas selection model still indicates that the treatment is effective.   □

5.3.3 Adjustment by Regression

Stanley [51] seems to be the first author who proposed a regression-based treatment effect estimate adjusting for small-study effects. Moreno et al. [35] compared various regression-based approaches for adjusting the treatment effect for small-study effects and included also the trim-and-fill method in their extensive simulation study. The methods proposed by Moreno et al. [35] are based on regression tests discussed in Sect. 5.2. The slope of the weighted linear regression is used to construct a test for asymmetry of the funnel plot. Thereby the intercept is interpreted as the estimated treatment effect for a study of infinite precision, adjusted for small-study effects. This idea was also discussed by Copas and Malley [12].

Another paper by Moreno et al. [36] showed that their adjusted treatment effect estimates were able to predict the effect of the whole database of antidepressant trials in the FDA registry from a biased subset of these trials that were published.

Rücker et al. [45, 46] used a similar approach and combined it with a shrinkage procedure. The underlying model is an extended random effects model that takes account of possible small-study effects by allowing the effect to depend on the standard error.

$$\hat{\theta}_k = \theta_* + \sqrt{\sigma_k^2 + \tau^2}\, (\epsilon_k + \theta_B), \quad \epsilon_k \overset{\text{i.i.d.}}{\sim} N(0, 1). \tag{5.6}$$

Here $\hat{\theta}_k$ is the observed effect in study k, θ_* the global mean, σ_k^2 the within-study sampling variance, and τ^2 the between-study variance. The parameter θ_B represents the bias introduced by small-study effects, as is seen by either of the following considerations.

On the one hand, θ_B can be interpreted as the expected shift in the standardized treatment effect if precision is very small

$$\mathrm{E}\left[\frac{\hat{\theta}_k - \theta_*}{\sigma_k}\right] \to \theta_B \quad \text{if} \quad \sigma_k \to \infty.$$

On the other hand, $\theta_{\text{adj}} = \theta_* + \tau\theta_B$ is interpreted as the limit treatment effect if precision is infinite

$$\mathrm{E}[\theta_k] \to \theta_* + \tau\theta_B \quad \text{if} \quad \sigma_k \to 0.$$

Note that as θ_B is included in Eq. (5.6), θ_* has not the same interpretation as parameter θ in the usual random effects model (2.11). The two models are only the same if $\theta_B = 0$. If there are genuine small-study effects, i.e. $\theta_B \neq 0$, model (5.6) includes a component making the treatment effect depend on the standard error. The expected treatment effect of a study of infinite size is $\theta_* + \tau\theta_B$, whereas θ_* alone does not appropriately represent the treatment effect.

The ML estimates $\hat{\theta}_B$ and $\hat{\theta}_*$ can be interpreted as intercept and slope in a linear regression on a so-called generalised radial plot, which makes them readily available by standard software [12]. Using estimates for θ_B, θ_* and τ, the adjusted treatment effect is estimated by $\hat{\theta}_{\mathrm{adj}} = \hat{\theta}_* + \hat{\tau}\hat{\theta}_B$.

Example 5.14 Standard regression modelling applied to the generalised radial plot and using the `lm` function could be used for parameter estimation in the extended random effects model. However, it is more convenient to use the `limitmeta` function of R package **metasens**. We apply it to our example.

```
> l1 <- limitmeta(ms1)
> print(l1, digits=2)
Results for individual studies (left: original data;
                                 right: shrunken estimates)

          OR          95%-CI       OR          95%-CI
1      6.57 [2.11;   20.48]     2.65 [0.85;  8.25]
*** Output truncated ***
37     5.69 [1.51;   21.42]     1.96 [0.52;  7.38]

Result of limit meta-analysis:

 Random effects model    OR        95%-CI    z      pval
    Adjusted estimate 1.84 [1.26; 2.68] 3.17   0.0015
  Unadjusted estimate 3.73 [2.80; 4.97] 9.01 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.4670; I^2 = 68.3% [55.5%; 77.4%]; G^2 = 91.5%

Test of heterogeneity:
      Q d.f.  p.value
 113.52    36 < 0.0001

Test of small-study effects:
   Q-Q' d.f.  p.value
  44.20     1 < 0.0001

Test of residual heterogeneity beyond small-study effects:
     Q' d.f.  p.value
  69.32    35   0.0005

Details on adjustment method:
- expectation (beta0)
```

The printout shows the estimates and 95 % confidence intervals of the primary studies (left) and those of the shrunken estimates (right), followed by the result of the adjusted model, compared to that of the usual random effects model. The adjusted point estimate is comparable to that of the Copas selection model given in Sect. 5.3.2, OR = 1.82 [1.46; 2.26], however, the 95 % confidence interval is considerably wider. Both results differ substantially from the results of the

trim-and-fill method given in Sect. 5.3.1, OR = 2.45 [1.83; 3.28] for the random effects model.

Heterogeneity is quantified as usual, with an added measure G^2 that is interpreted as the proportion of unexplained variance after allowing for possible small-study effects in the limit meta-analysis [45]. In the example, $G^2 = 91.5\%$ is very large, which means that there is still unexplained variance even after adjusting for small-study effects. This is caused by the large studies at the top of the funnel plot shown in Fig. 5.4 which have very different point estimates that cannot be explained by a small-study effect. The various sources of heterogeneity are separated by partitioning $Q = 113.52$ into two parts $Q - Q' = 44.2$ and $Q' = 69.2$ as shown in the next part of the printout. Both parts are interpreted as tests and give significant results, that is, we have both a small-study effect and residual heterogeneity beyond that.

The limitmeta function was run with the default option method.adjust= "beta0" that represents one of three available adjustment methods; for details, see [45].

Finally, we show how the result of the regression-based adjusting method can be illustrated by inserting a curve into the funnel plot using the R command

```
> funnel(l1)
```

The resulting funnel plot, shown in Fig. 5.12, adds to the fixed effect and random effects estimates (dashed and dotted vertical lines, respectively) a grey curve. It
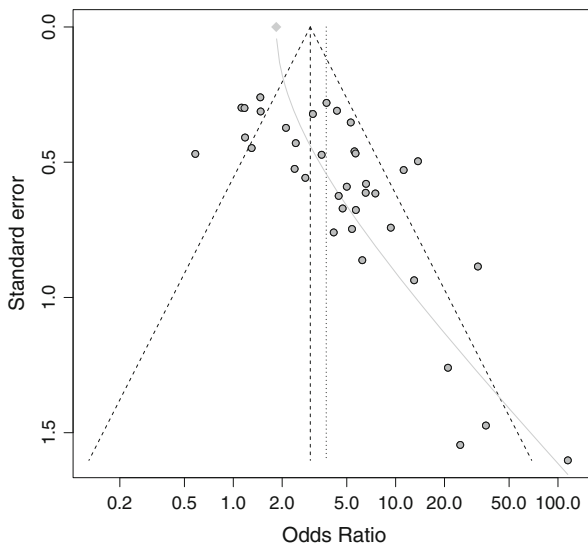
Fig. 5.12 Graphical output from regression-based adjustment method for meta-analysis on NSAIDS in acute pain [34]. The *grey curve* runs from a biased effect estimate for a very small-study (*bottom*) to the adjusted estimate for a study with infinite precision (*top*)

starts at the bottom from a very biased treatment effect estimate (thought resulting from a hypothetical small-study with infinite standard error), then runs to the top of the figure and finally ends at the adjusted treatment effect estimate $\exp(\hat{\theta}_{\text{adj}})$, indicated by a grey diamond. The grey curve corresponds to the regression line originating from the generalised radial plot, here transformed and transposed for fitting into the funnel plot. $\square$

5.4   Summary

In this chapter we have described how the extent of possible small-study effects such as publication bias can be investigated and—if appropriate—adjusted for. We began by showing how R could be used to create funnel and radial plots, and discussed their interpretation. Then, we reviewed the large number of statistical tests for funnel plot asymmetry, and illustrated their application. We concluded by illustrating three established methods which attempt to correct the treatment estimate for small-study effects.

References

1. A'hern, R.P.: Widening eligibility to phase II trials: constant arcsine difference phase II trials. Control. Clin. Trials **25**, 251–264 (2004)
2. Armitage, P., Berry, G., Matthews, J.N.S.: Statistical Methods in Medical Research. Blackwell Science, Oxford (1987)
3. Begg, C.B., Berlin, J.A.: Publication bias: a problem in interpreting medical data. J. R. Stat. Soc. Ser. A **151**, 419–445 (1988). (C/R: p445-463)
4. Begg, C.B., Mazumdar, M.: Operating characteristics of a rank correlation test for publication bias. Biometrics **50**, 1088–1101 (1994)
5. Berkey, C.S., Hoaglin, D.C., Mosteller, F., Colditz, G.A.: A random-effects regression model for meta-analysis. Stat. Med. **14**, 395–411 (1995)
6. Carpenter, J., Rücker, G., Schwarzer, G.: copas: an R package for fitting the Copas selection model. R J. **1**(2), 31–36 (2009)
7. Carpenter, J.R., Schwarzer, G., Rücker, G., Künstler, R.: Empirical evaluation showed that the Copas selection model provided a useful summary in 80 % of meta-analyses. J. Clin. Epidemiol. **62**, 624–631 (2009)
8. Chan, A.W., Hróbartsson, A., Haahr, M.T., Gøtzsche, P.C., Altman, D.G.: Empirical evidence for selective reporting of outcomes in randomized trials. comparison of protocols to published articles. J. Am. Med. Assoc. **291**, 2457–2465 (2004)
9. Chan, A.W., Krleza-Jeric, K., Schmid, I., Altman, D.G.: Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. Can. Med. Assoc. J. **171**, 735–740 (2004)
10. Copas, J.: What works?: Selectivity models and meta-analysis. J. R. Stat. Soc. Ser. A **162**, 95–109 (1999)
11. Copas, J., Lozada-Can, C.: The radial plot in meta-analysis: approximations and applications. Appl. Stat. **58**(3), 329–344 (2009)

12. Copas, J.B., Malley, P.F.: A robust p-value for treatment effect in meta-analysis with publication bias. Stat. Med. **27**(21), 4267–4278 (2008)
13. Copas, J., Shi, J.Q.: Meta-analysis, funnel plots and sensitivity analysis. Biostatistics **1**, 247–262 (2000)
14. Copas, J.B., Shi, J.Q.: A sensitivity analysis for publication bias in systematic reviews. Stat. Methods Med. Res. **10**, 251–265 (2001)
15. Duncan, B., Olkin, I.: Bias of estimates of the number needed to treat. Stat. Med. **24**, 1837–1848 (2005)
16. Duval, S., Tweedie, R.: A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. J. Am. Stat. Assoc. **95**, 89–98 (2000)
17. Duval, S., Tweedie, R.: Trim and Fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics **56**, 455–463 (2000)
18. Dwan, K., Altman, D.G., Arnaiz, J.A., Bloom, J., Chan, A., Cronin, E., Decullier, E., Easterbrook, P.J., Elm, E.V., Gamble, C., Ghersi, D., Ioannidis, J.P., Simes, J., Williamson, P.R.: Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE **3**(8), e3081 (2008). Doi: 10.1371/journal.pone.0003081
19. Easterbrook, P.J., Berlin, J.A., Gopalan, R., Matthews, D.R.: Publication bias in clinical research. Lancet **337**, 867–872 (1991)
20. Egger, M., Smith, G.D., Schneider, M., Minder, C.: Bias in meta-analysis detected by a simple, graphical test. Br. Med. J. **315**, 629–634 (1997)
21. Galbraith, R.F.: A note on graphical presentation of estimated odds ratios from several clinical trials. Stat. Med. **7**, 889–894 (1988)
22. Galbraith, R.F.: Graphical display of estimates having differing standard errors. Technometrics **30**, 271–281 (1988)
23. Gleser, L.J., Olkin, I.: Models for estimating the number of unpublished studies. Stat. Med. **15**, 2493–2507 (1996)
24. Harbord, R.M., Egger, M., Sterne, J.A.: A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat. Med. **25**(20), 3443–3457 (2006)
25. Higgins, J.P., Green, S. (eds.): Cochrane Handbook for Systematic Reviews of Interventions—Version 5.1.0 [updated March 2011]. The Cochrane Collaboration (2011). URL http://www.cochrane-handbook.org
26. Ioannidis, J.P., Trikalinos, T.A.: An exploratory test for an excess of significant findings. Clin. Trials **4**(3), 245–253 (2007)
27. Ioannidis, J.P.A., Trikalinos, T.A.: The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. Can. Med. Assoc. J. **176**(8), 1091–1096 (2007)
28. Kasuya, E.: Angular transformation—another effect of different sample sizes. Ecol. Res. **19**, 165–167 (2004)
29. Kendall, M., Gibbons, J.D.: Rank Correlation Methods, 5th edn. Edward Arnold, London (1990)
30. Lau, J., Ioannidis, J.P.A., Terrin, N., Schmid, C.H., Olkin, I.: The case of the misleading funnel plot. Br. Med. J. **333**, 597–600 (2006)
31. Light, R.J., Pillemer, D.B.: Summing up. In: The Science of Reviewing Research. Harvard University Press, Cambridge, MA (1984)
32. Macaskill, P., Walter, S.D., Irwig, L.: A comparison of methods to detect publication bias in meta-analysis. Stat. Med. **20**, 641–654 (2001)
33. McCullagh, P., Nelder, J.: Generalized Linear Models. Chapman & Hall, London (1989)
34. Moore, R.A., Tramer, M.R., Carroll, D., Wiffen, P.J., McQuay, H.J.: Quantive systematic review of topically applied non-steroidal anti-inflammatory drugs. Br. Med. J. **316**(7128), 333–338 (1998)
35. Moreno, S., Sutton, A., Ades, A., Stanley, T., Abrams, K., Peters, J., Cooper, N.: Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Med. Res. Methodol. **9**, 2 (2009). URL http://www.biomedcentral.com/1471-2288/9/2/abstract

36. Moreno, S.G., Sutton, A.J., Turner, E.H., Abrams, K.R., Cooper, N.J., Palmer, T.P., Ades, A.E.: Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. Br. Med. J. **339**, b2981 (2009). Doi: 10.1136/bmj.b2981

37. Orwin, R.G.: A fail-safe *N* for effect size in meta-analysis. J. Educ. Stat. **8**, 157–159 (1983)

38. Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L.: Comparison of two methods to detect publication bias in meta-analysis. J. Am. Med. Assoc. **295**, 676–680 (2006)

39. Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L.: Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat. Med. **27**, 4544–4562 (2007)

40. Peters, J.L., Sutton, A.J., Jones, D.J., Abrams, K.R., Rushton, L.: Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J. Clin. Epidemiol. **61**(10), 991–996 (2008)

41. Rosenthal, R.: The "file drawer problem" and tolerance for null results. Psychol. Bull. **86**, 638–641 (1979)

42. Rothstein, H.R., Sutton, A.J., Borenstein, M.: Publication Bias in Meta Analysis: Prevention, Assessment and Adjustments. Wiley, Chichester (2005)

43. Rücker, G., Schwarzer, G., Carpenter, J.R.: Arcsine test for publication bias in meta-analyses with binary outcomes. Stat. Med. **27**(5), 746–763 (2008)

44. Rücker, G., Schwarzer, G., Carpenter, J., Olkin, I.: Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. Stat. Med. **28**(5), 721–738 (2009)

45. Rücker, G., Schwarzer, G., Carpenter, J., Binder, H., Schumacher, M.: Treatment effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics **12**(1), 122–142 (2010). Doi:10.1136/jme.2008.024521.

46. Rücker, G., Carpenter, J., Schwarzer, G.: Detecting and adjusting for small-study effects in meta-analysis. Biom. J. **53**(2), 351–368 (2011)

47. Schwarzer, G., Antes, G., Schumacher, M.: 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 (2002)

48. Schwarzer, G., Antes, G., Schumacher, M.: A test for publication bias in meta-analysis with sparse binary data. Stat. Med. **26**, 721–733 (2007)

49. Schwarzer, G., Carpenter, J.R., Rücker, G.: Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. J. Clin. Epidemiol. **63**, 282–288 (2010)

50. Schwarzer, G., Carpenter, J., Rücker, G.: metasens: Advanced statistical methods to model and adjust for bias in meta-analysis (2014). URL http://cran.R-project.org/package=metasens. R package version 0.1-0

51. Stanley, T.D.: Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxf. Bull. Econ. Stat. **70**(105–127) (2008)

52. Sterling, T.D.: Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. J. Am. Stat. Assoc. **54**, 30–34 (1959). (Comment: V54 p593)

53. Sterling, T.D., Rosenbaum, W.L., Weinkam, J.J.: Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. Am. Stat. **49**, 108–112 (1995)

54. Sterne, J.A.C., Egger, M.: Funnel plots for detecting bias in meta-analysis: guideline on choice of axis. J. Clin. Epidemiol. **54**, 1046–1055 (2001)

55. Sterne, J.A.C., Gavaghan, D., Egger, M.: Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J. Clin. Epidemiol. **53**, 1119–1129 (2000)

56. Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rücker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D., Higgins, J.P.T.: Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. Br. Med. J. **343**, d4002 (2011). URL http://bmj.com/cgi/content/full/bmj.d4002. Doi: 10.1136/bmj.d4002

57. Terrin, N., Schmid, C.H., Lau, J., Olkin, I.: Adjusting for publication bias in the presence of heterogeneity. Stat. Med. **22**, 2113–2126 (2003)
58. Terrin, N., Schmid, C.H., Lau, J.: In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J. Clin. Epidemiol. **58**(9), 894–901 (2005)
59. Thompson, S.G., Sharp, S.J.: Explaining heterogeneity in meta-analysis: a comparison of methods. Stat. Med. **18**, 2693–2708 (1999)
60. Vevea, J.L., Hedges, L.V.: A general linear model for estimating effect size in the presence of publication bias. Psychometrika **60**, 419–435 (1995)
61. Williamson, P.R., Gamble, C.: Identification and impact of outcome selection bias in meta-analysis. Stat. Med. **24**(10), 1547–1561 (2005)

Chapter 6
Missing Data in Meta-Analysis

In this chapter we discuss issues raised by missing data. In Sect. 6.1 we discuss how to explore the robustness of our inference to different assumptions about missing outcome measures, while in Sect. 6.2 we describe an imputation approach which may be used when a study does not report the precision.

6.1 Missing Outcome Data: Some Considerations

Here we consider the case where one, or more, studies contributing to a meta-analysis have missing outcome data. Our intention is to outline some of the issues to consider, and illustrate how some of the resulting analyses may be performed using R. For a broader discussion of the issues, see White et al. [9].

Usually, a study report will detail the nature and extent of missing data in each intervention arm, following the CONSORT guidelines [4]. If the proportion of patients with missing outcome is non-trivial, the reported point estimate and standard error need to be interpreted in the light of this. In particular, reviewers need to look at the assumptions about the missing outcome data that underpin the reported results, and consider how they relate to the overall research question of the systematic review.

In many trials, outcome data will only be collected while patients broadly comply with the intervention and other requirements of the protocol; when they cease such compliance, they will be deemed to have withdrawn and subsequent outcome data will be missing. We call patients who complete the study according to such rules *de jure* patients. Thus, analysis of data from *de jure* patients will estimate the effect of the intervention if patients adequately comply with the intervention and other aspects of the trial protocol. This applies to analyses made under the *missing at random* assumption, which in this context assumes that the conditional distribution

of a patient's outcome data given baseline (and possibly intermediate follow-up data) is the same *whether or not* that outcome is actually observed.

By contrast, a pragmatic understanding of the intervention will often seek to estimate the effect of the intervention under *de facto* patient behaviour, representing the likely effect of use of the intervention in the health sector.

In this setting, we may wish to explore the sensitivity of a point estimate obtained from *de jure* patients to various assumptions about the difference in average outcomes between (often missing) *de facto* outcomes and (usually observed) *de jure* outcomes. Such analyses generally need to assume missing data are *missing not at random*, or that missing data are *informative*. Further discussion of these issues is given in [1] and Chapter 10 of [3].

In an ideal world, the assumptions made for the primary analysis would be clear from the report, and the authors would also explore and report the robustness of their primary analysis to different assumptions concerning the missing outcomes. Unfortunately, this is not always the case. The following approach described by [6], and developed in the context of meta-analysis of binary data by [7] may be helpful.

This is a two-stage approach. In the first stage, for contributing studies with missing outcome data we consider if any adjustment is needed to take account of this. Having made any adjustments, the second stage is simply to perform a meta-analysis in the usual way.

6.1.1 Study-Level Adjustment for Missing Data

Suppose study k reports a complete records analysis, i.e. treatment effect estimates $\hat{\theta}_{ke}$ and $\hat{\theta}_{kc}$ with corresponding standard errors S.E. $(\hat{\theta}_{ke})$ and S.E. $(\hat{\theta}_{kc})$ where "e" denotes experimental arm and "c" control arm. These quantities estimate the population parameters values

$$\left\{ \theta_{ke}, \sqrt{\text{Var}\,(\hat{\theta}_{ke})}, \theta_{kc}, \sqrt{\text{Var}\,(\hat{\theta}_{kc})} \right\} .$$

Accordingly, the treatment effect is defined as $\theta_k = \theta_{ke} - \theta_{kc}$. For ease of notation, we write s_e and s_c instead of S.E. $(\hat{\theta}_{ke})$ and S.E. $(\hat{\theta}_{kc})$, respectively, and omit the study index k when the context makes it clear we refer to a specific study.

Suppose that, for a particular study, treatment estimates are based on only n_{eo} out of n_e patients in the experimental arm and n_{co} out of n_c patients in the control arm.[1] Accordingly the proportion of missing data in the experimental and control arms

---

[1] Index *o* stands for *observed*.

respectively is given by

$$\pi_e = \frac{n_e - n_{eo}}{n_e} \quad \text{and} \quad \pi_c = \frac{n_c - n_{co}}{n_c}.$$

Let δ_e, δ_c be the average difference between observed and missing responses in the experimental and control arms. Then the average effect of intervention is

$$\begin{aligned}
\Delta &= \{(1 - \pi_e)\theta_e + \pi_e(\theta_e + \delta_e)\} - \{(1 - \pi_c)\theta_c + \pi_c(\theta_c + \delta_c)\} \\
&= (\theta_e - \theta_c) + (\delta_e\pi_e - \delta_c\pi_c),
\end{aligned} \tag{6.1}$$

where θ_e and θ_c are respectively the mean of the complete records analysis in the experimental and control arms.

From this we see that if the average difference between observed and missing outcomes is the same in the experimental and control group, then the difference between Δ and the complete records analysis $\theta_e - \theta_c$ depends on the difference in the proportions of missing data in the two arms.

When it is reasonable to believe that $\Delta \approx (\theta_e - \theta_c)$, then the study can be included in the meta-analysis in the usual way. When this is not reasonable, or if we wish to explore the sensitivity of our conclusions to this assumption, we need to specify a distribution for δ_e, δ_c. We suppose

$$\delta_e \sim N(\mu_e, v_e^2), \quad \delta_c \sim N(\mu_c, v_c^2), \quad \text{Cor}(\delta_e, \delta_c) = \rho. \tag{6.2}$$

Then, using the complete records, we have $(\hat{\theta}_e, s_e^2, \hat{\theta}_c, s_c^2)$ as estimates of treatment effects and variances as well as estimates of the probability of missing data in each arm $\hat{\pi}_e$ and $\hat{\pi}_c$. Substituting these estimates into (6.1), we have

$$\hat{\Delta} = (\hat{\theta}_e - \hat{\theta}_c) + (\mu_e\hat{\pi}_e - \mu_c\hat{\pi}_c), \tag{6.3}$$

with

$$\text{Var}(\hat{\Delta}) = s_e^2 + s_c^2 + V_1 + V_2. \tag{6.4}$$

Here, $(s_e^2 + s_c^2)$ represents the usual variance estimate, calculated using the complete records, and using the conditional variance formula

$$V_1 = v_e^2\hat{\pi}_e^2 + v_c^2\hat{\pi}_c^2 - 2\rho v_e v_c \hat{\pi}_e \hat{\pi}_c, \tag{6.5}$$

and

$$V_2 = (\mu_e^2 + v_e^2)\frac{\hat{\pi}_e(1 - \hat{\pi}_e)}{n_{eo}} + (\mu_c^2 + v_c^2)\frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_{co}}. \tag{6.6}$$

For reasonable sample sizes V_2 will negligible compared to V_1.

In order to use this approach, we need to specify (or obtain expert opinion on) $\mu_e, \mu_c, v_e^2, v_c^2$, and ρ. In practice obtaining expert opinion on plausible values of these parameters is often not possible. Given that our aim is to explore the sensitivity of the meta-analytic estimate of the treatment effect to different assumptions about the missing outcome data, we can instead use the approach below.

6.1.2   Sensitivity Analysis Strategies

Following [7], we suggest four choices for the sensitivity parameters $\mu_e, \mu_c, v_e^2, v_c^2$, and ρ, which explore sensitivity to four different features of the studies in the meta-analysis.

For each choice we (i) re-estimate all the treatment effects and standard errors for studies with missing data and (ii) re-fit the random effects meta-analysis to the resulting data. Note that because of the heterogeneity of missing data and its effects across studies, we do not typically consider the fixed effect model appropriate in this context. The four choices, or sensitivity strategies, are:

1. *Fixed Equal:* assume $\mu_e = \mu_c = \mu$, and $v_c^2 = v_e^2 = 0$, so that δ is common across all treatment arms within each study, and the same for all studies.

   As this analysis assumes the average difference between missing and observed outcomes is the same across all arms and studies, it is therefore sensitive to imbalance in missing data between treatment arms.
2. *Fixed Opposite:* assume $\mu_e = -\mu_c$, and $v_c^2 = v_e^2 = 0$, so that across all studies in the active arms observed and missing outcomes differ by δ, whereas they differ by $-\delta$ in the control arms.

   By assuming the differences between missing and observed arms are equal and opposite across the studies, this analysis is sensitive to the overall proportion of missing data in the studies.
3. *Random Equal:* assume $\delta_e = \delta_c = \delta$, with $\delta \sim N(\mu, v^2)$. That is, differences are common across arms within a study, but random across studies.

   This analysis builds on (1) by increasing the standard errors of study estimates according to the imbalance in the proportion of missing observations. Studies with unbalanced proportions of missing data are therefore further down-weighted in the subsequent meta-analysis.
4. *Random Uncorrelated:* assume δ_e independent of δ_c, i.e. $\rho = 0$. With this strategy it may be appropriate to build on strategy 2 by choosing $\mu_e = -\mu_c$. Relative to strategy 2, this analysis thus further increases the standard errors of study estimates according to the proportion of missing data in the study. Such studies are then further down-weighted in the subsequent meta-analysis.

```
> # 1. Do meta-analysis
> m <- metacont(Ne, Me, Se, Nc, Mc, Sc,
+               studlab=paste(author, year),
+               data=data1)
> # 2. Extract dataset
> mdata <- as.data.frame(m)
> mdata$studlab <- as.character(mdata$studlab)
> # 3. Select seven studies with smallest standard error
> data12 <- mdata[rank(mdata$seTE)<=7, c(7,1:6)]
> names(data12) <- c("study", "Ne", "Me", "Se", "Nc", "Mc", "Sc")
> # 4. Print dataset
> data12
                study Ne    Me    Se Nc    Mc    Sc
2           Boner 1989 20 15.70 13.10 20 22.70 16.47
5   DeBenedictis 1994a 17 14.40 11.10 17 27.40 17.30
12      Novembre 1994f 24 15.42  8.35 24 28.46 13.84
13      Novembre 1994s 19 11.00 12.40 19 26.10 14.90
14          Oseid 1995 20 14.10  9.50 20 28.90 18.00
16           Shaw 1985  8 10.27  7.02  8 34.43 10.96
17         Todaro 1993 13 10.10  8.90 13 23.50  4.00
```

Fig. 6.1 Code to read in data for the bronchoconstriction meta-analysis [5] and to extract studies with seven smallest standard errors

Example 6.1 We consider data from the meta-analysis introduced in Sect. 1.6, comparing Nedocromil sodium (experimental treatment) with placebo (control) for preventing exercise-induced bronchoconstriction. As before, the response is the maximum fall in the forced expiratory volume in 1 s (FEV$_1$) over the course of follow-up, expressed as a percentage. For each study, the mean value, standard deviation and sample size are reported for both experimental and control group.

To illustrate the four strategies for sensitivity analysis described above, we consider only data from the studies with the seven smallest standard errors. R code to create the analysis dataset and the data are shown in Fig. 6.1.

To explore the effect of missing data, we make eight patients missing in each arm of the `Novembre 1994f` study and five patients missing in each arm of the `Oseid 1995` study.

```
> data12$Nem <- rep(0, length(data12$Ne))
> data12$Nem[data12$study=="Novembre 1994f"] <- 8
> data12$Nem[data12$study=="Oseid 1995"] <- 5
> #
> data12$Ncm <- data12$Nem
> #
> data12$Neo <- data12$Ne - data12$Nem
> data12$Nco <- data12$Nc - data12$Ncm
```

Next, we conduct a meta-analysis using the observed number of patients (objects `Neo` and `Nco`) for the seven studies. Then we print the calculated mean differences and corresponding standard errors (objects `TE` and `seTE`) for the complete records analysis.

```
> mm1 <- metacont(Neo, Me, Se, Nco, Mc, Sc,
+                 data=data12, studlab=study)
> data12$TE   <- mm1$TE
> data12$seTE <- mm1$seTE
> data12[, c("study", "TE", "seTE", "Neo", "Nco")]
               study     TE     seTE Neo Nco
2           Boner 1989  -7.00 4.705693  20  20
5   DeBenedictis 1994a -13.00 4.985272  17  17
12      Novembre 1994f -13.04 4.040947  16  16
13      Novembre 1994s -15.10 4.447175  19  19
14          Oseid 1995 -14.80 5.255156  15  15
16            Shaw 1985 -24.16 4.601657   8   8
17          Todaro 1993 -13.40 2.706261  13  13
```

The random effects estimate for the complete report analysis is -14.22 with 95 % confidence interval of $[-17.20; -11.13]$. The funnel plot is shown in Fig. 6.2.   □
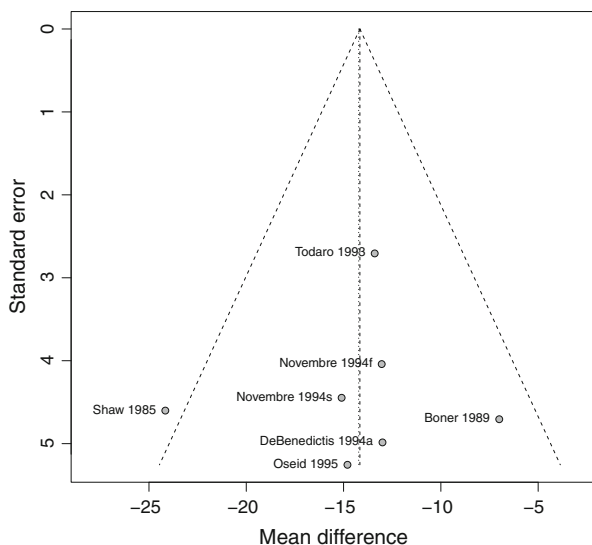
Fig. 6.2 Funnel plot for selected data from the bronchoconstriction meta analysis [5], with missing values introduced in `Novembre 1994f` and `Oseid 1995`

Before giving R code to implement each of the four strategies, we write an R function to calculate the standard error given in (6.4).

```
> semiss <- function(se, n.e, p.e, n.c, p.c,
+                     mu.e, nu.e, mu.c, nu.c, rho){
+   V1 <- function(p.e, p.c, nu.e, nu.c, rho)
+     nu.e^2*p.e^2 + nu.c^2*p.c^2 - 2*rho*nu.e*nu.c*p.e*p.c
+   V2 <- function(n.e, p.e, n.c, p.c, mu.e, nu.e, mu.c, nu.c)
+   (mu.e^2+nu.e^2)*p.e*(1-p.e)/n.e +
+     (mu.c^2+nu.c^2)*p.c*(1-p.c)/n.c
+   #
+   sqrt(se^2 +
+        V1(p.e, p.c, nu.e, nu.c, rho) +
+        V2(n.e, p.e, n.e, p.c,
+           mu.e, nu.e, mu.c, nu.c))
+}
```

6.1.3   Strategy 1: Fixed Equal

This strategy has a fixed difference between the mean of the observed and missing data, which is moreover the same in each arm. Thus, in the notation of (6.2), $\mu_e = \mu_c = \mu$, say, and $v_e^2 = v_c^2 = 0$. Looking back at (6.3) shows that this analysis explores sensitivity to different proportions of missing data in study arms. The variance component V_1 is zero for $v_e^2 = v_c^2 = 0$, see (6.5). Accordingly, the variance of the treatment estimate, (6.4), will only be slightly larger than for the complete records analysis.

Example 6.2 For this analysis, we choose $\mu_e = \mu_c = \mu = 8$, that is about half the treatment effect estimated from the observed data.

```
> # 1. Define parameters
> mu.e <- mu.c <- 8
> nu.e <- nu.c <- 0
> rho <- 0
> # 2. Calculate proportion missing in each study arm
> data12$Pe <- with(data12, Pe <- (Ne - Neo) / Ne)
> data12$Pc <- with(data12, Pc <- (Nc - Nco) / Nc)
> # 3. Calculate the mean effect for each study under strategy 1
> data12$TEs1 <- with(data12, TE + mu.e*Pe - mu.c*Pc)
> data12$seTEs1 <- with(data12,
+                       semiss(seTE, Neo, Pe, Nco, Pc,
+                              mu.e, nu.e, mu.c, nu.c, rho))
> # 4. Create indicator for studies with missings
> selmiss <- data12$study %in% c("Novembre 1994f", "Oseid 1995")
```

Using these commands, the following values are generated for the two studies
with missing values. Note, corresponding values are identical for studies with no
missing data.

```
> data12[selmiss, c("study", "TE", "seTE", "TEs1", "seTEs1")]
              study     TE      seTE    TEs1    seTEs1
12 Novembre 1994f -13.04 4.040947 -13.04 4.255236
14     Oseid 1995 -14.80 5.255156 -14.80 5.405244
```

As expected, since the proportion of missing data is the same in each arm, we
see no change in the means of the studies, but some increase in the variance.

Accordingly, results for a meta-analysis using these values are very similar to
those given above.

```
> mm1.s1 <- metagen(TEs1, seTEs1, data=data12,
+                     studlab=study, comb.fixed=FALSE)
> print(summary(mm1.s1), digits=2)
*** Output truncated ***
                                        95%-CI     z   p.value
Random effects model -14.23 [-17.7; -10.77] -8.05 < 0.0001
*** Output truncated ***
```

We discuss the results in more detail after performing the calculations for the
other strategies.   □

6.1.4   Strategy 2: Fixed Opposite

This strategy has a fixed difference between the mean of the observed and missing
data, but one that is of opposite sign in the intervention and control arms.

Example 6.3 As with strategy 1, we choose this to be 8 units, corresponding to
about half the estimated treatment effect from the observed data. For a conservative
analysis, we assume that in the intervention arm the missing patients had a score
that was on average 8 units greater (i.e. treatment less effective) than that observed,
while in the placebo arm the missing patients had a score that was on average 8 units
lower (i.e. treatment more effective) than that observed. Thus $\mu_e = 8, \mu_c = -8$ and
again $v_e^2 = v_c^2 = 0$. This strategy is therefore sensitive to the overall proportion of
missing data.

```
> # 1. Define parameters
> mu.e <- 8
> mu.c <- -mu.e
> nu.e <- nu.c <- 0
> rho <- 0
> # 2. Calculate the mean effect for each study under strategy 2
> data12$TEs2 <- with(data12, TE + mu.e*Pe - mu.c*Pc)
> data12$seTEs2 <- with(data12,
+                       semiss(seTE, Neo, Pe, Nco, Pc,
+                                mu.e, nu.e, mu.c, nu.c, rho))
```

```
> # 3. Print calculate values
> data12[selmiss, c("study", "TE", "seTE", "TEs2", "seTEs2")]
             study     TE     seTE       TEs2   seTEs2
12 Novembre 1994f -13.04 4.040947  -7.706667 4.255236
14     Oseid 1995 -14.80 5.255156 -10.800000 5.405244
> # 4. Do meta-analysis
> mm1.s2 <- metagen(TEs2, seTEs2, data=data12,
+                   studlab=study, comb.fixed=FALSE)
> print(summary(mm1.s2), digits=2)
Number of studies combined: k=7

                                         95%-CI    z  p-value
Random effects model -13.08 [-17.08; -9.09] -6.42 < 0.0001
*** Output truncated ***
```

As expected, this strategy markedly decreases the treatment benefit in the studies with missing data, while slightly increasing the standard errors.   □

6.1.5   Strategy 3: Random Equal

Here we set $\mu_e = \mu_c = \mu$, $v_e^2 = v_c^2 = v$ and $\rho = 1$. This strategy is thus similar to strategy 1, but more sensitive to imbalances in the proportion of missing data between study arms, since while (6.5) is always zero with strategy 1, it is only zero with this strategy if (as in our example) the proportion of missing data is the same in each arm.

Example 6.4 Again, we choose $\mu = 8$, and now we have to specify v^2, which we choose as 5, so that approximately 95 % of the time δ is in the interval $(3.5, 12.5)$.

```
> # 1. Define parameters
> mu.e <- mu.c <- 8
> nu.e <- nu.c <- sqrt(5)
> rho <- 1
> # 2. Calculate the mean effect for each study under strategy 2
> data12$TEs3 <- with(data12, TE + mu.e*Pe - mu.c*Pc)
> data12$seTEs3 <- with(data12,
+                    semiss(seTE, Neo, Pe, Nco, Pc,
+                          mu.e, nu.e, mu.c, nu.c, rho))
> # 3. Print calculate values
> data12[selmiss, c("study", "TE", "seTE", "TEs3", "seTEs3")]
             study     TE     seTE  TEs3   seTEs3
12 Novembre 1994f -13.04 4.040947 -13.04 4.271525
14     Oseid 1995 -14.80 5.255156 -14.80 5.416795
> # 4. Do meta-analysis
> mm1.s3 <- metagen(TEs3, seTEs3, data=data12,
+                   studlab=study, comb.fixed=FALSE)
> print(summary(mm1.s3), digits=2)
Number of studies combined: k=7

                                         95%-CI    z  p.value
```

```
Random effects model -14.24 [-17.7; -10.77] -8.05 < 0.0001
*** Output truncated ***
```

Since the proportion of missing data is exactly balanced in the two arms, variance component V_1, see (6.5), is zero, so in this example the results are almost the same as for strategy 1.   □

6.1.6 Strategy 4: Random Uncorrelated

In this strategy v_e^2 and v_c^2 are non-zero whereas correlation $\rho = 0$.

Example 6.5 Here we set $\mu_e = 8, \mu_c = -8, v_e^2 = v_c^2 = 5$ and $\rho = 0$. This explores the effect of the overall proportion of missing data on both the mean (as in strategy 2) but also the variance, since variance component V_1 in (6.5) is now non-zero, even with the same proportion of missing data in each arm, because $\rho = 0$.

```
> # 1. Define parameters
> mu.e <- 8
> mu.c <- -mu.e
> nu.e <- nu.c <- sqrt(5)
> rho <- 0
> # 2. Calculate the mean effect for each study under strategy 2
> data12$TEs4 <- with(data12, TE + mu.e*Pe - mu.c*Pc)
> data12$seTEs4 <- with(data12,
+                       semiss(seTE, Neo, Pe, Nco, Pc,
+                              mu.e, nu.e, mu.c, nu.c, rho))
> # 3. Print calculate values
> data12[selmiss, c("study", "TE", "seTE", "TEs4", "seTEs4")]
            study     TE     seTE       TEs4    seTEs4
12 Novembre 1994f -13.04 4.040947  -7.706667 4.399663
14     Oseid 1995 -14.80 5.255156 -10.800000 5.474182
> # 4. Do meta-analysis
> mm1.s4 <- metagen(TEs4, seTEs4, data=data12,
+                   studlab=study, comb.fixed=FALSE)
> print(summary(mm1.s4), digits=2)
Number of studies combined: k=7

                                   95%-CI      z  p.value
Random effects model -13.12 [-17.12; -9.12] -6.43 < 0.0001
*** Output truncated ***
```

                                                                                  □

6.1.7   Discussion of the Four Strategies

The results of the four strategies are summarised in Table 6.1, and four corresponding funnel plots are shown in Fig. 6.3.[2] As the proportion of missing data is the same in each study arm, strategy 1 (fixed equal) gives the same estimate of the mean as the complete records; however, the standard error is slightly inflated. This can also be seen by comparing Fig. 6.2 and the top left panel of Fig. 6.3. The corresponding random effects meta-analysis is virtually unchanged from the complete records analysis.

Strategy 2 (fixed opposite) results in a marked move towards the null for the point estimates from the studies with missing data, but no change in the standard

Table 6.1 Summary of results from the "complete records" (i.e. data remaining after the introduction of missing data), and under the four sensitivity analysis strategies

| | Complete records | Sensitivity strategy | | | |
		1: fixed equal	2: fixed opposite	3: random common	4: random opposite
Point estimate					
Boner 1989	−7.00	⋆	⋆	⋆	⋆
DeBenedictis 1994a	−13.00	⋆	⋆	⋆	⋆
Novembre 1994f	−13.04	−13.04	−7.71	−13.04	−7.71
Novembre 1994s	−15.10	⋆	⋆	⋆	⋆
Oseid 1995	−14.80	−14.80	−10.80	−14.80	−10.80
Shaw 1985	−24.16	⋆	⋆	⋆	⋆
Todaro 1993	−13.40	⋆	⋆	⋆	⋆
Standard error					
Boner 1989	4.71	⋆	⋆	⋆	⋆
DeBenedictis 1994a	4.99	⋆	⋆	⋆	⋆
Novembre 1994f	4.04	4.26	4.26	4.27	4.40
Novembre 1994s	4.45	⋆	⋆	⋆	⋆
Oseid 1995	5.26	5.41	5.41	5.42	5.47
Shaw 1985	4.60	⋆	⋆	⋆	⋆
Todaro 1993	2.71	⋆	⋆	⋆	⋆
Random effects meta-analysis					
Estimate	−14.22	−14.23	−13.08	−14.24	−13.12
Std. Err.	1.75	1.77	2.04	1.77	2.04
z-value	−8.13	−8.05	−6.42	−8.05	−6.43
τ^2	3.84	3.90	10.40	3.91	10.20

A "⋆" indicates results must be equal to the observed data for that study, as there are no missing data in that study

---

[2]R code to generate the funnel plots is given in the web-appendix.
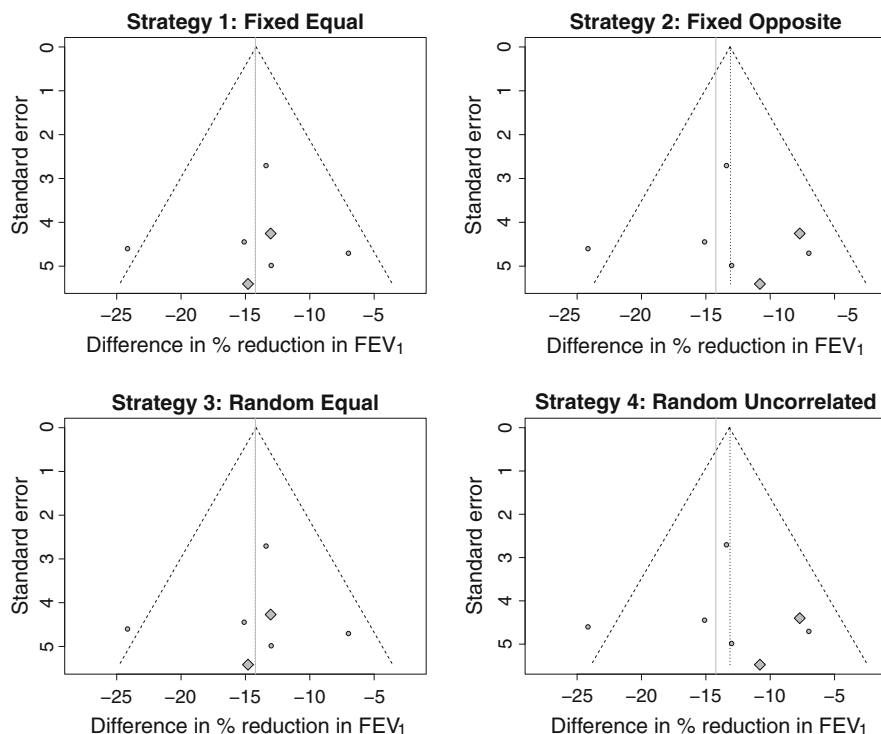
Fig. 6.3 Funnel plot showing results of strategy 1 (*top left*), strategy 2 (*top right*), strategy 3 (*bottom left*) and strategy 4 (*bottom right*). For ease of comparison with Fig. 6.2, studies with missing data are indicated by a *diamond*. The *solid line* is the random effects meta-analysis mean from the complete records

errors; compare Fig. 6.2 and the top right panel of Fig. 6.3. In the corresponding meta-analysis, this results in a marked increase in heterogeneity (although the heterogeneity statistic is still not significant at the 10 % level), slight reduction in the point estimate and increase in the standard error, and therefore a slight attenuation of the z-score.

Strategy 3 (random equal) would increase the standard error for the studies with missing data relative to strategy 1 if the proportion of missing data was different in the two arms; here it gives virtually identical results to strategy 1.

Lastly, strategy 4 (random opposite) increases the standard error of the estimates from studies with missing data relative to strategy 2 (bottom right panel Fig. 6.3). In these data this gives essentially the same results as strategy 2. If anything, the increased standard error for the studies with missing data downweights them in the random effects meta-analysis, resulting in a fractionally smaller estimate of τ^2, and slight increase in the absolute z-score, relative to strategy 2.

Overall, the test for heterogeneity does not reach significance at the 10 % level in any of the analyses, and the results from the complete records are robust to these

four sensitivity analyses, which give point estimates well within one standard error of the complete records analysis.

While the above example is artificial, it nevertheless illustrates the relative ease with which such sensitivity analyses can be performed, and how the different strategies can be used to highlight key features of the missing data and the corresponding impact on inferences about the treatment effect.

This approach may be applied to binary outcomes by replacing the mean effect and its standard error with the log odds ratio and its standard error. Such an approach clearly relies on the log odds being approximately normally distributed in the sample size at hand. For rare events, especially in smaller sample sizes, this may be inappropriate. Moreover, investigators may well be more comfortable discussing differences in the odds of an outcome between patients whose responses are observed and missing.

Thus [7] give details of how this approach may be applied to studies with binary outcomes. In this setting, μ_e, μ_c are replaced by informative missing odds ratios (IMORs). Paralleling the discussion above, the IMOR is the ratio of the odds of a good outcome in the missing data to that in the observed data. In this setting a closed formula for the study specific mean and variance of the log odds of success under strategies 3 and 4 above is more awkward; several options are discussed in the paper.

The approach outlined above is a two-stage approach: we first derive the impact of our assumptions about the missing data on the study specific estimates and standard errors, then follow this through to the meta-analysis. However, if desired a one stage process can be implemented, using a (generalised) linear mixed model. This approach has attractions in the context of rare events, and can allow for covariates and/or incorporating the sensitivity analysis into a wider evidence synthesis model. Details are given by [8].

6.2 Missing Precision

Here we consider the case where one or more studies we would like to include in a meta-analysis does not report the measure of precision. We assume that the number of patients enrolled is reported, but our multiple imputation approach can potentially extend to the scenario when both are missing.

We consider continuous outcome measures, since discrete data are typically modelled with the binomial or Poisson distribution where the variance is a known function of the mean.

6.2.1   Multiple Imputation Approach

We will tackle this problem using multiple imputation (MI), a practical and well-established tool for the analysis of partially observed data. A practical introduction to the key steps involved in MI is given by [2, pp. 37–46]. It involves the following steps:

1. choose a model for the missing data given the observed data;
2. taking full account of the uncertainty, draw the missing data from this model, creating M imputed datasets;
3. fit the substantive model to each of these M datasets, obtaining M point estimates and their associated standard errors, and
4. combine the results for final inference using Rubin's combination rules [2, p. 39].

This algorithm is very general, as is the applicability of Rubin's combination rules, so it provides a natural way to approach this problem.

Suppose that we have K studies, and each reports the estimated mean difference $\hat{\theta}_k = \hat{\theta}_{ke} - \hat{\theta}_{kc}$ and the number of participants n_{ke}, n_{kc}, $k = 1, \ldots, K$, where "e" denotes experimental arm and "c" control arm. However, suppose we only have the estimated standard error of the treatment difference for the first $K - 1$ studies, $s_1, \ldots, s_{K-1}$, as neither the estimated standard error s_K nor the estimated response variances $\hat{\sigma}_{Ke}^2, \hat{\sigma}_{Kc}^2$ are reported for study K.

Basic Idea of Multiple Imputation Algorithm

We propose a multiple imputation approach for this setting:

1. Using data from the studies where the standard error is reported, estimate the response variance for the experimental and control arms, $\hat{\sigma}_e^2, \hat{\sigma}_c^2$.
2. Apply multiple imputation as follows:

   a. Draw the missing study variances from an appropriate imputation distribution and calculate the study standard error.
   b. Fit the meta-analysis to the imputed data.
   c. Repeat steps 2a and 2b M times, and then summarise the results using Rubin's rules.

Further Details

Our approach rests on the assumption that the underlying variance of the response in the experimental and control arms, σ_e^2, σ_c^2, are each approximately common across studies, so that it is reasonable to assume that the standard error for the estimated treatment effect in study k is approximately $\sigma_e^2/n_{ke} + \sigma_c^2/n_{kc}$. When this is not appropriate, we need to identify a subset of studies which plausibly

have approximately the same response standard deviation as the study with missing precision, and apply the approach below in this subset.

We may also wish to explore whether the assumption of common underlying response variances σ_e^2, σ_c^2, across studies is plausible. This is most easily done by performing an F-test for equality of variances between one of the larger studies and all the other studies in turn. For example, if studies 1 and 2 have respectively n_{1c}, n_{2c} patients in the control arm with reported variances $\hat{\sigma}_{1c}^2$, $\hat{\sigma}_{2c}^2$, then under the null hypothesis of equal variances $\hat{\sigma}_{1c}^2/\hat{\sigma}_{2c}^2$ follows an F-distribution with $n_{1c}-1$, $n_{2c}-1$ degrees of freedom.

Provided that the assumption of a common underlying variance is reasonable, recall that in our setting the Kth study has missing standard error. Thus in the first step we estimate the value of σ_e^2 as

$$\hat{\sigma}_e^2 = \left[\sum_{k=1}^{K-1}(n_{ke}-1)\right]^{-1}\left\{\sum_{k=1}^{K-1}(n_{ke}-1)\hat{\sigma}_{ke}^2\right\},$$

and analogously estimate σ_c^2.

One option would be to set the missing $s_K^2 = \hat{\sigma}_e^2/n_{Ke} + \hat{\sigma}_c^2/n_{Kc}$, and then carry out the meta-analysis as usual. However, as this ignores any uncertainty in estimating s_K^2, treating it the same as all the reported standard errors, this pretends we have more information than is actually the case.

Instead, we use the multiple imputation approach outlined above. Under our assumptions we note that

$$\left\{\sum_{k=1}^{K-1}(n_{ke}-1)\hat{\sigma}_{ke}^2\right\} \approx \sigma_e^2 \chi^2_{\sum_{k=1}^{K-1}(n_{ke}-1)} \tag{6.7}$$

and

$$(n_{Ke}-1)s_{Ke}^2 \sim \sigma_e^2 \chi^2_{n_{Ke}-1}$$

and similarly for the control arm.

Let s_e^2 denote the left-hand side of (6.7) and define s_c^2 analogously. For imputation $m = 1, \ldots, M$ we

1. draw

$$\tilde{\sigma}_e^2 \sim s_e^2/\chi^2_{\sum_{k=1}^{K-1}(n_{ke}-1)} \quad \text{and} \quad \tilde{\sigma}_c^2 \sim s_c^2/\chi^2_{\sum_{k=1}^{K-1}(n_{kc}-1)}$$

2. and draw

$$\tilde{\sigma}_{em}^2 \sim \tilde{\sigma}_e^2 \frac{\chi^2_{n_{Ke}-1}}{n_{Ke}-1} \quad \text{and} \quad \tilde{\sigma}_{cm}^2 \sim \tilde{\sigma}_c^2 \frac{\chi^2_{n_{Kc}-1}}{n_{Kc}-1},$$

   and calculate $\tilde{s}_{Km}^2 = \tilde{\sigma}_{em}^2/n_{Ke} + \tilde{\sigma}_{cm}^2/n_{Kc}$.

3. Take $\tilde{s}_{Km}$ as the imputed standard error for the treatment estimate for study K and fit the meta analysis (fixed effect or random effects model), obtaining a pooled estimate $\hat{\theta}_{m*}$ with corresponding standard error s^2_{m*}.

We combine the M results for final inference using Rubin's rules [2, p. 39], so that

$$\hat{\theta}_* = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{m*}$$

and

$$\widehat{\mathrm{Var}}\,(\hat{\theta}_*) = s^2_w + \left(1 + \frac{1}{M}\right) s^2_b,$$

where

$$s^2_w = \frac{1}{M} \sum_{m=1}^{M} s^2_{m*} \quad \text{and} \quad s^2_b = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\theta}_{m*} - \hat{\theta}_*)^2.$$

For inference, we replace the normal distribution with a t_ν distribution with degrees of freedom ν given by

$$\nu = (M-1)\left[1 + \frac{s^2_w}{(1 + 1/M)s^2_b}\right]^2$$

In practice, $M = 10$ imputations are typically sufficient [2, p. 54], but if the results are close to statistical significance we may want to increase the number of imputations so we are not misled by Monte-Carlo variation.

We now give an example, and then briefly consider how we might proceed if only the overall treatment estimate is available in study K.

Example 6.6 To illustrate the multiple imputation approach, we once again use the bronchoconstriction meta-analysis used in Example 6.1. As in that example, we take the subset of seven studies with the smallest treatment effect standard errors.

We will set the standard deviations for the experimental and control arm to missing in Shaw 1985, impute them, and then compare the results of (1) the original (full) data analysis; (2) the analysis of excluding Shaw 1985 and (3) the analysis with the precision of Shaw 1985 imputed.[3]

```
> # 1. Select seven studies with smallest standard error
> data13 <- mdata[rank(mdata$seTE)<=7, c(7, 1:6)]
> names(data13) <- c("study", "Ne", "Me", "Se", "Nc", "Mc", "Sc")
> # 2. Set missing standard deviations for Shaw 1985 study
```

---

[3] See Fig. 6.1 for creation of dataset mdata.

```
> data13.noshaw <- data13
> data13.noshaw$Se[data13.noshaw$study=="Shaw 1985"] <- NA
> data13.noshaw$Sc[data13.noshaw$study=="Shaw 1985"] <- NA
> # 3. Print dataset
> data13.noshaw
                study  Ne    Me    Se Nc    Mc    Sc
2           Boner 1989 20 15.70 13.10 20 22.70 16.47
5   DeBenedictis 1994a 17 14.40 11.10 17 27.40 17.30
12      Novembre 1994f 24 15.42  8.35 24 28.46 13.84
13      Novembre 1994s 19 11.00 12.40 19 26.10 14.90
14          Oseid 1995 20 14.10  9.50 20 28.90 18.00
16            Shaw 1985  8 10.27    NA  8 34.43    NA
17         Todaro 1993 13 10.10  8.90 13 23.50  4.00
```

Looking at the data above suggests the standard deviation is markedly smaller for the `Todaro 1993` study, especially in the control arm. This observation is confirmed in the F-tests using `Boner 1989` as the reference study.

```
> # First study, i.e., Boner 1989, as reference group
> f.tests <- with(data13.noshaw,
+                 c(pf(Se^2/Se[1]^2, Ne-1, Ne[1]-1),
+                   pf(Sc^2/Sc[1]^2, Nc-1, Nc[1]-1)))
> # Drop first study (reference group)
> f.tests <- matrix(f.tests, ncol=2)[-1,]
> # Define row and column names and print results
> rownames(f.tests) <- data13.noshaw$study[-1]
> colnames(f.tests) <- c("pval.e", "pval.c")
> round(f.tests, 2)
                   pval.e pval.c
DeBenedictis 1994a   0.25   0.59
Novembre 1994f       0.02   0.21
Novembre 1994s       0.41   0.34
Oseid 1995           0.09   0.65
Shaw 1985              NA     NA
Todaro 1993          0.09   0.00
```

Overall, the results of the F-tests for equality of outcome variances against the variance in `Boner 1989` suggest that the variances are reasonably comparable. However, the `Todaro 1993` study has smaller variances, particularly in the control arm. Thus we will not consider data from `Todaro 1993` in imputing the precision of `Shaw 1985`. We set the number of imputations $M = 100$ in order to use the normal distribution, instead of the t-distribution for post-imputation inference.

```
> # Define analysis dataset miss
> miss <- data13.noshaw
> # Set number of imputations
> M <- 100
> # Create data frame to hold results of each imputation
> imp.shaw <- data.frame(seTE=rep(NA, M),
+                        TE.fixed=NA, seTE.fixed=NA,
+                        TE.random=NA, seTE.random=NA,
+                        tau=NA)
```

```
> # Set seed so results are reproducible
> set.seed(10)
> # Select studies for imputation
> selimp <- !(miss$study %in% c("Shaw 1985", "Todaro 1993"))
> selshaw <- miss$study=="Shaw 1985"
> # Form pooled estimate of variability:
> S2.e <- with(miss, sum((Ne[selimp]-1)*Se[selimp]^2))
> S2.c <- with(miss, sum((Nc[selimp]-1)*Sc[selimp]^2))
> # Calculate degrees of freedom:
> df.e <- sum(miss$Ne[selimp]-1)
> df.c <- sum(miss$Nc[selimp]-1)
> #
> miss$Se.shaw <- miss$Se
> miss$Sc.shaw <- miss$Sc
> #
> for (m in 1:M) {
+   # Draw sigma2.e, sigma2.c
+   sigma2.e <- S2.e / rchisq(1, df=df.e)
+   sigma2.c <- S2.c / rchisq(1, df=df.c)
+   # Draw standard deviations for Shaw 1985
+   sd.e.shaw <- sigma2.e *
+     rchisq(1, df=miss$Ne[selshaw]-1)/(miss$Ne[selshaw]-1)
+   sd.c.shaw <- sigma2.c *
+     rchisq(1, df=miss$Nc[selshaw]-1)/(miss$Nc[selshaw]-1)
+   # Store imputed standard error for illustrative funnel plot
+   imp.shaw$seTE[m] <- sqrt(sd.e.shaw/miss$Ne[selshaw] +
+                            sd.c.shaw/miss$Nc[selshaw])
+   # Meta-analysis of current imputed dataset
+   miss$Se.shaw[selshaw] <- sqrt(sd.e.shaw)
+   miss$Sc.shaw[selshaw] <- sqrt(sd.c.shaw)
+   #
+   m.shaw <- metacont(n.e=Ne, mean.e=Me, sd.e=Se.shaw,
+                      n.c=Nc, mean.c=Mc, sd.c=Sc.shaw,
+                      data=miss)
+   # Store results
+   imp.shaw$TE.fixed[m]     <- m.shaw$TE.fixed
+   imp.shaw$seTE.fixed[m]   <- m.shaw$seTE.fixed
+   imp.shaw$TE.random[m]    <- m.shaw$TE.random
+   imp.shaw$seTE.random[m]  <- m.shaw$seTE.random
+   imp.shaw$tau[m]          <- m.shaw$tau
+ }
```

Having performed the imputation, we now apply Rubin's rules to get the point estimates and standard errors. First, we calculate the between and within variances as well as the degrees of freedom for the fixed effect and random effects models.

```
> # Calculate between and within variances
> s2.b.fixed  <- var(imp.shaw$TE.fixed)
> s2.b.random <- var(imp.shaw$TE.random)
> s2.w.fixed  <- mean(imp.shaw$seTE.fixed^2)
> s2.w.random <- mean(imp.shaw$seTE.random^2)
> # Determine number of imputations
> M <- length(imp.shaw$TE.fixed)
> # Fixed effect estimate using multiple imputation
```

```
> TE.fixed.imp   <- mean(imp.shaw$TE.fixed)
> seTE.fixed.imp <- sqrt(var(imp.shaw$TE.fixed)*(1 + 1/M) +
+                        mean(imp.shaw$seTE.fixed^2))
> # Random effects estimate using multiple imputation
> TE.random.imp   <- mean(imp.shaw$TE.random)
> seTE.random.imp <- sqrt(var(imp.shaw$TE.random)*(1 + 1/M) +
+                         mean(imp.shaw$seTE.random^2))
> # Calculate degrees of freedom
> df.fixed  <- (M-1)*(1 + s2.w.fixed /((1+1/M)*s2.b.fixed) )^2
> df.random <- (M-1)*(1 + s2.w.random/((1+1/M)*s2.b.random))^2
> # Print degrees of freedom
> df.fixed
[1] 134360.2
> df.random
[1] 119359.6
```

In both cases we see that by choosing $M = 100$ imputations we have ensured that the reference t-distribution has sufficient degrees of freedom that the normal distribution can be used instead. Accordingly, we can calculate confidence limits based on the normal approximation.

```
> round(unlist(ci(TE.fixed.imp, seTE.fixed.imp))[1:5], 2)
    TE   seTE  lower  upper      z
-13.55   1.54 -16.56 -10.54  -8.83
> round(unlist(ci(TE.random.imp, seTE.random.imp))[1:5], 2)
    TE   seTE  lower  upper      z
-13.56   1.56 -16.61 -10.50  -8.70
```

Finally, we conduct a meta-analysis of the original data and create a funnel plot (see Fig. 6.4) to show the actual treatment effect standard error for Shaw 1985 and the imputed standard errors.

```
> # 1. Do meta-analysis of original data
> mm2 <- metacont(Ne, Me, Se, Nc, Mc, Sc,
+                 studlab=study, data=data13)
> TE.shaw   <- mm2$TE[mm2$studlab=="Shaw 1985"]
> seTE.shaw <- mm2$seTE[mm2$studlab=="Shaw 1985"]
> # 2. Generate funnel plot
> funnel(mm2,
+        xlim=c(-30, 0), ylim=c(max(imp.shaw$seTE), 0))
> # 2a. Label Shaw 1985 study
> text(TE.shaw-0.25, seTE.shaw-0.1, "Shaw 1985", cex=0.8, adj=1)
> # 2b. Add small triangles representing imputed standard errors
> set.seed(456) # Set seed so results are reproducible
> points(jitter(rep(TE.shaw,
+                    length(imp.shaw$seTE)), 0.5),
+        imp.shaw$seTE, pch=2, cex=0.2)
```

Figure 6.4 shows that the great majority of imputed standard errors are far greater than that actually observed for Shaw 1985. Looking back at the data (see Fig. 6.1), this makes sense, because all studies used to impute the missing standard deviations for Shaw 1985 had higher standard deviations for both experimental and control
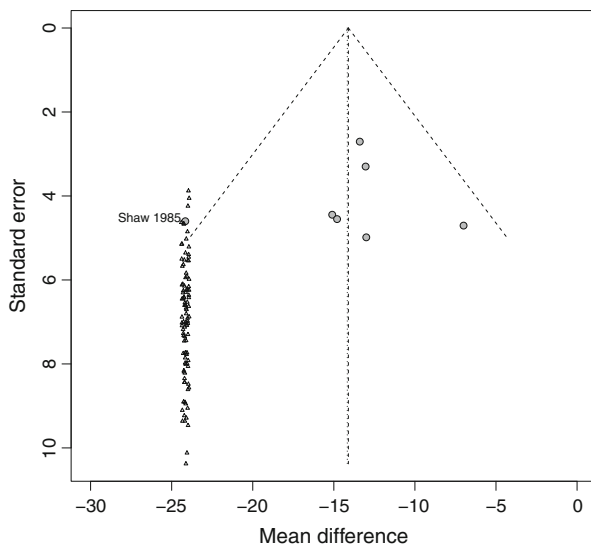
Fig. 6.4 Funnel plot of the seven studies, showing the range of imputed standard errors for `Shaw 1985`

Table 6.2 Results of fixed and random effects meta-analysis of full data, excluding `Shaw 1985` and using multiple imputation for the missing precision

	Fixed effect		Random effects	
Analysis	Estimate (std. err.)	z-score	Estimate (std. err.)	z-score
Full data (all seven studies)	−14.10 (1.47)	−9.56	−14.17 (1.67)	−8.50
Excluding `Shaw 1985`	−12.95 (1.56)	−8.32	−12.95 (1.56)	−8.32
Multiple imputation	−13.55 (1.54)	−8.83	−13.56 (1.56)	−8.70

arms. Note, the `jitter` function is used to separate overlapping points by adding a small random error to each value on the x-axis.

In Table 6.2, results for the fixed effect analysis show that, compared with the analysis of all seven studies, omitting `Shaw 1985` increases the point estimate by about 1 unit from −14.10 to −12.95, increases the standard error and thus results in a smaller z-score. This reflects the loss of information. After imputation, the point estimate is decreased (−13.55) towards the original, and the standard error slightly decreased. The z-score is larger than for the analysis omitting `Shaw 1985` but not as large as for the analysis of all seven studies. Thus, as we would hope, multiple imputation has successfully recovered some information, but not more than we would have had if we had seen the original data.

The picture for the random effects meta-analysis is slightly more complicated. When `Shaw 1985` is omitted, the fixed effect and random effects analysis are identical: not unexpectedly, as we are omitting the only study outside the 95 % confidence interval limits (diagonal dashed lines) in Fig. 6.4. This figure also

shows that, of the imputed standard errors, almost all put `Shaw 1985` inside the confidence limits. Thus the result that 87 of 100 imputed datasets result in $\hat{\tau}^2 = 0$ is not unexpected. Further, in this example, fitting the random effects meta-analysis if we increase the standard error of `Shaw 1985` over its actual value reduces $\hat{\tau}^2$ and hence increases the z-score for the treatment effect.

Taking these points together explains the random effect results in Table 6.2: after multiple imputation, as the large majority of the imputed datasets have larger precision values putting `Shaw 1985` within the confidence limits, the imputed estimate of τ^2 is lower than in the original meta-analysis, with the consequence that the z-score for the treatment effect is very slightly higher than for the original data.   □

6.2.2   Missing Participant Numbers

The above method provides a straightforward approach to including information from studies with missing precision, assuming that the mean effect is known within each arm as well as the number of patients in each arm.

If we only have the mean difference reported for a study and the study size n, then we can apply the method above to impute values of $\sigma_e^2 + \sigma_c^2$ for that study, from which we can calculate the treatment effect standard error

$$\sqrt{\frac{\sigma_e^2}{n/2} + \frac{\sigma_c^2}{n/2}},$$

assuming that there are $n/2$ patients in each arm of the study.

If, in addition the number of patients is unknown, then at each imputation step a value for this can be drawn at random from the other studies, or a subset of the studies expected to be of similar size. Of course, the more that has to be imputed, the less the information recovered by imputation will be.

6.3   Summary

In this chapter we have described how to approach the two main missing data issues in meta-analysis of summary data. The first, and most common, is that contributing studies have missing outcome data. Here, we may well wish to explore the robustness of our meta-analytic inferences to plausible departures from the assumptions underlying the primary study authors' analysis. The framework developed by White, Higgins and Wood [7], presented in Sect. 6.1, provides a practical approach for doing this. Less commonly, we may be missing a study precision. We have described a flexible multiple imputation approach for this issue.

We hope the examples and associated R code will enable readers to apply both these approaches in a wide range of settings.

References

1. Carpenter, J.R., Kenward, M.G.: Missing Data in Clinical Trials — A Practical Guide. National Health Service Co-ordinating Centre for Research Methodology, Birmingham (2008). http://www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03_JH17_MK.shtml
2. Carpenter, J.R., Kenward, M.G.: Multiple Imputation and Its Application. Wiley, Chichester (2013)
3. Carpenter, J.R., Roger, J.H., Kenward, M.G.: Analysis of longitudinal trials with protocol deviations: a framework for relevant, accessible assumptions and inference via multiple imputation. J. Biopharm. Stat. **23**, 1352–1371 (2013)
4. Schulz, K.F., Altman, G.D., Moher, D. for the CONSORT Group: Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. Br. Med. J. **340**, 1144–1146 (2010)
5. Spooner, C., Saunders, L.D., Rowe, B.H.: Nedocromil sodium for preventing exercise-induced bronchoconstriction. Cochrane Database Syst. Rev. (2002). doi:10.1002/14651858.CD001183
6. White, I., Carpenter, J., Evans, S., Schroter, S.: Eliciting and using expert opinions about non-response bias in randomised controlled trials. Clin. Trials **4**, 125–139 (2007)
7. White, I.R., Higgins, J.P.T., Wood, A.M.: Allowing for uncertainty due to missing data in meta-analysis—part 1: two-stage methods. Stat. Med. **27**, 711–727 (2008)
8. White, I.R., Welton, N.J., Wood, A.M., Ades, A.E., Higgins, J.P.T.: Allowing for uncertainty due to missing data in meta-analysis—part 2: hierarchical models. Stat. Med. **27**, 728–745 (2008)
9. White, I.R., Horton, N.J., Carpenter, J.R., Pocock, S.J.: Strategy for intention to treat analysis in randomised trials with missing outcome data. Br. Med. J. **342**, d40 (2011)

Chapter 7
Multivariate Meta-Analysis

In many clinical areas, there is no single, widely accepted outcome measure. For example, in rheumatoid arthritis three different outcome measures are in widespread use: the Health Assessment Questionnaire (HAQ) designed for measuring the severity of inflammatory joint disorders, the Disease Activity Score (DAS-28) to assess the level of disease activity and the American College of Rheumatology (ACR) response criteria. Similarly, in respiratory disease, we have (among other measures) Forced Expiratory Volume in one second (FEV$_1$), Forced Vital Capacity (FVC) and Peak Expiratory Flow (PEF).

Given data of this kind, it is natural to consider a joint meta-analysis of the various outcome measures simultaneously, especially when—as will typically be the case—not all the contributing studies have reported data on a common outcome scale.

7.1 Fixed Effect Model

We consider first the multivariate extension of the fixed effect model, and then the corresponding extension of the random effects model. For ease of presentation we describe the bivariate case, which extends directly if more outcomes are available.

Example 7.1 Lloyd et al. [8] report a meta-analysis of the effectiveness of anti-TNF-α inhibitors (e.g. etanercept, infliximab and adalimumab) in the treatment of rheumatoid arthritis. We take a non-random sample of five of these studies to illustrate the methods in this chapter. Figure 7.1 shows R code to read and print the data from these five studies. The data shown are mean change from baseline under treatment for two outcomes (DAS-28 and HAQ score), not comparisons between different arms. □

```
> # 1. Read in the data
> data14 <- read.csv("dataset14.csv")
> # 2. Print dataset
> data14
            author year mean.das se.das mean.haq se.haq
1           Bennet 2005    -1.7   0.25    -0.31   0.13
2          Bingham 2009    -1.6   0.10    -0.35   0.05
3       Bombardieri 2007   -1.9   0.05    -0.48   0.02
4 Navarro-Sarabia 2009    -1.1   0.18    -0.21   0.07
5    Van der Bijl 2008    -1.5   0.25    -0.21   0.08
```

Fig. 7.1 Illustrative subset of studies from the Lloyd data [1, 8] that report both DAS-28 score and HAQ score. Effects are mean change from baseline

Let $\hat{\boldsymbol{\theta}}_k = (\hat{\theta}_{k1}, \hat{\theta}_{k2})^T$ denote the bivariate outcome from study $k = 1, \ldots, K$. The multivariate fixed effect model is

$$\hat{\theta}_{k1} = \theta_{F1} + \epsilon_{k1}$$

$$\hat{\theta}_{k2} = \theta_{F2} + \epsilon_{k2}$$

$$\begin{pmatrix} \epsilon_{k1} \\ \epsilon_{k2} \end{pmatrix} \sim N \left\{ \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega}_k = \begin{pmatrix} \sigma_{k1}^2 & \sigma_{k12} \\ \sigma_{k12} & \sigma_{k2}^2 \end{pmatrix} \right\}, \tag{7.1}$$

where $\boldsymbol{\theta}_F = (\theta_{F1}, \theta_{F2})^T$ is the fixed effect parameter vector for the two outcomes.

As $\boldsymbol{\Omega}_k$ is unknown, it has to be estimated

$$\hat{\boldsymbol{\Omega}}_k = \begin{pmatrix} \hat{\sigma}_{k1}^2 & \hat{\sigma}_{k12} \\ \hat{\sigma}_{k12} & \hat{\sigma}_{k2}^2 \end{pmatrix}. \tag{7.2}$$

Whereas $\hat{\sigma}_{k1}^2$ and $\hat{\sigma}_{k2}^2$ are usually reported directly (or can be derived from confidence intervals), information on the covariance $\hat{\sigma}_{k12}$ is quite often not reported in publications. If this is the case, then we may need to use an estimate of the covariance obtained from elsewhere (perhaps a subset of studies on which individual participant data are available). When we do this, we should explore the robustness of our inferences to a range of plausible estimates of the covariance.

Let $\mathbf{W}_k = \hat{\boldsymbol{\Omega}}_k^{-1}$, then the fixed effect estimate is

$$\hat{\boldsymbol{\theta}}_F = \left(\sum_{k=1}^K \mathbf{W}_k \right)^{-1} \left(\sum_{k=1}^K \mathbf{W}_k \hat{\boldsymbol{\theta}}_k \right). \tag{7.3}$$

with corresponding covariance matrix

$$\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}_F) = \left(\sum_{k=1}^{K} \mathbf{W}_k\right)^{-1}. \tag{7.4}$$

The multivariate analogue of the heterogeneity statistic, Q, is defined as

$$Q = \sum_{k=1}^{K} (\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_F)^T \mathbf{W}_k (\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_F). \tag{7.5}$$

It can be shown that Q is approximately distributed as χ_j^2, where $j = (Kp) - 2$, where p is the number of outcomes. As in the univariate case, this can be used to evaluate whether there is more between study heterogeneity than would be expected under the fixed effect model.

Similarly, the I^2 measure of heterogeneity is defined as

$$I^2 = \frac{H^2 - 1}{H^2}, \text{ where } H^2 = \frac{Q}{j},$$

with j defined above.

Example 7.2   Figure 7.2 shows forest plots for both outcomes for the selection of the Lloyd data in Fig. 7.1. The following R code can be used to generate these figures using the `metagen` function and `forest.meta` function from R package **meta**.

```
> # Univariate meta-analysis of the DAS-28 outcome
> m.das <- metagen(mean.das, se.das,
+                   data=data14, sm="MD",
+                   studlab=paste(author, year),
+                   comb.random=FALSE)
> # Univariate meta-analysis of the HAQ outcome
> m.haq <- metagen(mean.haq, se.haq,
+                   data=data14, sm="MD",
+                   studlab=paste(author, year),
+                   comb.random=FALSE)
> forest(m.das, hetstat=FALSE)
> forest(m.haq, hetstat=FALSE)
```

We will compare univariate meta-analyses with bivariate meta-analyses. While, in this example, the data are change in outcome from baseline, the analysis proceeds in the same way if the data are estimated treatment differences from a randomised controlled trial. As the software does not handle treatment estimates from each arm in a randomised study directly, we need to derive estimated treatment differences (and associated standard errors) first.

Lloyd data: DAS-28 outcome

Study	TE	seTE	Mean difference	MD	95%–CI	W(fixed)
Bennet 2005	−1.7	0.25		−1.70	[−2.19; −1.21]	2.8%
Bingham 2009	−1.6	0.10		−1.60	[−1.80; −1.40]	17.8%
Bombardieri 2007	−1.9	0.05		−1.90	[−2.00; −1.80]	71.1%
Navarro−Sarabia 2009	−1.1	0.18		−1.10	[−1.45; −0.75]	5.5%
Van der Bijl 2008	−1.5	0.25		−1.50	[−1.99; −1.01]	2.8%
Fixed effect model				−1.79	[−1.87; −1.70]	100%

−2   −1   0   1   2

Lloyd data: HAQ outcome

Study	TE	seTE	Mean difference	MD	95%–CI	W(fixed)
Bennet 2005	−0.31	0.13		−0.31	[−0.56; −0.06]	1.8%
Bingham 2009	−0.35	0.05		−0.35	[−0.45; −0.25]	12.0%
Bombardieri 2007	−0.48	0.02		−0.48	[−0.52; −0.44]	75.3%
Navarro−Sarabia 2009	−0.21	0.07		−0.21	[−0.35; −0.07]	6.1%
Van der Bijl 2008	−0.21	0.08		−0.21	[−0.37; −0.05]	4.7%
Fixed effect model				−0.43	[−0.47; −0.40]	100%

−0.4 −0.2   0   0.2   0.4

Fig. 7.2 Forest plots for the DAS-28 and HAQ outcomes in Fig. 7.1

To fit the bivariate fixed effect meta-analysis outcome model, we use R package **mvmeta**, written by Antonio Gasparrini [3].[1] First, we make R package **mvmeta** available and print the list of arguments for the mvmeta function.

```
> library(mvmeta)
This is mvmeta 0.4.5. For an overview type: help('mvmeta-package').
> args(mvmeta)
function (formula, S, data, subset, method = "reml", bscov = "unstr",
 model = TRUE, contrasts = NULL, offset, na.action, control = list())
```

The first two arguments formula and S are mandatory. In the bivariate setting without covariates, argument formula is a $K \times 2$ matrix of estimated treatment effects

$$\begin{pmatrix} \hat{\theta}_{11} & \hat{\theta}_{12} \\ \vdots & \vdots \\ \hat{\theta}_{K1} & \hat{\theta}_{K2} \end{pmatrix}$$

---

[1]To install R package **mvmeta** use R command install.packages("mvmeta").

and argument S is a $K \times 3$ matrix of variances and covariances

$$\begin{pmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{112} & \hat{\sigma}_{12}^2 \\ \vdots & \vdots & \vdots \\ \hat{\sigma}_{K1}^2 & \hat{\sigma}_{K12} & \hat{\sigma}_{K2}^2 \end{pmatrix},$$

whose rows come from $\hat{\boldsymbol{\Omega}}_k$, $k = 1, \ldots, K$ [see (7.2)].

Note the Lloyd dataset does not include any information about the covariance or correlation between the outcomes. To illustrate the approach, we therefore explore various values of the correlation (since these are more intuitive than the corresponding covariances). As the mvmeta function requires the input of covariances, we have to calculate covariances from correlations recalling the association between covariances and correlations.[2] We write our own R function for this purpose:

```
> cor2cov <- function(sd1, sd2, cor) sd1*sd2*cor
```

For our first bivariate meta-analysis, we assume a correlation of 0 and create the necessary matrices.[3]

```
> theta <- cbind(data14$mean.das, data14$mean.haq)
> dimnames(theta) <- list(data14$author,
+                         c("mean.das", "mean.haq"))
> rho <- 0
> S.arth <- cbind(data14$se.das^2,
+                 cor2cov(data14$se.das, data14$se.haq, rho),
+                 data14$se.haq^2)
> dimnames(S.arth) <- list(data14$author,
+                         c("var.das", "cov", "var.haq"))
> S.arth
                 var.das cov var.haq
Bennet            0.0625   0  0.0169
Bingham           0.0100   0  0.0025
Bombardieri       0.0025   0  0.0004
Navarro-Sarabia   0.0324   0  0.0049
Van der Bijl      0.0625   0  0.0064
```

Using these matrices we conduct the multivariate meta-analysis with the mvmeta function.

```
> m.arth <- mvmeta(theta, S.arth, method="fixed")
> print(summary(m.arth), digits=2)
Call:  mvmeta(formula = theta ~ 1, S = S.arth, method = "fixed")

Multivariate fixed-effects meta-analysis
Dimension: 2

Fixed-effects coefficients
```

---

[2] $\mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$, so $\mathrm{Cov}(X, Y) = \mathrm{Cor}(X, Y)\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$.

[3] Assigning row and column names to matrices theta and S.arth using the base R function dimnames is optional, however it makes the printouts easier to follow!

```
          Estimate Std. Error       z Pr(>|z|)  95%ci.lb  95%ci.ub
mean.das    -1.79       0.04 -42.37     0.00     -1.87     -1.70 ***
mean.haq    -0.43       0.02 -24.89     0.00     -0.47     -0.40 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multivariate Cochran Q-test for heterogeneity:
Q = 51.70 (df = 8), p-value = 0.00
I-square statistic = 84.5%

5 studies, 10 observations, 2 fixed and 0 random-effects parameters
logLik    AIC     BIC
-11.12   26.24   26.85
```

The output from this model should be familiar from the results in earlier
chapters and the definitions given above. Notice that the univariate (see Fig. 7.2)
and multivariate results agree when the correlation is zero. We now experiment with
different correlations.

```
> # All studies have a correlation of 0.9 between outcomes
> rho <- 0.9
> S.arth2 <- cbind(data14$se.das^2,
+                  cor2cov(data14$se.das, data14$se.haq, rho),
+                  data14$se.haq^2)
> dimnames(S.arth2) <- dimnames(S.arth)
> # All studies have a correlation of -0.9 between outcomes
> rho <- -0.9
> S.arth3 <- cbind(data14$se.das^2,
+                  cor2cov(data14$se.das, data14$se.haq, rho),
+                  data14$se.haq^2)
> dimnames(S.arth3) <- dimnames(S.arth)
> # A mix of modest positive correlations between outcomes
> rho <- c(0.5, 0.2, 0.1, 0.6, 0.4)
> S.arth4 <- cbind(data14$se.das^2,
+                  cor2cov(data14$se.das, data14$se.haq, rho),
+                  data14$se.haq^2)
> dimnames(S.arth4) <- dimnames(S.arth)
> # Do meta-analyses
> m.arth2 <- mvmeta(theta, S.arth2, method="fixed")
> m.arth3 <- mvmeta(theta, S.arth3, method="fixed")
> m.arth4 <- mvmeta(theta, S.arth4, method="fixed")
```

We can use the generic `coef` and `vcov` functions to extract the fixed effect
estimates as well as the estimated covariance matrix from an object of class
`mvmeta`. The base R function `diag` is convenient to extract the corresponding
standard errors. For example, taking the results with an assumed correlation of 0.9
(i.e. R object `m.arth2`), we proceed as follows:

```
> # Fixed effect means
> round(coef(m.arth2), 3)
mean.das.(Intercept) mean.haq.(Intercept)
              -1.776               -0.427
> # Covariance matrix
```

Table 7.1 Results from univariate and bivariate fixed effect meta-analysis, with different assumed correlations

Method	Assumed outcome correlation	Fixed effect estimate (std. err.)	
		DAS-28	HAQ
Univariate `metagen`	0	−1.786 (0.0422)	−0.432 (0.0174)
Multivariate `mvmeta`	0	−1.786 (0.0422)	−0.432 (0.0174)
Multivariate `mvmeta`	0.9	−1.776 (0.0413)	−0.427 (0.0170)
Multivariate `mvmeta`	−0.9	−1.762 (0.0413)	−0.442 (0.0170)
multivariate `mvmeta`	mix	−1.805 (0.0416)	−0.439 (0.0171)

Estimates are change in mean score from baseline

```
> vcov(m.arth2)
                       mean.das.(Intercept)  mean.haq.(Intercept)
mean.das.(Intercept)          0.0017048052          0.0006286747
mean.haq.(Intercept)          0.0006286747          0.0002890715
> # Standard errors of fixed effect means
> round(sqrt(diag(vcov(m.arth2))), 4)
mean.das.(Intercept) mean.haq.(Intercept)
              0.0413               0.0170
```

Table 7.1 shows the results of univariate and bivariate fixed effect meta-analyses for both outcomes using data from our selection of five studies from the Lloyd dataset, given in Fig. 7.1. We see that the results of the univariate and multivariate meta-analysis agree when the correlation is zero, but differ by within one standard error as the correlation moves from 0.9 to −0.9. Note too, that for a common correlation, the standard errors vary with the absolute value of that correlation.

To complete this example we illustrate the analysis when $\Omega_k = \Sigma/n_k$, $k = 1, \ldots, K$, for a common response covariance matrix Σ. If we take

$$\Sigma = \begin{pmatrix} 2.146 & \rho\sqrt{2.146 \cdot 0.352} \\ \rho\sqrt{2.146 \cdot 0.352} & 0.352 \end{pmatrix}$$

and the sample sizes for the five studies as $(27, 188, 810, 68, 41)$, then the calculations are as follows:

```
> rho <- 0.9
> sample.sizes <- c(27,188,810,68,41)
> # We use matrix multiplication (of a 5-by-1 and a 1-by-3 matrix)
> # as a concise way of generating the argument S for mvmeta:
> S.arth.common <- matrix(1/sample.sizes, ncol=1) %*%
+    matrix(c(2.146, rho*sqrt(2.146*0.352), 0.352), nrow=1)
> dimnames(S.arth.common) <- list(data14$author,
+                          c("var.das", "cov", "var.haq"))
> round(S.arth.common, 4)
               var.das    cov var.haq
Bennet          0.0795 0.0290  0.0130
Bingham         0.0114 0.0042  0.0019
Bombardieri     0.0026 0.0010  0.0004
```

Table 7.2 Results from univariate and bivariate fixed effect meta-analysis, with common response covariance matrix Σ

Method	Assumed outcome correlation	Fixed effect estimate (std. err.)	
		DAS-28	HAQ
Univariate `metagen`	0.0	−1.786 (0.0422)	−0.432 (0.0174)
Multivariate `mvmeta`	0.9	−1.783 (0.0435)	−0.428 (0.0176)
Multivariate `mvmeta`	−0.9	−1.783 (0.0435)	−0.428 (0.0176)

Estimates are change in mean DAS-28, HAQ, score from baseline

```
Navarro-Sarabia  0.0316 0.0115  0.0052
Van der Bijl     0.0523 0.0191  0.0086
> # Conduct and print bivariate meta-analysis
> m.arth.common <- mvmeta(theta, S.arth.common, method="fixed")
> print(summary(m.arth.common), digits=2)
*** Output truncated ***
        Estimate Std. Error     z Pr(>|z|) 95%ci.lb 95%ci.ub
mean.das   -1.78       0.04 -40.99    0.00    -1.87    -1.70 ***
mean.haq   -0.43       0.02 -24.32    0.00    -0.46    -0.39 ***
*** Output truncated ***
5 studies, 10 observations, 2 fixed and 0 random-effects parameters
logLik    AIC     BIC
-0.22    4.43    5.04
```

Examination of the likelihood shows that, for a common response covariance matrix Σ, the univariate and multivariate fixed effect model results are very similar. The results in Table 7.2 confirm this.   □

The above results suggest that, when both outcomes are available on all studies, there is little to be gained from a multivariate meta-analysis. While, if the correlation is available for each study, the results may be fractionally more precise, this is unlikely to be practically important. Moreover, this gain comes because the outcomes are assumed to follow a multivariate normal distribution. This assumption cannot be checked from summary data alone, and the more outcomes there are, the less likely it is to be true. Likewise, if we are only interested in one outcome, which is observed on all studies, multivariate meta-analysis is unlikely to be useful.

However, there are two other general reasons why we might wish to perform multivariate meta-analysis. The first is that our ultimate summary is some function of the meta-analytic means of the multivariate outcomes, e.g. some $f(\theta_{F1}, \theta_{F2})$ in (7.1). For example, [3] consider multivariate meta-analysis of the non-linear relationship between temperature and mortality. Here, data from each of K studies (urban centres) results in K separate estimates of a set of p parameters describing the non-linear relationship between mortality and temperature, together with their covariance matrix. These parameters are then the multivariate responses in a subsequent p-dimensional meta-analysis; the set of p mean parameters estimated in this meta-analysis and the associated covariance matrix, are then used to provide the meta-analytic summary of the relationship between temperature and mortality.

The second reason is that many studies may not report all the outcomes. We consider this next.

7.2 Dealing with Unbalanced Data

It will often be the case that not all studies report all outcomes. For example, the five studies shown in Fig. 7.1, which report both DAS-28 and HAQ, are our non-random subset of the 21 studies in [8], the remaining 16 of which report one or other of DAS-28 and HAQ, but not both. In this case, a multivariate meta-analysis of the full data has the potential to reduce bias and gain precision. Naturally, this comes at the price of additional assumptions.

The first assumption concerns the mechanism driving which studies report which outcomes. This could be completely random, i.e. completely unrelated to the outcome values. In this case, the outcomes are said to be "Missing Completely At Random", and a univariate analysis is unbiased, but a multivariate meta-analysis may gain precision.

However, it could be that within a study the chance of a particular outcome being seen, or reported, *is* associated with its underlying potentially unseen value, *but* that this association is broken given the other observed, or reported, outcomes from that study. For example, it may be that study k was in a higher risk patient population, and that outcome A, which is more invasive to ascertain, is not really ethical to collect in this patient group. If the risk in the patient population is associated with values of outcome A, then the chance of reporting outcome A is associated with its value. However, suppose that outcomes B and C are less invasive, and are reported (irrespective of their values). Then it is plausible to assume (for we can never be sure) that the chance of outcome A being reported, *given* reported outcomes B and C, is no longer associated with the value of outcome A. This is an example of data being "Missing At Random" (see [2, Chap. 1] for more details). In this case, a univariate meta-analysis of outcome A will generally be biased, and an appropriate multivariate meta-analysis has the potential to both reduce bias and gain precision.

If we are prepared to assume that unreported outcomes are missing at random given reported outcomes, then we do not have to explicitly model the chance of seeing an outcome. Of course, we can never know if this assumption is correct, so in certain contexts we may need to be cautious in our inferences. Multiple imputation provides a natural way to explore the robustness of inferences to this assumption, for example applying the approaches described in [2, Chap. 10].

The second key assumption that the gains of a multivariate meta-analysis rest on is that the joint distribution of the summary outcomes from a study is multivariate normal. If the underlying outcomes themselves have a multivariate normal distribution it follows that the summary outcomes (e.g. means) will. However, even if the underlying outcomes are not normally distributed, the central limit theorem implies that—if the sample size is large enough—summary measures (means, log odds ratios, log risk ratios, log hazard ratios) will have a multivariate normal distribution.

Unfortunately, this assumption cannot be checked from summary data. In practice, we are reasonably comfortable with this assumption for study k if it has hundreds of participants, but not if it has tens of participants. If in doubt, we should explore the robustness of our conclusions to omitting study k.

We therefore proceed under the assumption that unreported outcomes are missing at random, and that within each study the outcomes (reported or not) are multivariate normal. As we describe below, this makes the analysis relatively straightforward. First though, since any reductions in bias and gains in precision are going to arise from pooling information, it makes sense to look at the data carefully.

Example 7.3 Returning to the extract from the Lloyd data, Fig. 7.3 shows funnel plots for the two outcomes of the five studies from the Lloyd data in Fig. 7.1. The following R code was used to generate this figure.

```
> funnel(m.das,
+        xlab="Mean difference in DAS-28 score from baseline")
> sel <- m.das$studlab %in% c("Bennet 2005", "Bombardieri 2007",
+                             "Navarro-Sarabia 2009")
> text(m.das$TE[sel]-0.015, m.das$seTE[sel]-0.002,
+      m.das$studlab[sel],
+      adj=1, cex=0.8)
> text(m.das$TE[!sel]+0.015, m.das$seTE[!sel]-0.002,
+      m.das$studlab[!sel],
+      adj=0, cex=0.8)
> funnel(m.haq,
+        xlab="Mean difference in HAQ score from baseline")
> text(m.haq$TE-0.005, m.haq$seTE-0.002,
+      m.haq$studlab,
+      adj=1, cex=0.8)
```
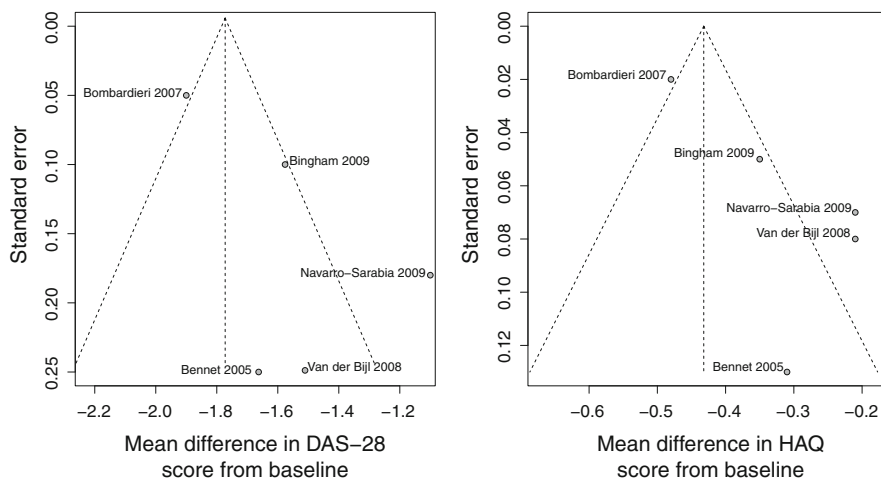

Fig. 7.3 Funnel plot of change in DAS-28 and HAQ scores for the five studies from the Lloyd data. Numbers correspond to study numbers in Fig. 7.1

We see that the `Bombardieri` study has considerably more information than the others, and that omitting the DAS-28 score for the `Navarro-Sarabia` study will have a marked effect on the overall DAS-28 mean.

We now show how to use R to obtain a bivariate plot of the outcomes, showing the 95 % confidence regions. To illustrate this, we assume a common correlation between the outcomes is 0.25 across all five studies. The R code is as follows[4]:

```
> # Load R library ellipse which provides R function ellipse
> library(ellipse)
> # Plot the study means, setting the x-limits and y-limits
> # so that the confidence regions will be visible
> plot(data14$mean.das, data14$mean.haq,
+      xlim=c(-2.5, -0.5), ylim=c(-0.7, 0), pch="+",
+      xlab="Mean difference in DAS-28 score from baseline",
+      ylab="Mean difference in HAQ score from baseline")
> # Add confidence regions
> rho <- 0.25
> with(data14,
+      for (i in seq(along=mean.das)){
+        S <- matrix(c(se.das[i]^2,
+                      se.das[i]*se.haq[i]*rho,
+                      se.das[i]*se.haq[i]*rho,
+                      se.haq[i]^2),
+                   byrow=TRUE, ncol=2)
+        lines(ellipse(S, centre=c(mean.das[i], mean.haq[i]),
+                      level=0.95 ), col="grey")
+      })
> # Add study labels
> studlab <- paste(data14$author, data14$year)
> # Study label from Van der Bijl needs a line break
> # in order not to overwrite another label
> bijl <- data14$author=="Van der Bijl"
> studlab[bijl] <- paste(data14$author[bijl], "\n",
+                        data14$year[bijl])
> text(data14$mean.das+0.015, data14$mean.haq, studlab, adj=0)
```

The resulting Fig. 7.4 again indicates that the DAS-28 score from the `Navarro-Sarabia` study is somewhat of an outlier. Notice that the common correlation ($\rho = 0.25$) does not imply a common covariance because the marginal variances are different for the different studies, hence the difference between the ellipses. We also see that if the DAS-28 score from the `Navarro-Sarabia` study were to be missing, then the multivariate meta-analysis would lean heavily on the `Bombardieri` study, whose estimates are much more precise. The figure also suggests that, were this outcome missing, multivariate meta-analysis may imply a lower value than the truth, consistent with the relationship between HAQ and DAS-28 score estimated from the studies with both outcomes. Below we therefore set the DAS-28 score from the `Navarro-Sarabia` study to missing, and explore how multivariate meta-analysis performs.   □

---

[4]To install the `ellipse` package use R command `install.packages("ellipse")`.
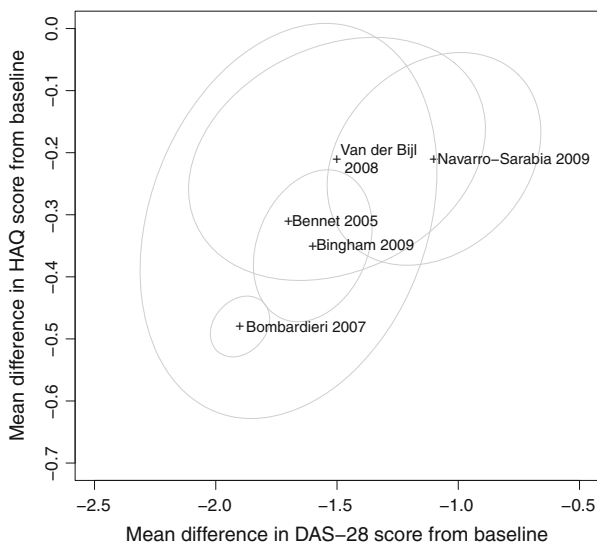
Fig. 7.4 Plot of change in DAS-28 and HAQ scores with 95 % confidence regions for the five studies from the Lloyd data

Under the two assumptions discussed above, the outcomes reported by each study can readily be included in the multivariate meta-analysis. Each study's contribution to the likelihood for the model is simply the marginal likelihood of its observed outcomes. For example, in the bivariate fixed effect model (7.1), if the first outcome is missing in study k, the likelihood contribution comes from the marginal model of the second outcome, see also Eq. (7.6):

$$\hat{\theta}_{k2} = \theta_{F2} + \epsilon_{k2}, \quad \epsilon_{k2} \sim N(0, \sigma_{k2}^2).$$

Another way to describe the handling of missing outcomes is as follows. We set

$$\hat{\boldsymbol{\theta}}_k = \begin{pmatrix} 0 \\ \hat{\theta}_{k2} \end{pmatrix} \text{ and } \mathbf{W}_k = \begin{pmatrix} 0 & 0 \\ 0 & 1/\hat{\sigma}_{k2}^2 \end{pmatrix} \tag{7.6}$$

and then apply Eq. (7.3) as before.

Calculation of the heterogeneity statistic, Q, is slightly more complicated in case of missing outcomes in studies. Suppose there are K studies, and that there are p outcomes. Suppose further that study k reports the subset of $p_k \leq p$ outcomes which we denote by the $p_k \times 1$ vector $\hat{\boldsymbol{\theta}}_k$ with corresponding $p_k \times p_k$ covariance matrix $\hat{\boldsymbol{\Omega}}_k$. Let $\mathbf{W}_k = \hat{\boldsymbol{\Omega}}_k^{-1}$ and $\boldsymbol{\theta}_{kF}$ be the $p_k \times 1$ vector of the corresponding fixed effect

estimates from (7.1). Then the heterogeneity statistic is defined as

$$Q = \sum_{k=1}^{K}(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_{kF})^T \mathbf{W}_k(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_{kF}). \tag{7.7}$$

It can then be shown that Q is approximately distributed as χ_j^2, where $j = (\sum_{k=1}^{K} p_k) - 2$.

For studies with missing outcomes, R package **mvmeta** automatically uses the appropriate marginal likelihood without any prompting from the user.

Example 7.4 We now set the DAS-28 score for the `Navarro-Sarabia` study to missing:

```
> theta.miss <- theta
> selnava <- rownames(theta.miss)=="Navarro-Sarabia"
> theta.miss[selnava, "mean.das"] <- NA
> theta.miss
                mean.das mean.haq
Bennet              -1.7    -0.31
Bingham             -1.6    -0.35
Bombardieri         -1.9    -0.48
Navarro-Sarabia       NA    -0.21
Van der Bijl        -1.5    -0.21
```

For the first bivariate meta-analysis we use `S.arth` defined above, which sets the outcome correlations to zero.

```
> m.arth.miss <- mvmeta(theta.miss, S.arth, method="fixed")
> print(summary(m.arth.miss), digits=2)
Call:  mvmeta(formula = theta.miss ~ 1, S = S.arth, method = "fixed")

Multivariate fixed-effects meta-analysis
Dimension: 2

Fixed-effects coefficients
         Estimate Std. Error      z Pr(>|z|) 95%ci.lb 95%ci.ub
mean.das    -1.83       0.04 -42.11     0.00    -1.91    -1.74 ***
mean.haq    -0.43       0.02 -24.89     0.00    -0.47    -0.40 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multivariate Cochran Q-test for heterogeneity:
Q = 36.34 (df = 7), p-value = 0.00
I-square statistic = 80.7%

5 studies, 9 observations, 2 fixed and 0 random-effects parameters
logLik    AIC     BIC
 -4.24  12.47   12.87
```

Table 7.3 Comparison of univariate and multivariate meta-analyses of the Lloyd data when the DAS-28 score from the `Navarro-Sarabia` study is set to missing

Data	Analysis	Assumed outcome correlation	Fixed effect estimate (std. err.)	
			DAS-28	HAQ
Full	Univariate	–	−1.786 (0.042)	−0.432 (0.017)
Full	Multivariate	(0.25,0.25,0.25,0.25,0.25)	−1.786 (0.042)	−0.432 (0.017)
Missing	Univariate	–	−1.826 (0.042)	−0.432 (0.017)
Missing	Multivariate	(0.25,0.25,0.25, – ,0.25)	−1.817 (0.043)	−0.431 (0.017)
Missing	Multivariate	(0.90,0.90,0.90, – ,0.90)	−1.783 (0.041)	−0.426 (0.017)
Missing	Multivariate	(0.00,0.00,0.90, – ,0.00)	−1.752 (0.040)	−0.418 (0.015)
Missing	Multivariate	(0.90,0.90,0.00, – ,0.90)	−1.849 (0.035)	−0.436 (0.015)

The output shows that five studies with nine observations were considered in the multivariate meta-analysis. Note, results for the HAQ score are identical to the univariate results reported above as we assume a correlation of 0.

Table 7.3 shows results comparing univariate and multivariate meta-analysis under a range of correlations. These may be readily performed by following the examples above; full R code is available in the web-appendix. The first row shows the univariate meta-analysis of the DAS-28 score when no data are missing, and the second row the result of a bivariate meta-analysis of the DAS-28 and HAQ score, again when no data are missing and the HAQ and DAS-28 scores are assumed to have a common correlation of 0.25 across studies. This is close to the value estimated by [1]. This relatively weak correlation is typical of many settings, and the univariate and bivariate analyses agree to three significant figures.

With the DAS-28 score missing from the `Navarro-Sarabia` study, the mean DAS-28 score from the univariate analysis decreases by about 1 standard error to −1.826. The multivariate meta-analysis again with correlation 0.25, which includes the HAQ outcome but not the DAS-28 outcome for the `Navarro-Sarabia` study, gives a mean DAS-28 of −1.817, which is fractionally lower. This illustrates that any bias reduction with a multivariate meta-analysis is likely to be small if the correlation is moderate. Notice that because the DAS-28 score from the `Navarro-Sarabia` study is missing, the correlation between outcomes for this study does not enter the likelihood.

The last three rows of Table 7.3 illustrate that—assuming unreported outcomes are missing at random—a relatively large correlation is needed for multivariate meta-analysis to correct any bias in a univariate meta-analysis. They also indicate what the analysis says about the missing outcomes is dominated by the information coming from the larger studies. Lastly, the low standard error (0.035) from the final analysis cautions that information is bought with assumptions.   □

7.3   Random Effects Model

The multivariate fixed effect meta-analysis model extends to include random effects in an analogous way to the univariate case. For simplicity, we again focus on the bivariate case. The random effects counterpart to model (7.1) is

$$\hat{\theta}_{k1} = \theta_{R1} + u_{k1} + \epsilon_{k1}$$

$$\hat{\theta}_{k2} = \theta_{R2} + u_{k2} + \epsilon_{k2}$$

$$\begin{pmatrix} u_{k1} \\ u_{k2} \end{pmatrix} \sim N \left\{ \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{T} = \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix} \right\}, \tag{7.8}$$

$$\begin{pmatrix} \epsilon_{k1} \\ \epsilon_{k2} \end{pmatrix} \sim N \left\{ \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega}_k = \begin{pmatrix} \sigma_{k1}^2 & \sigma_{k12} \\ \sigma_{k12} & \sigma_{k2}^2 \end{pmatrix} \right\}, \tag{7.9}$$

where $\boldsymbol{\theta}_R = (\theta_{R1}, \theta_{R2})^T$ is the random effects parameter vector for the two outcomes. Here $\mathbf{T}$ is the covariance matrix of the random effects (u_{k1}, u_{k2}) describing the between-study heterogeneity in the outcomes.

7.3.1   Fitting the Random Effects Model

Fitting the multivariate random effects model requires estimation of $\mathbf{T}$. This can be done by Maximum Likelihood (ML), Restricted Maximum Likelihood (REML), or a multivariate version of the DerSimonian–Laird method of moments estimator [6, 9]. All three options are available in R package **mvmeta** and we now describe them briefly.

Let $\mathbf{W}_k^\star = (\hat{\boldsymbol{\Omega}}_k + \mathbf{T})^{-1}$. The likelihood corresponding to (7.8) and (7.9) is

$$L(\boldsymbol{\theta}_F, \mathbf{T}) = -\frac{1}{2} \left[\sum_{k=1}^{K} \left\{ \log |\hat{\boldsymbol{\Omega}}_k + \mathbf{T}| + (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_R)^T \mathbf{W}_k^\star (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_R) \right\} \right] + c \tag{7.10}$$

with a normalising constant c. For studies with only a subset of the outcomes observed, their contribution to (7.10) is replaced by the appropriate marginal likelihood of the observed studies, as discussed on page 176. Estimates of $\boldsymbol{\theta}_R$ and $\mathbf{T}$ are obtained by maximizing the likelihood.

REML seeks to correct downward bias in the ML estimates of $\mathbf{T}$ when the number of studies, K, is small (as is often the case in meta-analysis). It does this by adding a penalty to the likelihood:

$$L_{\text{REML}} = L - \frac{1}{2} \log \left| \sum_{k=1}^{K} \mathbf{W}_k^\star \right|. \tag{7.11}$$

REML is usually preferable to ML in applications.

There are various method of moments estimators for $\mathbf{T}$. We outline the recent proposal by [6], which is implemented in R package **mvmeta**. This method is equivalent to the DerSimonian–Laird method in the univariate case.

The method of moments first requires us to calculate the fixed effect estimate, $\hat{\boldsymbol{\theta}}_F$ using Eq. (7.3). Then, setting $\mathbf{W}_k = \hat{\boldsymbol{\Omega}}_k^{-1}$, we calculate the $p \times p$ matrix

$$\hat{\mathbf{V}} = \sum_{k=1}^{K} \mathbf{W}_k (\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_F)(\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_F)^T. \tag{7.12}$$

The method of moments estimator is derived by evaluating the expectation, $\mathrm{E}[\hat{\mathbf{V}}]$, which in the case where each study reports all p outcomes is

$$\mathrm{E}[\hat{\mathbf{V}}] = (K-1)\mathbf{I}_p \left(\mathbf{W}_+ - \sum_{k=1}^{K} \mathbf{W}_k \mathbf{W}_+^{-1} \mathbf{W}_k \right) \mathbf{T}, \tag{7.13}$$

where $\mathbf{W}_+ = \sum_{k=1}^{K} \mathbf{W}_k$. To obtain the estimate $\hat{\mathbf{T}}$, we replace $\mathrm{E}[\hat{\mathbf{V}}]$ in (7.13) by its observed value (7.12) and solve for $\mathbf{T}$.

Once we have $\hat{\mathbf{T}}$, generalising from the univariate random effects model we set $\mathbf{W}_k^{\star} = (\hat{\mathbf{T}} + \hat{\boldsymbol{\Omega}}_k)^{-1}$ and have

$$\hat{\boldsymbol{\theta}}_R = \left\{ \sum_{k=1}^{K} \mathbf{W}_k^{\star} \right\}^{-1} \sum_{k=1}^{K} \mathbf{W}_k^{\star} \hat{\boldsymbol{\theta}}_k \tag{7.14}$$

with variance–covariance matrix

$$\mathrm{Var}\left[\hat{\boldsymbol{\theta}}_R\right] \approx \left\{ \sum_{k=1}^{K} \mathbf{W}_k^{\star} \right\}^{-1}. \tag{7.15}$$

When a subset of studies do not report all the outcomes, the formula for obtaining the method of moments estimator of $\mathbf{T}$ needs to be modified, using the ideas described for the fixed effect estimator (7.6). Details of the appropriate modification, and the extension to include covariates, are given by [6].

Example 7.5 We now illustrate the use of the `mvmeta` function to fit multivariate random effects meta-analysis. As before, we have to specify the correlation between the DAS-28 and HAQ outcomes.

First we assume a correlation of 0.25 for all the studies, and show the results of using all three methods to estimate $\mathbf{T}$.

```
> rho <- 0.9
> S.arth.r <- cbind(data14$se.das^2,
+                   cor2cov(data14$se.das, data14$se.haq, rho),
```

```
+                      data14$se.haq^2)
> m.arth.reml <- mvmeta(theta, S.arth.r, method="reml")
> m.arth.ml   <- mvmeta(theta, S.arth.r, method="ml")
> m.arth.mm   <- mvmeta(theta, S.arth.r, method="mm")
```

The use of argument `method="reml"` is optional for the REML method as this is the default for the `mvmeta` function. The REML method gives the following results

```
> print(summary(m.arth.reml), digits=2)
Call:  mvmeta(formula = theta ~ 1, S = S.arth.r, method = "reml")

Multivariate random-effects meta-analysis
Dimension: 2
Estimation method: REML

Fixed-effects coefficients
         Estimate Std. Error      z Pr(>|z|)  95%ci.lb  95%ci.ub
mean.das    -1.55       0.14 -11.45     0.00     -1.82     -1.29 ***
mean.haq    -0.33       0.06  -5.74     0.00     -0.44     -0.22 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Between-study random-effects (co)variance components
      Structure: General positive-definite
    Std. Dev   Corr
mean.das    0.27 mean.das
mean.haq    0.11      1

Multivariate Cochran Q-test for heterogeneity:
Q = 42.29 (df = 8), p-value = 0.00
I-square statistic = 81.1%

5 studies, 10 observations, 2 fixed and 3 random-effects parameters
logLik     AIC     BIC
  3.19    3.62    4.02
```

We see that, in addition to the output for the fixed effect model, the program returns the following:

1. Information on the structure of the variance covariance matrix of the random effects, which we call **T** above. By default this matrix is unstructured (argument `bscov="unstr"` in the `mvmeta` function) which results in the printout 'general positive-definite'.
2. The standard deviation of the variance components and their correlation. These give the estimated elements of **T** as:

$$\begin{pmatrix} 0.27^2 & 0.27 \cdot 0.11 \cdot 1 \\ 0.27 \cdot 0.11 \cdot 1 & 0.11^2 \end{pmatrix}$$

Table 7.4 Results from bivariate random effects meta-analysis of the Lloyd data, for ML, REML and MM estimates of **T**, taking within study correlation between outcomes as 0.25 for all studies

Method	Random effects estimate (std. err.)	
	DAS-28	HAQ
REML	−1.552 (0.136)	−0.330 (0.058)
ML	−1.558 (0.123)	−0.332 (0.052)
MM	−1.561 (0.143)	−0.323 (0.067)

Estimates are change in mean DAS-28 and HAQ score from baseline

In this example, we have assumed that the between-study correlation is 0.25. This leads to an estimated value of 1 for the correlation between the random effects for DAS-28 and HAQ (column `corr` in output). When we assume a higher between-study correlation, the estimated correlation between the random effects is no longer 1.

Results of bivariate meta-analyses based on the three different methods to estimate **T** are given in Table 7.4. We see that estimates obtained using ML have a noticeably smaller standard error than those obtained using REML. This is expected, as the ML estimate of **T** is downwardly biased (with the bias declining as the number of studies, K, increases). Since the number of studies in a meta-analysis is often quite small, the bias from an ML analysis can be quite marked. REML removes much of this bias, and is therefore preferable to ML. However, the standard errors from REML are also downwardly biased when the number of studies is small. This problem can be addressed using the "Kenward Roger" correction [7], but this is not available in R package **mvmeta**, version 0.4.5. In the absence of this, we prefer the results from the method of moments.

We now compare the results of the fixed effect model with various correlations (Table 7.1) with those from the random effects analysis in Table 7.5. R code for these calculations are available in the web-appendix. For each outcome correlation, the random effects estimates are closer to zero than the fixed effect estimates. This is because the random effects analysis gives more weight to the smaller studies, and as Fig. 7.3 shows these have estimates that are closer to zero. The random effects analysis also gives markedly larger standard errors because of the substantial heterogeneity in the data. In general, marked differences between the fixed effect and random effects analysis can be an indicator of funnel plot asymmetry and possible publication bias (Chap. 5); in this example all the analyses definitively reject the null hypothesis.

In some analyses it will be useful to estimate and plot the study specific random effects, here (u_{k1}, u_{k2}), $k = 1, \ldots K$. These are readily extracted. The function `blup` returns $\hat{\boldsymbol{\theta}} + \hat{\mathbf{u}}_k$, so to obtain the study specific random effects we need to subtract $\hat{\boldsymbol{\theta}}$.

```
> blup(m.arth.mm)
             mean.das   mean.haq
Bennet       -1.604610 -0.3435751
Bingham      -1.604856 -0.3436913
Bombardieri  -1.890980 -0.4786708
```

Table 7.5 Results from univariate and bivariate random effects meta-analysis of the Lloyd data, using a range of correlations and the MM estimates of **T** throughout

Method	Assumed outcome correlation	Random effects estimate (std. err.) DAS-28	HAQ
Univariate `metagen`	–	−1.582 (0.147)	−0.323 (0.066)
Multivariate `mvmeta`	0.00	−1.564 (0.140)	−0.320 (0.068)
Multivariate `mvmeta`	0.25	−1.561 (0.143)	−0.323 (0.067)
Multivariate `mvmeta`	0.90	−1.566 (0.147)	−0.325 (0.066)
Multivariate `mvmeta`	−0.90	−1.665 (0.084)	−0.273 (0.094)

Estimates are change in mean DAS-28 and HAQ score from baseline

```
Navarro-Sarabia -1.292164 -0.1961779
Van der Bijl    -1.411576 -0.2525105


> blup(m.arth.mm) - fitted(m.arth.mm)
                         y1          y2
                   mean.das    mean.haq
Bennet          -0.04377303 -0.02064998
Bingham         -0.04401904 -0.02076614
Bombardieri     -0.33014273 -0.15574565
Navarro-Sarabia  0.26867336  0.12674718
Van der Bijl     0.14926143  0.07041459
```

As expected from Fig. 7.3, we see that the residuals from `Navarro-Sarabia` and `Van der Bijl` are positive. The software will also calculate standard errors for the $\hat{\mathbf{u}}_k$'s, which are needed to produce a "caterpillar". See `help(blup)` for more details.   □

7.4 Discussion

In this chapter we have described how the generic fixed effect and random effects meta-analysis models can be extended to multiple responses, and shown how the R package **mvmeta** can be used to fit such models.

If study outcomes are plausibly multivariate normal, then theory suggests we will obtain estimates with better statistical properties [10]. As the analyses in this chapter illustrate, the practical gain depends on (1) whether information on the correlation between the outcomes is available and (2) whether each study contributing to the meta-analysis reports all the outcomes.

For the case when our outcome of interest is reported by all the contributing studies the results in this chapter suggest there is no practical benefit of a multivariate meta-analysis; essentially the same conclusion was reached by Riley et al. [10].

When different studies report different outcomes, then multivariate meta-analysis has potential. As usual in statistics, though, additional information is bought at the price of additional assumptions. We discussed these in Sect. 7.2 and they need to be critically appraised; see also [5]. Graphical exploration of the data is therefore essential. Care also needs to be taken with random effects modelling in these settings, as (particularly when there is some funnel plot asymmetry) information can be borrowed across studies in unexpected ways.

Unfortunately, in many applications the use of multivariate meta-analysis will be hindered by lack of information about the correlation between outcomes, at least for some studies. Appropriate pooling of correlation information across studies is possible, requires careful modelling and entails further assumptions (e.g. [1], who also consider different types of outcomes); in applications the potential for practically relevant gain needs to be assessed before these are undertaken. In practice, it may be simplest to investigate the sensitivity to a range of different outcome correlation values, particularly the largest and smallest contextually plausible values.

Just as univariate meta-analysis extends to meta-regression, so multivariate meta-analysis can be extended to include covariates (which may have fixed or random coefficients). This is possible with the `mvmeta` package, although we have not explored this here. This has an additional potential gain if not all studies report all outcomes, since covariates can increase the plausibility of the underlying missing at random assumption [2, Chap. 1]. Such covariates, which are associated both with the chance of outcomes being reported and their underlying values, should be included.

The number of recent papers on multivariate meta-analysis is testament to the recent research activity in this area [4]. The more individual patient data are available, the more attractive multivariate meta-analysis becomes; therefore its relevance is likely to increase with the passage of time.

References

1. Bujkiewicz, S., Thompson, J.R., Sutton, A.J., Cooper, N.J., Harrison, M.J., Symmons, D.P.M., Abrams, K.R.: Multivariate meta-analysis of mixed outcomes: a Bayesian approach. Stat. Med. **32**(22), 3926–43 (2013). doi:10.1002/sim.5831
2. Carpenter, J.R., Kenward, M.G.: Multiple Imputation and Its Application. Wiley, Chichester (2013)
3. Gasparrini, A., Armstrong, B., Kenward, M.G.: Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat. Med. **31**(29), 3821–3839 (2012). doi:10.1002/sim.5471
4. Jackson, D., Riley, R., White, I.R.: Multivariate meta-analysis: potential and promise. Stat. Med. **30**, 2481–2498 (2011)
5. Jackson, D., White, I.R., Riley, R.D.: Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat. Med. **31**(29), 3805–3820 (2012)
6. Jackson, D., White, I.R., Riley, R.R.: A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. Biom. J. **55**(2), 231–245 (2013). doi: 10.1002/bimj.201200152

7. Kenward, M.G., Roger, J.H.: Small sample inference for fixed effects from restricted maximum likelihood. Biometrics **53**, 983–997 (1997)
8. Lloyd, S., Bujkiewicz, S., Wailoo, A.J., Sutton, A.J., Scott, D.: The effectiveness of anti-TNF-α therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. Rheumatology (Oxford) **49**(12), 2313–21 (2010). doi:10.1093/rheumatology/keq169
9. Pinto, E.M., Willan, A.R., O'Brien, B.J.: Cost-effectiveness analysis for multinational clinical trials. Stat. Med. **24**(13), 1965–82 (2005). doi:10.1002/sim.2078
10. Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R.: An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat. Med. **26**, 78–97 (2013)

Chapter 8
Network Meta-Analysis

Network meta-analysis (also known as multiple treatment comparison or mixed treatment comparison) seeks to combine information from all randomised comparisons among a set of treatments for a given medical condition. It is therefore a key tool for evidence-based medicine [24] and is currently a very active research topic [1, 21, 31, 32].

Network meta-analysis is a generalisation of pairwise meta-analysis. It aims to answer, in a statistically principled way, the natural clinical question of how a number of existing treatments for a patient with a given diagnosis compare to each other. To do this, network meta-analysis combines direct and indirect evidence on treatment effects, as we explain in Sect. 8.1. However, problems of heterogeneity and potential inconsistency are ever present and potentially even more challenging than in pairwise meta-analysis. In Sect. 8.5 we therefore discuss graphical tools for presenting and understanding heterogeneity.

There are a variety of different methods and software for network meta-analysis [3]. In this chapter, we describe and illustrate a frequentist weighted least squares approach, described by Rücker [27, 28] and implemented in the R package **netmeta** [30]. This stable, computationally fast, and widely applicable approach is essentially equivalent to maximum likelihood estimation. Other approaches are described briefly in Appendix A.3.

8.1 Concepts and Challenges of Network Meta-Analysis

To introduce the ideas, suppose we wish to compare three treatments, say two active treatments A and B and a control C. Then each randomised comparison in a study having arms A and C provides a *direct* estimate of the difference of the treatment effects of A and C, measured on some scale, e.g., as a log odds ratio. Suppose we denote this $\hat{\theta}_{AC}^{\text{direct}}$. Other studies may provide information on the direct comparison

between treatment B and the same control C, denoted $\hat{\theta}_{BC}^{\text{direct}}$. Such studies provide *indirect* evidence for the comparison of A and B from the treatment difference $A - C$ and $B - C$ as follows:

$$\hat{\theta}_{AB}^{\text{indirect}} \;=\; \hat{\theta}_{AC}^{\text{direct}} \;-\; \hat{\theta}_{BC}^{\text{direct}} \tag{8.1}$$

with variance

$$\text{Var}\,(\hat{\theta}_{AB}^{\text{indirect}}) \;=\; \text{Var}\,(\hat{\theta}_{AC}^{\text{direct}}) \;+\; \text{Var}\,(\hat{\theta}_{BC}^{\text{direct}}). \tag{8.2}$$

Of course, in addition to this indirect evidence, we may have direct evidence from studies comparing A and B, denoted by $\hat{\theta}_{AB}^{\text{direct}}$. We wish to combine the direct and indirect evidence to get the most precise estimates of the treatment differences and associated standard errors. In order to do this we need to make some assumptions, principally

1. that the studies are independent, and
2. that the underlying effects are—in some sense—*consistent*. This is also known as transitivity assumption [31, 35].

We will see in Sect. 8.3.1 that the data available can be summarised in a *network graph*. Formally, a network is said to be consistent if the sum of direct treatment effects over all closed circuits in the graph is zero. In practice, consistency means that indirect evidence for the difference between any two treatments does not differ from the direct evidence. Consistency in our network of treatments A, B, and control C means that $\theta_{AB}^{\text{direct}} = \theta_{AB}^{\text{indirect}}$ for the comparison of A and B, $\theta_{AC}^{\text{direct}} = \theta_{AC}^{\text{indirect}}$ for the comparison of A and C, and $\theta_{BC}^{\text{direct}} = \theta_{BC}^{\text{indirect}}$ for the comparison of B and C. A consequence of this is that if data giving direct comparisons between A and B are available alongside data giving indirect comparisons, the extent of inconsistency, $\hat{\theta}_{AB}^{\text{direct}} - \hat{\theta}_{AB}^{\text{indirect}}$, can be assessed. Further, the proportion of the evidence coming from direct and indirect comparisons can be calculated and used to help interpret the results.

Example 8.1 Figure 8.1 shows a slightly more complex network with four treatments, named A, B, C and D. The treatments, called *nodes* in graph theory, which are joined with a line, called *edge* in graph theory, correspond to those for which direct evidence is available. Thus, we see that direct evidence is available for all comparisons except $A - C$ which must be estimated indirectly. Furthermore, for example, the comparison of A and D can draw on both direct information, and also indirect information following the paths $A \to B \to D$ and $A \to B \to C \to D$. Returning to the $A - C$ comparison without direct evidence, this is built up by combining information of both paths $A \to B \to C$ and $A \to D \to C$.   □

The statistical problem is to estimate the treatment differences, in our example $\theta_{AB}, \theta_{AC}, \theta_{AD}, \theta_{BC}, \theta_{BD}$ and θ_{CD}, and their standard errors from the available studies. To do this we shall assume in the first instance that all the direct and indirect effects

Fig. 8.1 A network with four
treatments, *A*, *B*, *C* and *D*.
Lines indicate we have data
from one or more studies
comparing the two treatments
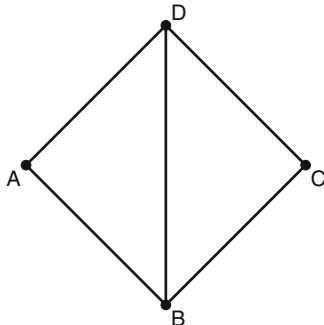
are consistent with each other and that differences we see in the data are due to
random error. We will explore how to assess this assumption and modify the analysis
when it may not be realistic.

8.2   Model and Estimation in Network Meta-Analysis

There has been a considerable literature on frequentist estimation of treatment
effects in network meta-analysis. The approach we now present draws on the
following references: [19, 22, 27, 33]. We begin by considering data consisting of
only two-arm trials. The extension to multi-arm studies is detailed on page 192.

Suppose there are n treatments, corresponding to the nodes in a network graph.
Suppose further we wish to estimate the n treatment effects, which we denote
by the $n \times 1$ vector $\boldsymbol{\theta}^{\text{treat}}$. To do this we have data from $K \geq (n-1)$ two-arm
studies,[1] which we denote by $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)$ with associated standard errors
$\mathbf{s} = (s_1, s_2, \ldots, s_K)$. Then our model is

$$\hat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\theta}^{\text{treat}} + \epsilon, \quad \epsilon \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{8.3}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix whose ith entry is s_i^2.

Note, $\hat{\boldsymbol{\theta}}$ is a $K \times 1$ vector in a network meta-analysis consisting only of two-arm
trials. In the situation of multi-arm studies, this vector is of dimension $m \times 1$ with m
denoting the total number of pairwise comparisons.

Example 8.2 Consider again the network shown in Fig. 8.1, and suppose we have
$K = 5$ studies each providing a single pairwise treatment comparison: $A-B, B-C,$
$C-D, A-D$ and $B-D$, i.e. the data consist of $m = 5$ pairwise treatment comparisons.
Then $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1^{AB}, \hat{\theta}_2^{BC}, \hat{\theta}_3^{CD}, \hat{\theta}_4^{AD}, \hat{\theta}_5^{BD})^T$, and $\boldsymbol{\theta}^{\text{treat}} = (\theta_A, \theta_B, \theta_C, \theta_D)^T$ is the vector of

---

[1]A minimum of $n-1$ two-arm studies is necessary to create a connected network graph with n
treatments (nodes).

treatment effects ($n = 4$). Based on this information, we can easily build the design matrix $\mathbf{X}$ and write down (8.3) for this special case:

$$\begin{pmatrix} \hat{\theta}_1^{AB} \\ \hat{\theta}_2^{BC} \\ \hat{\theta}_3^{CD} \\ \hat{\theta}_4^{AD} \\ \hat{\theta}_5^{BD} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \theta_D \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{pmatrix}$$

$$= \mathbf{X}\boldsymbol{\theta}^{\,\text{treat}} + \epsilon.$$

□

We note that, since—as usual in meta-analysis of summary data—we only have data on treatment differences, we cannot estimate each of the elements in $\boldsymbol{\theta}^{\,\text{treat}}$. However, as we describe in the next subsection, we can nevertheless estimate the fitted values from (8.3), and then use these to estimate the treatment comparisons and assess the extent of heterogeneity.

8.2.1   Fixed Effect Model

Here we describe details of the frequentist weighted least squares approach by Rücker [27]. Less technically minded readers may wish to skip to the worked example in Sect. 8.3.

We have already noted that (8.3) is over-parameterised, as we only have at most $(n-1)$ independent treatment comparisons, but the model has a parameter for each of the n treatments, $\boldsymbol{\theta}^{\,\text{treat}}$. Thus the matrix $\mathbf{X}$ is not of full rank, so its inverse does not exist, and we cannot obtain the weighted least squares estimates of $\boldsymbol{\theta}^{\,\text{treat}}$ directly.

We could re-parameterise the model so that $\mathbf{X}$ is of full rank. However, as Rücker [27] shows we can avoid this by taking the approach outlined below. This approach was originally developed using graph-theoretical methods that have a long history in electrical network theory [7, 9, 26, 27, 34]. In fact, it is also equivalent to weighted least squares methodology developed for experimental design as far back as the 1930s [4, 5, 25, 27, 33, 36]. Indeed, network meta-analyses are examples of incomplete block designs [33]. The difference from agricultural designs is that the decisions about which comparisons to estimate, i.e. trials to perform, are not made in advance with the network meta-analysis in mind!

As before, we denote by n the number of different treatments (nodes) in the network and let m be the number of pairwise treatment comparisons (and hence the

number of studies, since thus far we are only considering two-arm studies).[2] We number the treatments $1, \ldots, n$ and pairwise comparisons $1, \ldots, m$ in an arbitrary order which remains unchanged throughout the calculation.

Then, let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T$ and $\mathbf{s} = (s_1, \ldots, s_m)^T$ be the vectors of observed treatment differences and their standard errors, respectively. As is common in meta-analysis, we treat the standard errors as fixed.

As in (8.3), the network structure is defined by the design matrix $\mathbf{X}$ which is an $m \times n$ matrix. Here, if we only include two-arm studies, each row corresponds to a study and m is the number of studies. Then, as illustrated in Example 8.2, we have a "1" in the column that corresponds to the first treatment and a "−1" in the column that belongs to the second treatment. Each row of $\mathbf{X}$ should then sum to zero.

As previously mentioned, $\mathbf{X}^T\mathbf{X}$ is not of full rank, so to estimate the treatment effects we need to construct the Moore–Penrose pseudoinverse matrix [2, 28]. To do this, we define the $n \times n$ Laplacian matrix $\mathbf{L}$ (which plays a central role in graph theory) as

$$\mathbf{L} = \mathbf{X}^T\mathbf{W}\mathbf{X}, \tag{8.4}$$

where $\mathbf{W}$ is a diagonal matrix of dimension $m \times m$ whose diagonal elements are the inverse variance study weights $(1/s_1^2, \ldots, 1/s_m^2)$.

$\mathbf{L}$, the Laplacian matrix, has rank $n-1$ and is not invertible. However, its Moore–Penrose pseudoinverse $\mathbf{L}^+$ [10, 26] is defined and can be calculated by

$$\mathbf{L}^+ = (\mathbf{L} - \mathbf{J}/n)^{-1} + \mathbf{J}/n, \tag{8.5}$$

where $\mathbf{J}$ is the $n \times n$ matrix whose elements are all 1.

Estimation of Treatment Effects

Once we have $\mathbf{L}^+$, we can calculate our estimates of the fitted values $\hat{\boldsymbol{\theta}}^{nma}$ as

$$\hat{\boldsymbol{\theta}}^{nma} = \mathbf{X}\mathbf{L}^+\mathbf{X}^T\mathbf{W}\hat{\boldsymbol{\theta}},$$
$$= \mathbf{H}\hat{\boldsymbol{\theta}}, \tag{8.6}$$

where $\mathbf{H}$ is known as the *hat matrix* in regression. Equation (8.6) means that the elements of $\hat{\boldsymbol{\theta}}^{nma}$ (the network estimates) are linear combinations of the elements of $\hat{\boldsymbol{\theta}}$ (the observed estimates) with coefficients coming from the rows of $\mathbf{H}$. In other words, each network estimate is constituted by the observed estimates, weighted

---

[2]Note that if we allow multi-edge graphs, so, for example, if there are two studies comparing A and B we have two edges connecting nodes A and B, we may call m the number of *edges*.

with the elements of the corresponding row of $\mathbf{H}$. Accordingly, the elements of this row of $\mathbf{H}$ are interpreted as generalised weights.

We can also calculate the variance–covariance matrix of $\hat{\boldsymbol{\theta}}^{nma}$, $\mathbf{X}\mathbf{L}^+\mathbf{X}^T$. From this we can estimate all treatment contrasts and associated standard errors of interest. These estimates and standard errors are the same as those obtained by (weighted) maximum likelihood [25, 33, 36].

The hat matrix, $\mathbf{H}$, is a projection matrix which maps $\hat{\boldsymbol{\theta}}$ onto the consistent $(n-1)$-dimensional subspace. This gives the fitted values $\hat{\boldsymbol{\theta}}^{nma}$, which can be interpreted as the values that minimize the quadratic form

$$Q_{\text{total}} = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma})^\top \mathbf{W}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{nma}). \tag{8.7}$$

The Q_{total} statistic (8.7) therefore measures the extent of heterogeneity within the network. When all studies are two-arm studies, under the null hypothesis of no heterogeneity it is approximately χ^2-distributed with $m-(n-1)$ degrees of freedom. This therefore provides an approximate test of consistency. For a standard pairwise meta-analysis, Q_{total} corresponds to Cochran's Q statistic [12].

Variance Estimation

After fitting the network meta-analysis model, for any treatment comparison, we can calculate an estimate of the treatment effect using the direct evidence, and each piece of indirect evidence. The variance of the resulting treatment estimate is estimated by

$$V_{ij} = L_{ii}^+ + L_{jj}^+ - 2L_{ij}^+, \tag{8.8}$$

where V_{ij} denotes the variance of the resulting (potentially indirect) comparison of treatments i and j [7]. We note explicitly that by Eq. (8.8) variances are also estimated for pairs of treatments for which no direct comparison exists. Assuming that treatment estimates are consistent across the network, by using all the information in the data through the network meta-analysis, we obtain the most precise estimates of all comparisons.

Multi-Arm Studies

Usually, we have a number of multi-arm studies (i.e. studies with more than two treatment groups) to include in our network meta-analysis. We can do this most easily by including each multi-arm study in the dataset as a series of two-arm comparisons. However, the standard error of each two-arm comparison from a multi-arm study needs to be adjusted to reflect the fact that comparisons within multi-arm studies are correlated.

Consider a multi-arm study of p treatments with known variances. For this study, in the `netmeta` function, the user needs to supply treatment effects and standard errors for each of $p(p-1)/2$ possible comparisons. For instance, a three-arm study contributes three pairwise comparisons, $3(3-1)/2 = 3$, and a four-arm study six possible pairwise comparisons, $4(4-1)/2 = 6$.

We have to take care to account for within-study correlation before computing the weighted least squares estimator (8.6). To do this we inflate the standard errors for comparisons within each multi-arm study by back-calculation. Using these back-calculated standard errors in the weighted least squares estimator then gives results that correctly reflect the within-study correlation. To achieve this, we use a theorem by Gutman and Xiao [10, Theorem 7] to determine $\mathbf{L}_s^+$ and $\mathbf{L}_s = (\mathbf{L}_s^+)^+$ for each multi-arm study, also called sub-network s, with $s = 1, \ldots, S$, and S denoting the number of multi-arm studies, as described by [27]. The calculation proceeds as follows.

For multi-arm study s with $p_s > 2$ arms, for each of the $p_s(p_s-1)/2$ comparisons we have a variance for the comparison of each treatment contrast i and j ($i \neq j$), say v_{sij}^2. Let $\mathbf{V}_s$ be a $p_s \times p_s$ symmetric matrix with zeros on the diagonal and with (i,j) entry v_{sij}^2. Let $\mathbf{X}_s$ be the design matrix for the $p_s(p_s-1)/2$ comparisons in multi-arm study s, formed as in (8.3) above.

Then calculate

$$\mathbf{L}_s^+ = -\frac{1}{2p_s^2}\, \mathbf{X}_s^\top \mathbf{X}_s \mathbf{V}_s \mathbf{X}_s^\top \mathbf{X}_s, \tag{8.9}$$

and from this, calculate $\mathbf{L}_s = (\mathbf{L}_s^+)^+$ using (8.5). Denote the elements of $\mathbf{L}_s$ by l_{sij}. The adjusted variances for the comparison of treatment i and j are $-1/l_{sij}^{-1}$. Rücker et al. [28] show this method leads to results that are identical to those of the standard approach [19, 33].

Example 8.3 We illustrate the calculation for a network with a single four-arm study ($p_1 = 4$), so that $\mathbf{V} = \mathbf{V}_1$ has dimension 4×4. If v_{ij} is the variance of the comparison of treatment i and j, then we set

$$\mathbf{V} = \begin{pmatrix} 0 & v_{12} & v_{13} & v_{14} \\ v_{12} & 0 & v_{23} & v_{24} \\ v_{13} & v_{23} & 0 & v_{34} \\ v_{14} & v_{24} & v_{34} & 0 \end{pmatrix}.$$

Then,

$$\mathbf{X}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

and we calculate $\mathbf{L}_1^+$ as given by (8.9).   □

I-Squared for Network Meta-Analysis

The generalised heterogeneity statistic Q_{total} given in Eq. (8.7) is used to measure heterogeneity/inconsistency across the whole network. It can be partitioned in various ways, to help identify and understand the sources of heterogeneity. We return to this in Sect. 8.4.

We now consider the degrees of freedom of Q_{total}. Each p-arm study contributes $p - 1$ degrees of freedom to the total Q_{total} statistic. The total degrees of freedom are given by the sum of the degrees of freedom contributed by each study minus $n - 1$ (the number of treatments minus 1, which is the dimension of the consistent subspace). Denoting this by df, we can define a generalised I^2 statistic [12] as

$$I^2 = \max\left(\frac{Q_{\text{total}} - df}{Q_{\text{total}}}, 0\right). \tag{8.10}$$

8.2.2 Random Effects Model

A simple random effects model can be defined using the estimate of a common heterogeneity variance τ^2 for each pairwise treatment comparison. This is then added to each of the comparison variances, $s_i^2 + \hat{\tau}^2, i = 1, \ldots m$, before calculating $\mathbf{L}^+$ (8.5) which is used in (8.6). For multi-arm studies, the estimate $\hat{\tau}^2$ is added to the observed variance of each comparison before reducing the weights as described on page 192. Network meta-analysis is applied to the same observed treatment differences, now using the enlarged standard errors, as in standard pairwise meta-analysis.

In order to do this, we need an estimate of τ^2. For this we use a special case of the generalised DerSimonian–Laird estimate given in Jackson et al. [16] referred to in Sect. 7.3.1. It is estimated by

$$\hat{\tau}^2 = \max\left(\frac{Q_{\text{total}} - df}{tr\left((\mathbf{I} - \mathbf{H})\mathbf{UW}\right)}, 0\right), \tag{8.11}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix and *tr* denotes the trace of a matrix, which is the sum of its diagonal elements. The matrix $\mathbf{H}$ is defined in (8.6) and $\mathbf{W}$ is given in Sect. 8.2.1. $\mathbf{U}$ is obtained as a block diagonal matrix derived from the $m \times m$ matrix $\mathbf{X}\mathbf{X}^{\top}/2$ by selecting for each p-arm study a $p \times p$ block while setting the rest of the matrix elements to zero.

8.3   Using the R Package netmeta for Network Meta-Analysis

In this section we present an extended example using data from a network meta-analysis by Senn et al. [33] comparing different treatments for controlling blood glucose levels in patients with diabetes. This dataset comes with the R package **netmeta**.[3]

Example 8.4 To load the data, use the R code shown in Fig. 8.2. Patients enrolled in studies included in this dataset were treated with one of ten diabetes treatments, designed to reduce blood glucose levels. The effect measure was the mean difference of average plasma glucose concentration, referred to as HbA_{1c} and measured in mmol/mol. The full names of the various treatments can be obtained by typing

```
> help("Senn2013")
```

at the command line.

Variable `TE` contains the pairwise treatment effect comparing treatments `treat1` and `treat2`; variable `seTE` is the corresponding standard error. For example, the `DeFrozo1995` study is comparing metformin (`metf`) and placebo (`plac`). The average plasma glucose concentration is larger in the placebo group accordingly the mean difference is negative, -1.90.

The dataset contains one three-arm study (`Willms1999` in rows 3, 27, 28). The `netmeta` function which is described in the next subsection requires as input all pairwise comparisons for multi-arm studies. Therefore, we now show how the necessary information can be extracted from a publication using the `Willms1999` study as an example. This study reports sample sizes, group means, and corresponding standard deviations.

```
> willms <- data.frame(treatment=c("metf", "acar", "plac"),
+                      n=c(29, 31, 29),
+                      mean=c(-2.5, -2.3, -1.3),
+                      sd=c(0.862, 1.782, 1.831),
+                      stringsAsFactors=FALSE)
> willms
  treatment  n mean    sd
1      metf 29 -2.5 0.862
2      acar 31 -2.3 1.782
3      plac 29 -1.3 1.831
```

---

[3]To install R package **netmeta** use R command `install.packages("netmeta")`.

Using the `metacont` function we can calculate all three pairwise treatment comparisons:

```
> comp12 <- metacont(n[1], mean[1], sd[1], n[2], mean[2], sd[2],
+                    data=willms, sm="MD")
> comp13 <- metacont(n[1], mean[1], sd[1], n[3], mean[3], sd[3],
+                    data=willms, sm="MD")
> comp23 <- metacont(n[2], mean[2], sd[2], n[3], mean[3], sd[3],
+                    data=willms, sm="MD")
```

Next, we extract mean differences and corresponding standard errors from R objects `comp12`, `comp13` and `comp23`

```
> TE <- c(comp12$TE, comp13$TE, comp23$TE)
> seTE <- c(comp12$seTE, comp13$seTE, comp23$seTE)
```

and define R objects

```
> treat1 <- c(willms$treatment[1], willms$treatment[1],
+             willms$treatment[2])
> treat2 <- c(willms$treatment[2], willms$treatment[3],
+             willms$treatment[3])
```

with information on the two treatments.

These R objects can be combined in an R data set

```
> data.frame(TE, seTE=round(seTE, 4), treat1, treat2,
+            studlab="Willms1999")
    TE   seTE treat1 treat2    studlab
1 -0.2 0.3579   metf   acar Willms1999
2 -1.2 0.3758   metf   plac Willms1999
3 -1.0 0.4669   acar   plac Willms1999
```

which contains exactly the same information as given in rows 3, 27 and 28 of Fig. 8.2.

Note, a shortened version of this R code can be used to calculate the treatment comparison for a two-arm study reporting sample sizes, group means and standard deviations. Likewise, the code can be altered in order to calculate treatment comparisons for binary data using the `metabin` function instead of the `metacont` function.[4]

8.3.1   Basic Analysis and Network Plots

The `netmeta` function has the following arguments:

```
> args(netmeta)
function (TE, seTE, treat1, treat2, studlab, data = NULL,
```

---

[4]After finishing the book manuscript, a new R function `pairwise` to conduct these calculations automatically has been added to R package `meta`.

```
> # 1. Make R package netmeta available
> library(netmeta)
Loading required package: meta
Loading 'meta' package (version 4.0-2).
Loading 'netmeta' package (version 0.6-0).
> # 2. Load dataset
> data(Senn2013)
> data15 <- Senn2013
> # 3. Print dataset
> data15
      TE   seTE treat1 treat2           studlab
1  -1.90 0.1414   metf   plac       DeFronzo1995
2  -0.82 0.0992   metf   plac         Lewin2007
3  -0.20 0.3579   metf   acar         Willms1999
4  -1.34 0.1435   rosi   plac       Davidson2007
5  -1.10 0.1141   rosi   plac  Wolffenbuttel1999
6  -1.30 0.1268   piog   plac        Kipnes2001
7  -0.77 0.1078   rosi   plac        Kerenyi2004
8   0.16 0.0849   piog   metf       Hanefeld2004
9   0.10 0.1831   piog   rosi         Derosa2004
10 -1.30 0.1014   rosi   plac          Baksi2004
11 -1.09 0.2263   rosi   plac     Rosenstock2008
12 -1.50 0.1624   rosi   plac           Zhu2003
13 -0.14 0.2239   rosi   metf          Yang2003
14 -1.20 0.1436   rosi   sulf Vongthavaravat2002
15 -0.40 0.1549   acar   sulf         Oyama2008
16 -0.80 0.1432   acar   plac         Costa1997
17 -0.57 0.1291   sita   plac     Hermansen2007
18 -0.70 0.1273   vild   plac         Garber2008
19 -0.37 0.1184   metf   sulf          Alex1998
20 -0.74 0.1839   migl   plac      Johnston1994
21 -1.41 0.2235   migl   plac     Johnston1998a
22  0.00 0.2339   rosi   metf           Kim2007
23 -0.68 0.2828   migl   plac     Johnston1998b
24 -0.40 0.4356   metf   plac Gonzalez-Ortiz2004
25 -0.23 0.3467   benf   plac        Stucci1996
26 -1.01 0.1366   benf   plac        Moulin2006
27 -1.20 0.3758   metf   plac         Willms1999
28 -1.00 0.4669   acar   plac         Willms1999
```

Fig. 8.2 R code to load the diabetes example and view the data. For each treatment comparison (columns treat1 and treat2) in the network, the estimated treatment effect on HbA_{1c}, TE, and its standard error, seTE, are shown. The right-hand column shows the study labels. Note Willms1999 (rows 3, 27, 28) is a three-arm study

```
subset = NULL, sm = "", level = 0.95, level.comb = 0.95,
comb.fixed = TRUE, comb.random = FALSE, reference.group = "",
all.treatments = NULL, seq = NULL, tau.preset = NULL,
title = "", warn = TRUE)
```

Of these, the first four arguments: TE, seTE, treat1, and treat2 are mandatory. However, the study label argument studlab will be used in almost all analyses, and is essential for telling the software if there are multi-arm studies in the network. As with the Willms1999 study in Fig. 8.2, the treatment comparisons within a multi-arm study must have exactly the same study label, or the software will not be able to link them. A further important point is that it does not matter what order we put each study's treatment comparisons in; the software will re-order them (and change the sign of the treatment effect) as appropriate.

We can illustrate these points as follows.

```
> mn0 <- netmeta(TE, seTE, treat1, treat2, data=data15)
Warning messages:
1: In netmeta(TE, seTE, treat1, treat2, data = data15) :
  No information given for argument 'studlab'. Assuming that
  comparisons are from independent studies.
2: In netmeta(TE, seTE, treat1, treat2, data = data15) :
  Treatments within a comparison have been re-sorted in
  increasing order.
```

The first warning tells us that we did not give the study labels, and that the software is assuming that each row of the data matrix data15 comes from a different two-arm study. This is incorrect for this example, as we have one three-arm study, Willms1999. The second warning simply says that the software has re-ordered the treatment comparisons appropriately before performing any analyses.

As usual, we can print the object nm0 by entering it at the command line:

```
> mn0
Original data:

   treat1 treat2    TE    seTE
1    metf   plac -1.90  0.1414
2    metf   plac -0.82  0.0992
3    acar   metf  0.20  0.3579
4    plac   rosi  1.34  0.1435
*** Output truncated ***
Number of studies: k=28
Number of treatments: n=10
Number of pairwise comparisons: m=28
*** Output truncated ***
```

Comparing with Fig. 8.2, we see the number of studies is wrongly reported as 28, instead of 26. Furthermore, again comparing with Fig. 8.2, we see the fourth comparison (study Davidson2007) has been reversed, and the treatment effect correctly changed to 1.34.

The argument sm tells the software which summary measure to use for displaying the results of the network meta-analysis. Note that the treatment effects supplied to the software must be on the scale on which the analysis should be performed (e.g. log odds ratios, not odds ratios). For example, if log odds ratios are used as input to the netmeta function, odds ratios will be shown in printouts as well as forest plots if argument sm="OR" is used; otherwise, log odds ratios will be shown. The default

is `sm=""`, i.e. no information on the underlying summary measure is provided and thus no back-transformation is used in printouts and forest plots.

In many cases, the user will want to specify the reference group for making treatment comparisons. To do this, the argument `reference.group` (abbreviated `ref` or `r`) allows the user to specify a reference group. For example with the diabetes data we may specify `r="plac"`. To see all possible contrasts in the form of an effect matrix, the argument `all.treatments` (abbreviated `all` or `a`) is set to `TRUE`. This is the default.

The remaining arguments are the same as those we have met earlier in standard pairwise meta-analysis, and are discussed in detail in Part II. By default, only the results of a fixed effect network meta-analysis are printed. In order to also show results of a random effects analysis the argument `comb.random=TRUE` needs to be specified.

Before proceeding, we need to repeat the analysis, this time including the study labels. We also take the opportunity to specify the summary measure as mean difference (`MD`):

```
> mn1 <- netmeta(TE, seTE, treat1, treat2, studlab,
+                data=data15, sm="MD")
*** Warning message on study reordering omitted ***
```

8.3.2   A First Network Plot

We now show how to create a graphical representation of the network using the `netgraph` function.

```
> netgraph(mn1, seq=c("plac", "benf", "migl", "acar", "sulf",
+                      "metf", "rosi", "piog", "sita", "vild"))
```

The results are shown in Fig. 8.3. Note the use of the argument `seq` to specify the order of the sequence in which the treatments are shown anti-clockwise around the perimeter of the circle.[5] Treatments that are directly compared in at least one study are connected by a line. The thickness of this line is proportional to the inverse standard error of the direct treatment effect obtained using data from all studies which compared the two treatments. Thus, the thicker the line, the smaller the standard error, and the greater the evidence for that comparison. Of course, as this does not show the estimated treatment effect, this does not represent the statistical significance of the comparison.

From time to time we may wish to highlight a particular comparison, and this can be done with the argument `highlight`. We explore additional capabilities of the `netgraph` function in Sect. 8.3.4.

---

[5]We can also use argument `seq` in the `netmeta` function which would be considered in the `netgraph` and other functions, e.g. `print.netmeta`.
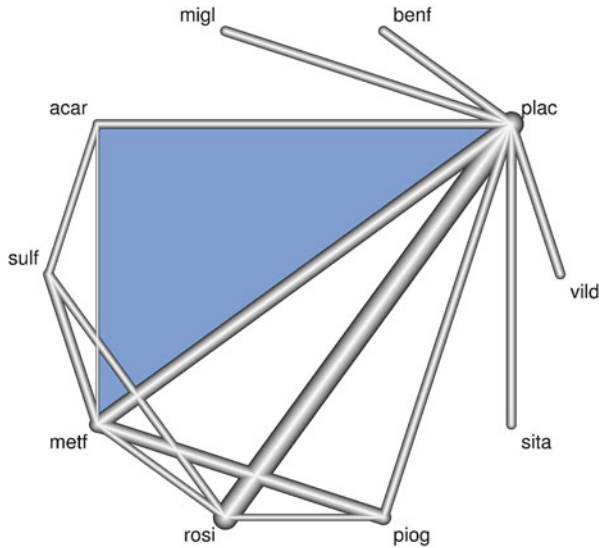
Fig. 8.3 A graph of the network for the diabetes data, generated using the `netgraph` function. The treatments are equally spaced on the perimeter of the circle. Any two treatments are connected by a line when there is at least one study comparing the two treatments. The thickness of the line is proportional to the inverse standard error of the direct treatment comparison. The *shading* indicates the three-arm study

8.3.3 A More Detailed Look at the Output

Next, we view the output stored in the R object `nm1` using the `print` function. We split this into several chunks which we discuss in turn.

```
> print(mn1, digits=2)
Original data (with adjusted standard errors for multi-arm studies):

            treat1 treat2     TE seTE seTE.adj narms multiarm
DeFronzo1995   metf   plac -1.90 0.14     0.14     2
Lewin2007      metf   plac -0.82 0.10     0.10     2
Willms1999     acar   metf  0.20 0.36     0.39     3        *
Davidson2007   plac   rosi  1.34 0.14     0.14     2
*** Output truncated ***
Moulin2006     benf   plac -1.01 0.14     0.14     2
Willms1999     metf   plac -1.20 0.38     0.41     3        *
Willms1999     acar   plac -1.00 0.47     0.82     3        *
...
```

First comes the data (treatments within a study have been re-ordered alphabetically by the program). Columns 1–5 correspond to the study labels, treatment groups (`treat1`, `treat2`), treatment effects (`TE`) and standard errors (`seTE`). The remaining three columns are only printed if at least one multi-arm study is

included in the meta-analysis. As discussed in Sect. 8.2.1, for multi-arm studies—
`Willms1999` in the example—standard errors have to be adjusted accordingly.
This is given in the column headed `seTE.adj`. Note that the standard error and
adjusted standard error are the same for two-arm studies. The other two columns
give the number of treatment arms per study (`narms`) and highlight multi-arm
studies using a star (`multiarm`).

   As we have already noted, it is important to use exactly identical study
labels for all treatment comparisons belonging to the same multi-arm study (here
`Willms1999`), otherwise the program will treat them as separate two-arm studies.
The reason for this special caution with multi-arm studies is, as mentioned before,
that the `netmeta` function automatically accounts for within-study correlation by
reweighing all pairwise comparisons of each multi-arm study.

   The next chunk of output gives a list of all the studies in alphabetical order with
information on the numbers of treatment arms per study (column `narms`). Once
again we see that `Willms1999` is the only study with three treatment arms.

```
...
Number of treatment arms (by study):
                    narms
Alex1998                2
*** Output truncated ***
Vongthavaravat2002      2
Willms1999              3
Wolffenbuttel1999       2
Yang2003                2
Zhu2003                 2
...
```

   The next chunk of output gives results from the fixed effect network meta-
analysis. For each study, and for each treatment comparison within that study, we
have the treatment effect (here the mean difference) fitted by the network meta-
analysis model, $\hat{\boldsymbol{\theta}}^{nma}$, see (8.6), and its 95 % confidence interval. Results are given
to two decimal places; this is controlled by the `digits` argument in the `print`
function. The corresponding contributions to the overall heterogeneity statistic,
Q_{total}, see (8.7), are then shown, followed by the leverage (which is given by the
corresponding diagonal element of the $m \times m$ hat matrix, **H**).

```
...
Data utilised in network meta-analysis (fixed effect model):

               treat1 treat2   MD          95%-CI     Q leverage
DeFronzo1995     metf   plac -1.11 [-1.23; -1.00] 30.89     0.18
Lewin2007        metf   plac -1.11 [-1.23; -1.00]  8.79     0.36
Willms1999       acar   metf  0.29 [ 0.06;  0.51]  0.05     0.09
Davidson2007     plac   rosi  1.20 [ 1.11;  1.30]  0.93     0.11
*** Output truncated ***
Hermansen2007    plac   sita  0.57 [ 0.32;  0.82]  0.00     1.00
Garber2008       plac   vild  0.70 [ 0.45;  0.95]  0.00     1.00
*** Output truncated ***
Gonzalez...2004  metf   plac -1.11 [-1.23; -1.00]  2.69     0.02
```

```
Stucci1996         benf   plac -0.91 [-1.15; -0.66]  3.79      0.13
Moulin2006         benf   plac -0.91 [-1.15; -0.66]  0.59      0.87
Willms1999         metf   plac -1.11 [-1.23; -1.00]  0.04      0.02
Willms1999         acar   plac -0.83 [-1.04; -0.61]  0.04      0.02

Number of studies: k=26
Number of treatments: n=10
Number of pairwise comparisons: m=28
...
```

Each of the heterogeneity statistics approximately follows a χ_1^2-distribution. They are useful to identify studies whose data differ markedly from what the model predicts. When there is only a single study evaluating a treatment comparison—e.g. `Hermansen2007` and `Garber2008` are the only studies evaluating sitagliptin and vildagliptin, respectively—the model fits perfectly and the corresponding Q statistic is zero. Conversely, we see that the results of `DeFronzo1995` differ markedly from what the rest of the network predicts. We can confirm this by comparing the estimated metformin vs placebo comparison from the network in the output immediately above (-1.11) with the original data from this study in the printout on page 200 (-1.90). The leverage of comparison i is the ith diagonal element of the hat matrix **H**. The model shows that this is the factor by which the variance of the estimate of a treatment comparison from a study is reduced by the information of the whole network. A small value of the leverage, close to 0, means a large variance reduction and thus a large gain in precision from the network. Conversely, a large value (close to 1) means almost no variance reduction and no gain in precision.

The mean leverage [27] depends only on the number of treatments n and the number of comparisons m: $(n-1)/m$. This can be checked by the following R commands

```
> mn1$n
[1] 10
> mn1$m
[1] 28
> mean(mn1$leverage.fixed)
[1] 0.3214286
> (mn1$n-1)/mn1$m
[1] 0.3214286
```

which are identical, and equal to $9/28 = 0.3214286$.

In the diabetes example the strongest gain from the network is seen in the `Gonzales-Ortiz2004` and two of the `Willms1999` comparisons with a leverage of 0.02. These comparisons have rather large standard errors and gain a lot from information through the network, particularly from other studies that have evaluated the same comparison, metformin or acarbose vs placebo. The largest possible leverage is 1.00, and two studies have this value: `Hermansen2007` and `Garber2008`. This occurs because they are the only studies evaluating sitagliptin and vildagliptin, respectively. Therefore these two drugs are not part of any loop

(circuit) in the network, see Fig. 8.3, and estimation cannot benefit from the additional data available within the network.

The next chunk of information gives the estimates for all the treatment contrasts in the network meta-analysis.

```
...
Fixed effect model

Treatment estimate (sm='MD'):
      acar  benf  metf  migl  piog  plac rosi  sita  sulf  vild
acar  0.00  0.08  0.29  0.12  0.24 -0.83 0.37 -0.26 -0.39 -0.13
*** Output truncated ***
vild  0.13  0.21  0.41  0.24  0.37 -0.70 0.50 -0.13 -0.26  0.00

Lower 95%-confidence limit:
      acar  benf  metf  migl  piog  plac  rosi  sita  sulf  vild
acar  0.00 -0.25  0.06 -0.21 -0.01 -1.04  0.15 -0.59 -0.61 -0.46
*** Output truncated ***
vild -0.20 -0.15  0.14 -0.11  0.08 -0.95  0.24 -0.49 -0.57  0.00

Upper 95%-confidence limit:
      acar  benf metf migl piog  plac rosi  sita  sulf  vild
acar  0.00  0.41 0.51 0.44 0.49 -0.61 0.60  0.07 -0.17  0.20
*** Output truncated ***
vild  0.46  0.56 0.69 0.60 0.66 -0.45 0.77  0.23  0.05  0.00
...
```

Above we have three $n \times n$ matrices (recall n is the number of treatments). These show the estimated treatment comparisons as well as lower and upper 95 % confidence limits. Results for each possible treatment comparisons are given.

Estimated treatment effects and confidence limits use information both from direct and indirect treatment comparisons. As the network is connected, all comparisons can be estimated.

The last chunk of output gives measures of heterogeneity/network inconsistency.

```
...
Quantifying heterogeneity/inconsistency:
tau^2 = 0.1087; I^2 = 81.4%

Test of heterogeneity/inconsistency:
    Q d.f.  p.value
 96.99   18 < 0.0001
```

Specifically we have the generalised DerSimonian–Laird estimator τ^2 [8, 15], Higgins' I^2, Cochran's Q_{total} (Q) with its degrees of freedom (d.f.), and a p-value for Q_{total} [12]. The results show that there is considerable heterogeneity in the network, and this needs to be explored further.

Finally, as an alternative to print.netmeta function, we can obtain a shorter summary of the analysis using the summary.netmeta function:

```
> print(summary(mn1))
Number of studies: k=26
```

```
Number of treatments: n=10
Number of pairwise comparisons: m=28
*** Output truncated ***
```

The resulting output includes matrices for treatment effects and confidence limits, as presented above.

8.3.4   Additional Network Plots

In Fig. 8.3 we have already seen a network plot produced by function `netgraph`. Here, we explore ways how to create alternative network plots using some of the additional functionality of `netgraph` [29].

We used the default settings to obtain Fig. 8.3. This placed the nodes (treatments) on a circle as, by default, argument `start.layout` is equal to `"circle"`.

We can use argument `iterate=TRUE` to further optimise the layout. First, "ideal" distances between each pair of nodes in the plane are specified. We followed a proposal in the literature to take the graph distance of nodes i and j, defined as the length, i.e. the number of edges, of the shortest path connecting i and j [14]. However, for most graphs this cannot be perfectly realised. Therefore, starting from an initial layout, the optimum is approximated in an iterative process called stress majorisation [14, 17, 23, 29], which is essentially a form of least squares optimisation. Users can choose whether network graphs generated for each iteration step are shown using argument `allfigures=TRUE`. A different starting layout than a circle can be chosen by argument `start.layout` (or abbreviated `start`): a starting layout obtained via eigenvectors of the Laplacian matrix (following Hall's algorithm [11]) or a random starting layout (`start="random"`). For Hall's algorithm, there are two possible options for setting the argument `start.layout`, `"eigen"`, or `"prcomp"`. These correspond to different procedures for computing eigenvectors by R function `eigen` or `prcomp` (via principal component analysis).

To see a series of random layouts, repeatedly execute the command (graph not shown):

```
> netgraph(mn1, start="random", iterate=TRUE,
+          col="darkgray", cex=1.5, multiarm=FALSE,
+          points=TRUE, col.points="green", cex.points=3)
```

If argument `points` is set to `TRUE`, nodes are marked. The marker type, size, and colour of the points can be specified using the arguments `pch.points`, `cex.points`, `col.points` which may be vectors. As usual, further generic plotting arguments are available, such as `cex` (determining the size of the treatment labels), `lwd` (determining the thickness of lines), `col` (determining the colour of lines), and others; see also help page of `netgraph` function.

In addition, the `netgraph` function allows us to identify multi-arm studies by coloured polygons. If argument `multiarm` is set to `FALSE` as in the command we have used, multi-arm studies are not identified in the figure.

Fig. 8.4 Network graph for the diabetes data, drawn by the stress majorisation algorithm. The *shading* indicates the three-arm study
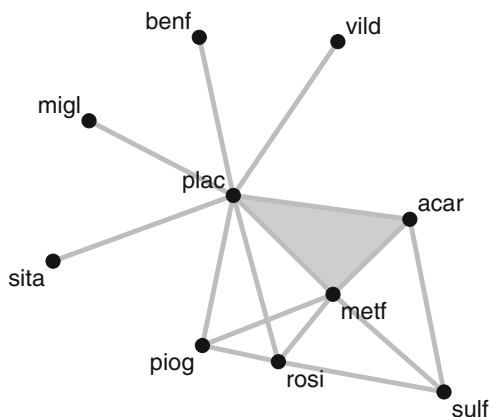
Using multiarm=TRUE (default) without explicitly specifying argument col.multiarm means that multi-arm studies are coloured using transparent colours from the R library **colorspace** if this is available, or colours from the rainbow scale otherwise. Unfortunately, transparent colours are not possible to display with all graphics devices. However, they can usually be incorporated into PDF-files. The function issues a warning if colours are not appropriately displayed.

As we have only one multi-arm study in our dataset, transparent colouring is not needed. Figure 8.4 is generated with the command

```
> netgraph(mn1, start="circle", iterate=TRUE,
+          col="darkgray", cex=1.5,
+          points=TRUE, col.points="black", cex.points=3,
+          col.multiarm="gray")
```

If you wish to view all the details of the iterations, then include the argument allfigures=TRUE (graphs not shown):

```
> netgraph(mn1, start="circle", iterate=TRUE,
+          col="darkgray", cex=1.5,
+          points=TRUE, col.points="black", cex.points=3,
+          col.multiarm="gray", allfigures=TRUE)
```

8.3.5   Forest Plots

Sometimes in a network meta-analysis the primary interest is to compare a number of treatments to a common treatment (also called reference or baseline treatment). This is usually placebo, usual care, no treatment, or a well-established standard treatment. When summarising the output, a reference treatment can be specified using the argument reference.group (abbreviated ref) in the netmeta, summary.netmeta, and the corresponding print functions.

```
> summary(mn1, ref="plac")
Number of studies: k=26
Number of treatments: n=10
Number of pairwise comparisons: m=28

Fixed effect model

Treatment estimate (sm="MD", reference.group="plac"):
           MD              95%-CI
acar -0.8274  [-1.0401; -0.6147]
benf -0.9052  [-1.1543; -0.6561]
metf -1.1141  [-1.2309; -0.9973]
migl -0.9439  [-1.1927; -0.6952]
piog -1.0664  [-1.2151; -0.9178]
plac  0.0000  [ 0.0000;  0.0000]
rosi -1.2018  [-1.2953; -1.1084]
sita -0.5700  [-0.8230; -0.3170]
sulf -0.4395  [-0.6188; -0.2602]
vild -0.7000  [-0.9495; -0.4505]
*** Output truncated ***
```

We see that in this printout all treatments are compared to the reference treatment, which is placebo.

A corresponding forest plot using reference group `"plac"` can be generated using the following R command:

```
> forest(mn1, ref="plac")
```

This produces a forest plot (graph not shown) for the fixed effect model because the analysis which created the R object `mn1` used the fixed effect model.

For the `forest.netmeta` function the argument `reference.group` is mandatory if this argument has not been used in the generation of the `netmeta` object. For example, the following R command will result in an error message.

```
> forest(mn1)
Error in forest.netmeta(mn1) :
  Argument 'reference.group' must match any of the following
  values: 'acar' - 'benf' - 'metf' - 'migl' - 'piog' - 'plac' -
          'rosi' - 'sita' - 'sulf' - 'vild'
```

Additional arguments are available in the `forest.netmeta` function to modify the figure. As always, the arguments of a function can be displayed using the `args` function.

```
> args(forest.netmeta)
function (x, pooled = ifelse(x$comb.random, "random", "fixed"),
    reference.group = x$reference.group, leftcols = "studlab",
    leftlabs = "Treatment", smlab = NULL, sortvar = x$seq, ...)
```

The argument `pooled` is used to explicitly specify whether the forest plot should be based on a fixed effect or random effects model. If `comb.random=FALSE` in the generation of the `netmeta` object, the fixed effect model is used. Otherwise, if `comb.random=TRUE`, the random effects model is used.
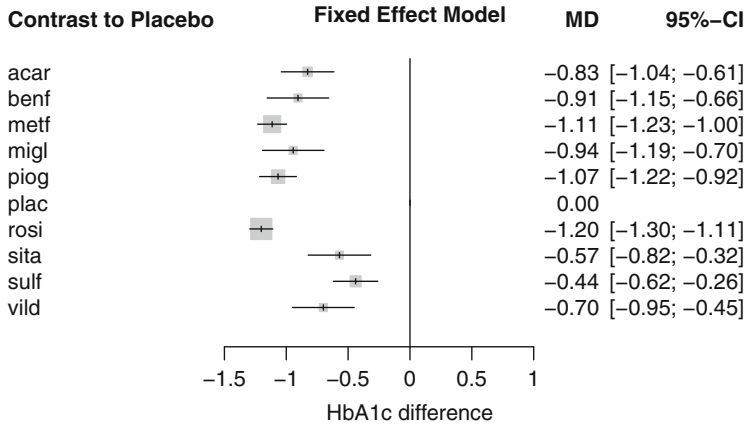
Fig. 8.5 Forest plot for the Senn data example, fixed effect model, with placebo as reference
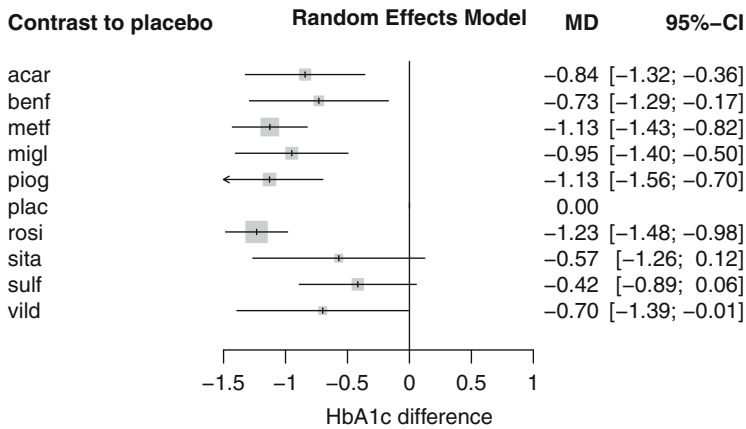
Fig. 8.6 Forest plot for the Senn data example, random effects model, with placebo as reference

Forest plots for fixed effect and random effects model are shown in Figs. 8.5 and 8.6, respectively. The following commands were used to produce these plots.

```
> forest(mn1, xlim=c(-1.5, 1), ref="plac",
+        leftlabs="Contrast to Placebo",
+        xlab="HbA1c difference")
> forest(mn1, xlim=c(-1.5, 1), ref="plac",
+        leftlabs="Contrast to placebo",
+        xlab="HbA1c difference",
+        pooled="random")
```

□

8.4   Decomposition of the Heterogeneity Statistic

The Q_{total} statistic (of the "whole network") can be decomposed into a Q statistic for assessing the heterogeneity between studies with the same design ("within designs") and a Q statistic for assessing the design inconsistency ("between designs"). Designs are defined by the subset of treatments compared in a study. For example, the comparison acarbose vs placebo in `Costa1997` and the comparison acarbose vs metformin vs placebo in `Willms1999` are two different designs in the Senn data example (see Fig. 8.2) even though the pairwise treatment comparison acarbose vs placebo is included in both designs.

Example 8.5 We can use the `decomp.design` function to calculate the Q statistics.

```
> round(decomp.design(mn1)$Q.decomp, 3)
                     Q df  pval
Whole network   96.986 18 0.000
Within designs  74.455 11 0.000
Between designs 22.530  7 0.002
```

We have used the fixed effect model for this analysis and we see there exists considerable heterogeneity/inconsistency within as well as between designs. We can further decompose the total within-design heterogeneity into the contribution from each design:

```
> print(decomp.design(mn1)$Q.het.design, digits=2)
           design     Q df    pval
1        acar:plac  0.00  0      NA
2        acar:sulf  0.00  0      NA
3        benf:plac  4.38  1 3.6e-02
4        metf:piog  0.00  0      NA
5        metf:plac 42.16  2 7.0e-10
6        metf:rosi  0.19  1 6.7e-01
7        metf:sulf  0.00  0      NA
8        migl:plac  6.45  2 4.0e-02
9        piog:plac  0.00  0      NA
10       piog:rosi  0.00  0      NA
11       plac:rosi 21.27  5 7.2e-04
12       plac:sita  0.00  0      NA
13       plac:vild  0.00  0      NA
14       rosi:sulf  0.00  0      NA
15 acar:metf:plac  0.00  0      NA
```

As we can see, 15 different designs are used in the 26 studies included in the network meta-analysis. Since there are only five designs for which we have more than one study, the remaining design specific Q statistics are equal to zero and have no degrees of freedom. Except for design `metf:rosi` ($p = 0.67$), for all the other four designs there is more heterogeneity between the contributing studies than we would expect by chance; in the case of `metf:plac` a substantial amount more ($p < 0.0001$). In a substantive application we would try and identify the sources of this and update the analysis appropriately.

The `decomp.design` function also gives the between-designs Q statistic based on a random effects model. This can be calculated based on a full design-by-treatment interaction random effects model [13]. Here, τ^2 is estimated by the method of moments [15]. Alternatively, the square-root of the between-study variance can be prespecified with `tau.preset`. This Q statistic can be displayed by entering:

```
> round(decomp.design(mn1)$Q.inc.random, 3)
                     Q df  pval tau.within
Between designs 2.194  7 0.948       0.38
```

□

8.5 The Net Heat Plot

We now introduce the net heat plot, proposed by Krahn, König, and Binder [18–20]. This is a graphical presentation which displays in a single plot two types of information. These are:

(1) for each network estimate, the contribution of each design to this estimate, and
(2) for each network estimate, the extent of inconsistency due to each design.

Taking (1) first, we have already seen in (8.6) that the elements in a row of the matrix **H** describe the contribution of the treatment comparison in each column to the network estimate in the row. The hat matrix **H** of dimension $m \times m$ was defined based on all individual pairwise treatment comparisons in the network, i.e. $m = 28$ treatment comparisons in the Senn data example. By contrast, the net heat plot is based on a condensed hat matrix with rows and columns corresponding to treatment comparisons within designs instead of single comparisons. This hat matrix therefore has lower dimension.

Example 8.6 Figure 8.7 shows a net heat plot for the fixed effect analysis of the Senn data example which was created using the following `netheat` command.

```
> netheat(mn1)
```

The rows and columns correspond to treatment comparisons within designs; treatment comparisons for which there is only one source of evidence are omitted. Pairwise comparisons corresponding to the three-arm design are designated by "_" following the treatment comparison label. The grey squares have area proportional to the contribution from the treatment comparison in the column to the treatment comparison in the row. For example, for the `acar:plac` treatment comparison, the largest grey square is on the diagonal indicating that the direct comparison is the greatest source of information. However, there are also moderate size squares in the same row for the `metf:sulf`, `rosi:sulf`, `plac:rosi`, `metf:plac`, and `acar:sulf` comparisons, indicating these are important sources of additional indirect evidence.
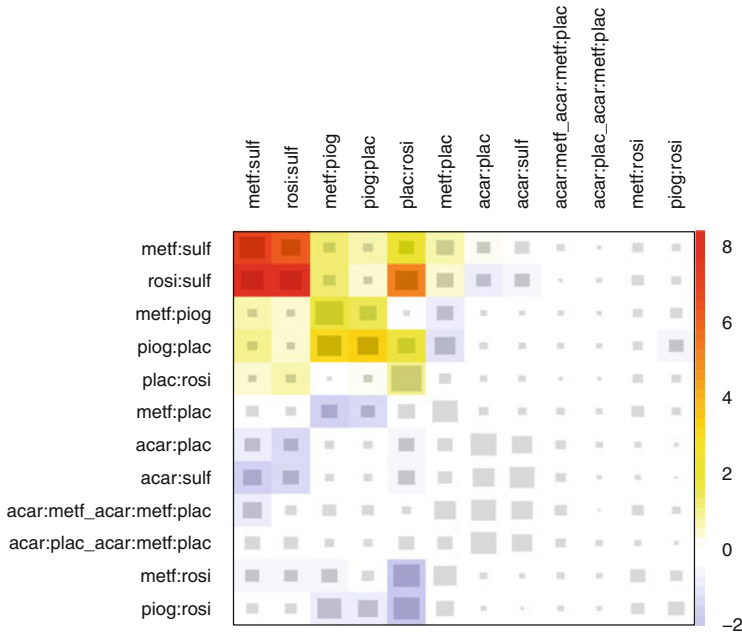
Fig. 8.7 Net heat plot of the Senn data example based on a fixed effect model

Now we consider (2). Recall that above we have decomposed the heterogeneity statistic Q_{total} into a within and between design component. We can further decompose the between design component into the contribution for each design. These are displayed in colour on the top-left to bottom-right diagonal of Fig. 8.7, with the largest heterogeneity shown in red in the top left corner. These correspond to the metf:sulf and rosi:sulf designs, which are together responsible for the majority of between-design heterogeneity.

We now turn to the off-diagonal colours. For each treatment comparison row, these are determined by the change in design inconsistency when a particular design is detached, i.e. after removing the consistency assumption for that specific design. For example, consider the top row of Fig. 8.7, corresponding to the metf:sulf treatment comparison. The (1,2) position is coloured red, which corresponds to a score of 6–8 in the colour scale on the right-hand side of the plot. This means that if we "remove" the assumption of consistency for the design in column 2 (rosi:sulf) and re-estimate the between-design inconsistency contribution to the design metf:sulf, it decreases. This means that the evidence for the treatment comparison metf:sulf from the design rosi:sulf is inconsistent with the other evidence.

Conversely, blue indicates that the evidence for the treatment comparison in the row from the design in the column is consistent. For example, for the

`acar:sulf` treatment comparison, the indirect evidence from the `metf:sulf` and `rosi:sulf` designs supports the direct evidence (coloured blue in the plot).

If the colours of a column corresponding to a design are identical to the colours on the diagonal, the detaching of the effect of this design removes the total inconsistency in the network.

As we have already noted, the diagonal colours show the designs `metf:sulf`, `rosi:sulf`, `metf:piog`, `piog:plac`, and `plac:rosi` contribute the most to the between-design inconsistency. The contributions of each design can also be printed as follows:

```
> round(decomp.design(mn1)$Q.inc.design, 2)
acar:plac       acar:sulf       benf:plac       metf:piog
     0.04            0.01            0.00            1.75
metf:plac       metf:rosi       metf:sulf       migl:plac
     0.20            0.01            6.62            0.00
piog:plac       piog:rosi       plac:rosi       plac:sita
     3.39            0.04            1.05            0.00
plac:vild       rosi:sulf acar:metf:plac acar:metf:plac
     0.00            9.29            0.01            0.13
```

We reiterate the point made above, that designs where only one treatment is involved in other designs of the network or where the removal of corresponding studies would lead to a splitting of the network do not contribute to the inconsistency assessment and are not incorporated into the net heat plot in Fig. 8.7. These are the four designs `benf:plac`, `migl:plac`, `plac:sita`, and `plac:vild`.

We see that the `metf:sulf` design contributes 6.62, and the `rosi:sulf` design contributes 9.29, consistent with the red colouring of the first and second diagonals of Fig. 8.7.

We have already commented on the red entries in positions (1,1), (1,2), (2,1) and (2,2) of Fig. 8.7. However, entries (4,3) and (4,4) are coloured orange, indicating inconsistency of evidence from the designs `metf:piog` and `piog:plac`.

Now consider the design `plac:rosi` (fifth row). There are six studies of this design in the network, and the within-design heterogeneity for this statistic, shown in Sect. 8.4 is large (21.27 on 5 degrees of freedom, $p < 0.001$). However, its large evidence base means it provides a lot of direct information (a large grey square in position (5,5)) as well as providing a lot of indirect information (as shown by the number of large grey squares in the `plac:rosi` column).

The colours in the `metf:plac` column are very light yellow in rows 1 and 2, then blue in rows 3 and 4, and then white. This means that relaxing the consistency assumption for this design slightly decreases the inconsistency contribution to the `metf:sulf` and `rosi:sulf` comparisons. However, the evidence from the `metf:plac` design is consistent with that from the `metf:piog` and `piog:plac` designs.

Overall, the heterogeneity in the network cannot be traced to one design. However, the single largest reduction in the whole network inconsistency is achieved by removing the `rosi:sulf` design (see the design inconsistency Q statistics above).
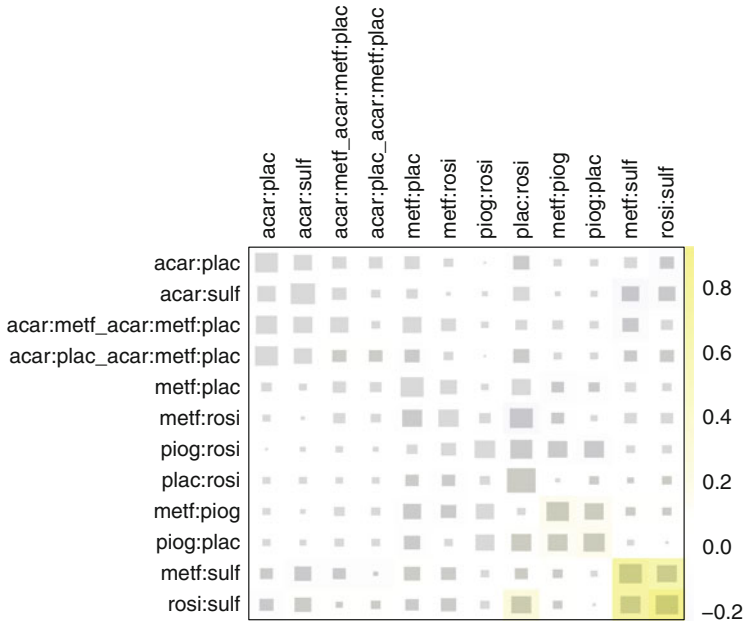
Fig. 8.8 Net heat plot of the Senn data example from a random effects model

Given the extent of heterogeneity, we may conclude a random effects analysis is more appropriate. The corresponding net heat plot is shown in Fig. 8.8, and shows a marked reduction in inconsistency. It is obtained by the following command:

```
> netheat(mn1, random=TRUE)
```

□

8.5.1 Bland–Altman Plot to Assess the Effect of Heterogeneity on Estimated Treatment Comparisons

In the Senn data example there is a substantial amount of heterogeneity. This is reflected in the fact that the estimated common heterogeneity variance $\hat{\tau}^2$ is larger than most of the study-specific sampling variances. Accordingly, alongside the fixed effect model, we might consider a random effects model.

Example 8.7 The results for the fixed effect and random effects model can be compared in a Bland–Altman plot [6] as follows:

```
> # Set seed so results are reproducible
> set.seed(125)
> fe <- mn1$TE.nma.fixed
> re <- mn1$TE.nma.random
```

```
> plot(jitter((fe+re)/2, 5), jitter(fe-re, 5),
+       xlim=c(-1.2, 1.2),
+       ylim=c(-0.25, 0.25),
+       xlab="Mean treatment effect (in fixed effect and random
                effects model)",
+       ylab="Difference of treatment effect (fixed effect minus
                random effects model)")
> abline(h=0)
```

The `jitter` function is used to separate overlapping points by adding a small random error to each x- and y-value. As the plot contains a random element, we use the `set.seed` command to generate a random, but reproducible, graph.

Figure 8.9 shows that estimated treatment effects of fixed effect and random effects model are similar with two exceptions. Estimated treatment effects in the random effects model are somewhat larger in the studies by `Stucci1996` and `Moulin2006` (bottom left of the plot).

A comparison of standard errors for fixed effect and random effects model shows that standard errors in the random effects model are much larger.

```
> summary(mn1$seTE.nma.random / mn1$seTE.nma.fixed)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.826   2.265   2.588   2.502   2.681   3.231
```

On average, standard errors in the random effects model (and accordingly confidence intervals) are about 2.5 times as large as those in the fixed effect model. However, treatment effect estimates are broadly similar. The results can be further illustrated using forest plots if desired (not shown). □
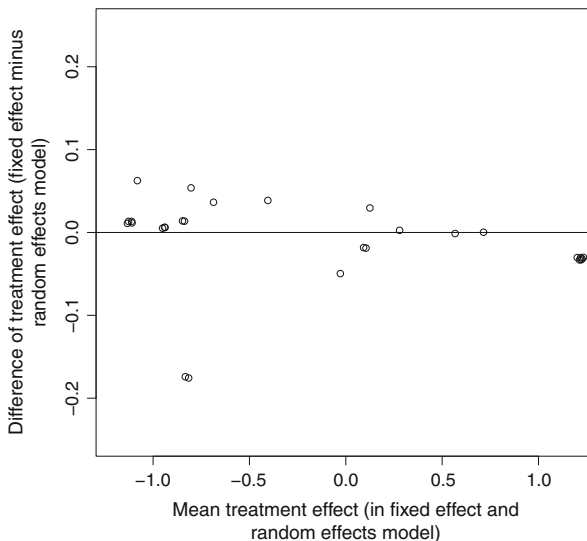
Fig. 8.9 Bland–Altman plot comparing individual treatment effects for fixed effect and random effects model

8.6   Summary

Network meta-analysis is a potentially powerful tool for using all the evidence in a particular area to estimate and compare treatment effects.

In this chapter, we have described a weighted least squares estimation approach which is implemented in the R package **netmeta**. The software can handle both single-arm and multi-arm studies; for the latter it accounts for the correlation appropriately. We have illustrated this approach using an example from diabetes [33], showing how the network can be graphed and a range of analyses explored.

Two important aspects of network meta-analysis are the extent of information gained on a particular treatment comparison through the indirect evidence and the extent of heterogeneity. Information about both of these is conveyed in the *net heat* plot; in addition the software provides a decomposition of heterogeneity within designs and between designs.

If there is clinically relevant heterogeneity, it should be explored further. As covariate adjustment is not currently possible with the software, one approach is to perform study specific (ideally individual participant data) analyses with appropriate covariate adjustment before using the software presented here to perform the network meta-analysis.

References

1. Achana, F., Hubbard, S., Sutton, A., Kendrick, D., Cooper, N.: An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial.   J. Clin. Epidemiol. (2013). doi:10.1016/j.jclinepi.2013.09.018
2. Albert, A.E.: Regression and the Moore-Penrose Pseudoinverse. Mathematics in Science and Engineering. Academic, New York (1972). ISBN:0-12-048450-1
3. Bafeta, A., Trinquart, L., Seror, R., Ravaud, P.: Analysis of the systematic reviews process in reports of network meta-analysis: methodological systematic review. Br. Med. J. **347**, f3675 (2013)
4. Bailey, R.A.: Designs for two-colour microarray experiments. J. R. Stat. Soc. Ser. C Appl. Stat. **56**(4), 365–394 (2007)
5. Bailey, R.A., Cameron, P.J.: Combinatorics of optimal designs. In: Huczynska, S., Mitchell, J.D., Roney-Dougal, C.M. (eds.) Surveys in Combinatorics.   Mathematical Society Lecture Notes, vol. 365, pp. 19–73. Cambridge University Press, London (2009)
6. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement.   Lancet **327**(8476), 307–310 (1986). doi:10.1016/S0140-6736(86)90837-8
7. Bollobás, B.: Modern Graph Theory. Springer, Heidelberg/New York (2002)
8. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Control. Clin. Trials **7**, 177–188 (1986)
9. Doyle, P.G., Snell, J.L.: Random Walks and Electric Networks. The Carus Mathematical Monographs. Mathematical Association of America, Washington, DC (1999)
10. Gutman, I., Xiao, W.: Generalized inverse of the Laplacian matrix and some applications. Bulletin T.CXXIX de l'Académie Serbe des Sciences et des Arts **29**, 15–23 (2004)

11. Hall, K.M.: An r-dimensional quadratic placement algorithm. Manage. Sci. **17**, 219–229 (1970)

12. Higgins, J.P.T., Thompson, S.G.: Quantifying heterogeneity in a meta-analysis. Stat. Med. **21**, 1539–1558 (2002)

13. Higgins, J.P.T., Jackson, D., Barrett, J.K., Lu, G., Ades, A.E., White, I.R.: Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res. Synth. Methods **3**(2), 98–110 (2012). doi:10.1002/jrsm.1044

14. Hu, Y.: Algorithms for visualizing large networks. In: U. Naumann, O. Schenk (eds.) Combinatorial Scientific Computing, pp. 525–549. Chapman and Hall/CRC Computational Science, Boca Raton, London, New York (2012). ISBN:9781439827352

15. Jackson, D., White, I.R., Riley, R.D.: Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat. Med. **31**(29), 3805–3820 (2012)

16. Jackson, D., White, I.R., Riley, R.D.: A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. Biom. J. **55**(2), 231–245 (2013). doi:10.1002/bimj.201200152

17. Kamada, T., Kawai, S.: An algorithm for drawing general undirected graphs. Inf. Process. Lett. **31**(1), 7–15 (1989)

18. König, J., Krahn, U., Binder, H.: Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat. Med. **32**, 5414–5429 (2013). doi:10.1002/sim.6001

19. Krahn, U., Binder, H., König, J.: A graphical tool for locating inconsistency in network meta-analyses. BMC Med. Res. Methodol. **13**(1), 35 (2013)

20. Krahn, U., Binder, H., König, J.: Visualizing inconsistency in network meta-analysis by independent path decomposition. BMC Med. Res. Methodol. **14**(1), 131 (2014)

21. Lee, A.W.: Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J. Clin. Epidemiol. **67**(2), 138–43 (2014). doi:10.1016/j.jclinepi.2013.07.014

22. Lu, G., Welton, N.J., Higgins, J.P.T., White, I.R., Ades, A.E.: Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. Res. Synth. Methods **2**, 43–60 (2011). doi:10.1002/jrsm.34

23. Michailidis, G., de Leeuw, J.: Data visualization through graph drawing. Comput. Stat. **16**(3), 435–450 (2001)

24. Nietert, P.J., Wahlquist, A.E., Herbert, T.L.: Characteristics of recent biostatistical methods adopted by researchers publishing in general/internal medicine journals. Stat. Med. **32**(1–10) (2013). doi:10.1002/sim.5311

25. Paterson, L.: Circuits and efficiency in incomplete block designs. Biometrika **70**(1), 215–225 (1983)

26. Rao, C., Mitra, S.K.: Generalized Inverse of Matrices and its Applications. Wiley, New York, London, Sydney, Toronto (1971). ISBN:0-471-70821-6

27. Rücker, G.: Network meta-analysis, electrical networks and graph theory. Res. Synth. Methods **3**, 312–324 (2012)

28. Rücker, G., Schwarzer, G.: Reduce dimension or reduce weights? Comparing two approaches to multi-armed studies in network meta-analysis. Stat. Med. **33**(25), 4353–4369 (2014). doi:10.1002/sim.6236

29. Rücker, G., Schwarzer, G.: Automated drawing of network plots in network meta-analysis. Res. Syn. Meth. (2015). doi:10.1002/jrsm.1143

30. Rücker, G., Schwarzer, G., Krahn, U., König, J.: netmeta: network meta-analysis with R (2014). www.cran.R-project.org/package=netmeta. R package version 0.6-0

31. Salanti, G.: Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res. Synth. Methods, 80–97 (2012). doi:10.1002/jrsm.1037

32. Salanti, G., Higgins, J.P., Ades, A.E., Ioannidis, J.P.: Evaluation of networks of randomized trials. Stat. Methods Med. Res. **17**(3), 279–301 (2008)

33. Senn, S., Gavini, F., Magrez, D., Scheen, A.: Issues in performing a network meta-analysis. Stat. Methods Med. Res. **22**, 169–89 (2013)
34. Spielman, D.: Spectral graph theory. In: U. Naumann, O. Schenk (eds.) Combinatorial Scientific Computing. Chapman and Hall/CRC Computational Science, Boca Raton (2012). ISBN:9781439827352
35. Veroniki, A.A., Vasiliadis, H.S., Higgins, J.P., Salanti, G.: Evaluation of inconsistency in networks of interventions. Int. J. Epidemiol. **42**(1), 332–345 (2013). doi:10.1093/ije/dys222
36. Yates, F.: The recovery of inter-block information in balanced incomplete block designs. Ann. Eugen. **10**(4), 317–325 (1940)

Chapter 9
Meta-Analysis of Diagnostic Test Accuracy Studies

Meta-analysis of diagnostic test accuracy (DTA) studies differs from meta-analysis of intervention studies in a number of respects. In this chapter, we explain the issues raised by meta-analysis of diagnostic accuracy studies and how these may be addressed. Alongside the statistical models, we present the R package **mada** [5] written for fitting these models and graphing the results.[1] Full details about the R package **mada** can be obtained by entering the R command `vignette("mada")`.

9.1 Special Challenges in Meta-Analysis of Diagnostic Test Accuracy Studies

The methodology of systematic reviews and meta-analyses of DTA studies is relatively new, with the first papers appearing in the 1990s, and recent important contributions [20]. Methods for meta-analysis of intervention studies cannot be readily translated to DTA studies. The reason is that in DTA studies the outcome is bivariate, typically consisting of the (sensitivity, specificity) pair for a number of studies. As we will outline, the large across-study correlation between these parameters is of central interest. Therefore, it is not appropriate to conduct separate meta-analyses for sensitivity and specificity [7]. Instead, bivariate modelling is necessary. Thus meta-analysis of DTA studies is a special application of multivariate meta-analysis, introduced in Chap. 7; however, we use a separate package to make the calculations easier. A further challenge is that heterogeneity is typically much larger in meta-analysis of DTA studies because of (1) variation between studies

---

[1]To install the R package **mada** use the R command `install.packages("mada")`. This will automatically install the package R package **mvmeta** which it depends on.

in how a continuous marker is dichotomised into a test classification and also (2) variation in the accuracy of tests across different settings.

9.2   Analysis of Diagnostic Test Accuracy Studies

In a typical DTA study, there are two groups of patients, those who truly have a certain condition or disease, and those who in truth are disease free. In this chapter, we assume that the presence or absence of the disease has been ascertained by a fully accurate "gold standard" procedure, so that the true condition of each patient is assumed known.

Now we are interested in a so-called *index test* which we hope will be a reliable guide as to whether the patient has the disease or not. The test may be based on a biomarker, a questionnaire, an imaging modality, or a more complex diagnostic procedure. However, the index test is not completely reliable. Thus, for a proportion of patients who have a positive index test result (denoted T_+) they will in truth *not have* the disease. Likewise, a proportion of patients who have a negative index test result (denoted T_-) will in truth *have* the disease. A generic representation of the resulting data is given in Table 9.1.

9.2.1   Definition of Sensitivity and Specificity

Two common measures to evaluate the performance of a diagnostic test are *sensitivity* and *specificity*.

The sensitivity is defined as the probability that a patient has a positive index test result *given he/she in truth has the disease*. It is estimated by

$$\widehat{Se} = \frac{TP}{n_1}.$$

Table 9.1 Generic representation of data from a diagnostic test accuracy study

	Disease status		
Test result	D_+	D_-	Total
T_+	TP	FP	TP+FP
T_-	FN	TN	FN+TN
Total	n_1	n_2	n

Here, n_1 patients truly have the disease, and n_2 are disease free

T_+ denotes a positive test result, and T_- a negative test result

TP true positives, *FP* false positives, *TN* true negatives, *FN* false negatives

The estimated sensitivity is also called true positive rate (TPR). The corresponding false negative rate (FNR) is given by $1 - \widehat{Se} = FN/n_1$.

The specificity is defined as the probability that a patient has a negative index test result *given he/she in truth does not have the disease*. This is estimated by the true negative rate (TNR)

$$\widehat{Sp} = \frac{TN}{n_2}.$$

The corresponding false positive rate (FPR) is $1 - \widehat{Sp} = FP/n_2$.

Ideally, both sensitivity and specificity of a diagnostic test are close to one. However, as we will see in Sect. 9.2.3 there is a trade-off between these two measures.

Example 9.1 Scheidler et al. [18] conducted a DTA meta-analysis to compare the utility of lymphangiography, computed tomography and magnetic resonance imaging (MR) for the diagnosis of lymph node metastasis in patients with cervical cancer. Here we restrict our attention to the MR data from ten studies. R code to read in and print the MR dataset is given in Fig. 9.1.

As shown in Fig. 9.1, sensitivity and specificity can be easily calculated from the available data. In this dataset there is substantial heterogeneity with sensitivity values ranging from 0 to 0.89. On the other hand, much less heterogeneity is seen in specificity values which range from 0.84 to 1.

```
> # 1. Load MR dataset (Scheidler 1997, Table 3)
> data16 <- read.csv("dataset16.csv", as.is=TRUE)
> # 2. Print dataset
> data16
     author year tp n1  tn  n2
1    Hricak 1988  9 11  44  46
2     Greco 1989  3  8  32  38
3     Janus 1989  3  4  16  18
4       Kim 1990  3 15  44  45
5        Ho 1992  0  5  15  15
6       Kim 1993  7 29 167 169
7   Hawnaur 1994 12 16  29  33
8       Kim 1994 23 37 230 235
9     Subak 1995  8 13  53  58
10    Heuck 1997 16 18  22  24
> # 3. Calculate sensitivity
> round(data16$tp / data16$n1, 4)
[1] 0.8182 0.3750 0.7500 0.2000 0.0000 0.2414 0.7500 0.6216
[9] 0.6154 0.8889
> # 4. Calculate specificity
> round(data16$tn / data16$n2, 4)
[1] 0.9565 0.8421 0.8889 0.9778 1.0000 0.9882 0.8788 0.9787
[9] 0.9138 0.9167
```

Fig. 9.1 Magnetic resonance imaging data for the diagnosis of lymph node metastasis in patients with cervical cancer [18]. See Table 9.1 for an explanation of the column headings tp, n1, tn and n2

We can use the `metaprop` function from R package **meta** to calculate confidence intervals for sensitivity

```
> metaprop(tp, n1, data=data16,
+          comb.fixed=FALSE, comb.random=FALSE,
+          studlab=paste(author, year))
                proportion          95%-CI
Hricak 1988        0.8182 [0.4822; 0.9772]
Greco 1989         0.3750 [0.0852; 0.7551]
Janus 1989         0.7500 [0.1941; 0.9937]
Kim 1990           0.2000 [0.0433; 0.4809]
Ho 1992            0.0000 [0.0000; 0.5218]
Kim 1993           0.2414 [0.1030; 0.4354]
Hawnaur 1994       0.7500 [0.4762; 0.9273]
Kim 1994           0.6216 [0.4476; 0.7754]
Subak 1995         0.6154 [0.3158; 0.8614]
Heuck 1997         0.8889 [0.6529; 0.9862]
*** Output truncated ***
```

and specificity

```
> metaprop(tn, n2, data=data16,
+          comb.fixed=FALSE, comb.random=FALSE,
+          studlab=paste(author, year))
                proportion          95%-CI
Hricak 1988        0.9565 [0.8516; 0.9947]
Greco 1989         0.8421 [0.6875; 0.9398]
Janus 1989         0.8889 [0.6529; 0.9862]
Kim 1990           0.9778 [0.8823; 0.9994]
Ho 1992            1.0000 [0.7820; 1.0000]
Kim 1993           0.9882 [0.9579; 0.9986]
Hawnaur 1994       0.8788 [0.7180; 0.9660]
Kim 1994           0.9787 [0.9510; 0.9931]
Subak 1995         0.9138 [0.8102; 0.9714]
Heuck 1997         0.9167 [0.7300; 0.9897]
*** Output truncated ***
```

These commands print the same values for sensitivity and specificity as given in Fig. 9.1. Furthermore, exact binomial confidence intervals are printed for sensitivity and specificity, respectively.

Note, we do not conduct a meta-analysis for sensitivity and specificity (arguments `comb.fixed=FALSE` and `comb.random=FALSE`) separately, as the (sensitivity, specificity) pairs from all studies are correlated, and this plays an important role in the analysis [7].   □

9.2.2 Additional Measures: Diagnostic Odds Ratio and Likelihood Ratios

The results of a DTA study are typically reported as a (sensitivity, specificity) pair. However, some attempts have been made to condense the result of a DTA study in

a single number. The most common approach is the use of the diagnostic odds ratio (DOR) [12].

The DOR is defined as

$$\text{DOR} = \frac{\text{TP}/\text{FN}}{\text{FP}/\text{TN}} = \frac{\text{TP} \times \text{TN}}{\text{FP} \times \text{FN}} = \frac{\widehat{\text{Se}} \times \widehat{\text{Sp}}}{(1 - \widehat{\text{Se}})\,(1 - \widehat{\text{Sp}})}.$$

This is exactly the same information as used in Eq. (3.2) which defines the odds ratio for studies with binary outcome, see also Table 3.1. Accordingly, the `metabin` function can be used to calculate DORs, as shown in the next example.

The DOR can also be defined in terms of the so-called likelihood ratios. The positive and negative likelihood ratio are defined as

$$\text{LR}_+ = \frac{\widehat{\text{Se}}}{1 - \widehat{\text{Sp}}}$$

and

$$\text{LR}_- = \frac{1 - \widehat{\text{Se}}}{\widehat{\text{Sp}}}.$$

A good diagnostic test should have both high sensitivity and specificity. Accordingly, the positive likelihood ratio should be large (as $\widehat{\text{Se}}$ should be close to 1 while $1 - \widehat{\text{Sp}}$ should be close to zero) and the negative likelihood ratio should be small (as $1 - \widehat{\text{Se}}$ should be close to zero while $\widehat{\text{Sp}}$ should be close to 1). Values close to one for LR_+ or LR_- would mean that the diagnostic test has no predictive ability with respect to positive (LR_+) or negative (LR_-) test results.

An alternative definition of the DOR using likelihood ratios is then

$$\text{DOR} = \frac{\widehat{\text{Se}} \times \widehat{\text{Sp}}}{(1 - \widehat{\text{Se}})\,(1 - \widehat{\text{Sp}})} = \frac{\widehat{\text{Se}} \big/ (1 - \widehat{\text{Sp}})}{(1 - \widehat{\text{Se}}) \big/ \widehat{\text{Sp}}} = \frac{\text{LR}_+}{\text{LR}_-}.$$

A good diagnostic test has a large positive likelihood ratio and a small negative likelihood ratio. Accordingly, a good diagnostic test has a large DOR.

Example 9.2 The DOR with 95 % confidence limits can be calculated using the `metabin` function.

```
> metabin(tp, n1, n2-tn, n2, data=data16, sm="OR",
+        comb.fixed=FALSE, comb.random=FALSE,
+        studlab=paste(author, year),
+        addincr=TRUE, allstudies=TRUE)
                    OR              95%-CI
Hricak 1988   67.6400 [10.2408; 446.7581]
Greco 1989     3.1818 [ 0.6536;  15.4906]
Janus 1989    15.4000 [ 1.4987; 158.2474]
```

```
Kim 1990       8.3067 [ 1.1097;  62.1816]
Ho 1992        2.8182 [ 0.0497; 159.9591]
Kim 1993      22.3333 [ 4.9958;  99.8391]
Hawnaur 1994 18.2099 [ 4.2099;  78.7671]
Kim 1994      67.9216 [23.3129; 197.8885]
Subak 1995    15.0331 [ 3.7599;  60.1067]
Heuck 1997    59.4000 [ 9.2046; 383.3273]
*** Output truncated ***
```

Note, we have used argument `addincr=TRUE` in order to add 0.5 to event counts in Table 9.1 for all the studies. In particular the argument `allstudies=TRUE` is necessary to calculate the DOR for Ho 1992 which has zero true positives and zero false positives.

We could use base R to calculate likelihood ratios. However, this can be done much more simply using function `madad` from R package **mada**:

```
> library(mada)
Loading required package: mvtnorm
Loading required package: ellipse
Loading required package: mvmeta
This is mvmeta 0.4.5. For an overview type:
    help('mvmeta-package').
> args(madad)
function (x = NULL, TP, FN, FP, TN, level = 0.95,
    correction = 0.5, correction.control = "all",
    method = "wilson", yates = TRUE, suppress = TRUE, ...)
```

This function like many other function in R package **mada** can be applied (1) to a data frame or matrix or (2) to a set of vectors `TP`, `FN`, `FP` and `TN`—corresponding to the number of true positive test results, etc. However, when applying R package **mada** to a data frame or matrix, this *must* have columns with the names `TP`, `FN`, `FP` and `TN`, otherwise an error is produced:

```
> md1 <- madad(data16)
Error in checkdata(origdata) :
  Data frame or matrix must have columns labelled TP, FN,
  FP and TN.
```

In order to use the `madad` function and other function in R package **mada**, we add four additional columns to dataset `data16`:

```
> data16$TP <- data16$tp
> data16$FN <- data16$n1 – data16$tp
> data16$FP <- data16$n2 – data16$tn
> data16$TN <- data16$tn
> md1 <- madad(data16)
```

These commands create an object of class `madad` for which the corresponding print function generates the following output. For better readability we split the output into two parts. The first part is:

```
> print(md1, digits=2)
Descriptive summary of data16 with 10 primary studies.
Confidence level for all calculations set to 95 %
Using a continuity correction of 0.5 if applicable
```

```
Diagnostic accuracies
      sens 2.5% 97.5% spec 2.5% 97.5%
 [1,] 0.79 0.51  0.93 0.95 0.84  0.98
 [2,] 0.39 0.15  0.69 0.83 0.69  0.92
 [3,] 0.70 0.30  0.93 0.87 0.65  0.96
 [4,] 0.22 0.08  0.46 0.97 0.87  0.99
 [5,] 0.08 0.01  0.48 0.97 0.76  1.00
 [6,] 0.25 0.13  0.43 0.99 0.95  1.00
 [7,] 0.74 0.50  0.89 0.87 0.72  0.94
 [8,] 0.62 0.46  0.76 0.98 0.95  0.99
 [9,] 0.61 0.36  0.81 0.91 0.81  0.96
[10,] 0.87 0.65  0.96 0.90 0.72  0.97

Test for equality of sensitivities:
X-squared = 38.2443, df = 9, p-value = 1.6e-05
Test for equality of specificities:
X-squared = 31.2101, df = 9, p-value = 0.00027
...
```

This first part of the output prints the sensitivity (column sens) and specificity (column spec) with corresponding 95 % confidence limits. Two χ^2-tests are conducted to test the hypotheses that sensitivities and specificities are equal across studies. These are calculated using the prop.test function from base R package **stats**. In this example, both hypotheses are rejected (p-values $<$ 0.001 and 0.0003 for sensitivity and specificity, respectively).

Looking closely, we see that sensitivity and specificity values differ from those in Sect. 9.2.1 above, which were obtained using the metaprop function. This discrepancy is because using a continuity correction is the default in the madad function but not for the metaprop function. To get the same estimates for sensitivity and specificity we can use the argument correction.control="none" in the madad function.[2]

The second part of the output prints DORs as well as likelihood ratios:

```
...
 Diagnostic OR and likelihood ratios
        DOR   2.5%  97.5% posLR  2.5%  97.5% negLR 2.5% 97.5%
 [1,] 67.64 10.24 446.76 14.88  4.30  51.46  0.22 0.07  0.66
 [2,]  3.18  0.65  15.49  2.33  0.79   6.86  0.73 0.43  1.26
 [3,] 15.40  1.50 158.25  5.32  1.46  19.32  0.35 0.09  1.33
 [4,]  8.31  1.11  62.18  6.71  1.08  41.66  0.81 0.62  1.05
 [5,]  2.82  0.05 159.96  2.67  0.06 119.92  0.95 0.73  1.22
 [6,] 22.33  5.00  99.84 17.00  4.29  67.42  0.76 0.62  0.94
 [7,] 18.21  4.21  78.77  5.56  2.24  13.76  0.31 0.14  0.68
 [8,] 67.92 23.31 197.89 26.54 11.20  62.89  0.39 0.26  0.59
 [9,] 15.03  3.76  60.11  6.51  2.65  16.03  0.43 0.22  0.84
[10,] 59.40  9.20 383.33  8.68  2.64  28.52  0.15 0.05  0.47
Correlation of sensitivities and FPRs:
   rho  2.5 % 97.5 %
  0.44  -0.26   0.84
```

---

[2]Confidence limits would be still different if we used the argument correction. control="none" as different methods are used to calculate these.

We get exactly the same values for DORs and corresponding confidence limits by using the `metabin` function command given above.

Finally, the Pearson correlation coefficient of the sensitivities (TPR) and FPR across studies is printed; here, it is 0.44. It is expected to be positive, but this is not always the case. In the example, the 95 % confidence interval for the correlation coefficient includes zero. Sensitivities, specificities, DORs and positive and negative likelihood ratios can all be displayed using forest plots, by analogy to meta-analyses of intervention studies. The `forest.madad` function which is hidden inside R package **mada** can be used for this purpose. Figure 9.2 showing forest plots for sensitivity and specificity generated using the following commands[3]:

```
> # Changes to plot layout:
> # - two plots (columns) in one row (argument mfrow)
> # - use maximal plotting region (argument pty)
> oldpar <- par(mfrow=c(1,2), pty="m")
> # Forest plot for sensitivities
> forest(md1, type="sens", main="Sensitivity")
> # Forest plot for specificities
> forest(md1, type="spec", main="Specificity")
> # Use previous graphical settings
> par(oldpar)
```
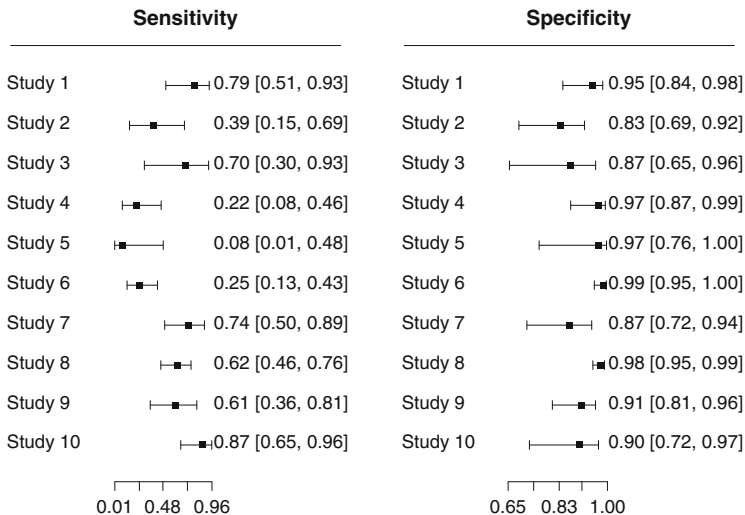

Fig. 9.2 Univariate forest plots for sensitivity and specificity in MR studies [18]

---

[3]We use R object `oldpar` in order to restore the settings of the graphics windows. This is recommended after changing these settings for a specific plot; however, from now on we do not display this command.

Fig. 9.3 Univariate forest plot for the log diagnostic odds ratio in MR studies [18]

Log diagnostic odds ratio

Study 1		4.21 [2.33, 6.10]
Study 2		1.16 [−0.43, 2.74]
Study 3		2.73 [0.40, 5.06]
Study 4		2.12 [0.10, 4.13]
Study 5		1.04 [−3.00, 5.07]
Study 6		3.11 [1.61, 4.60]
Study 7		2.90 [1.44, 4.37]
Study 8		4.22 [3.15, 5.29]
Study 9		2.71 [1.32, 4.10]
Study 10		4.08 [2.22, 5.95]

−3.00   1.55   6.10
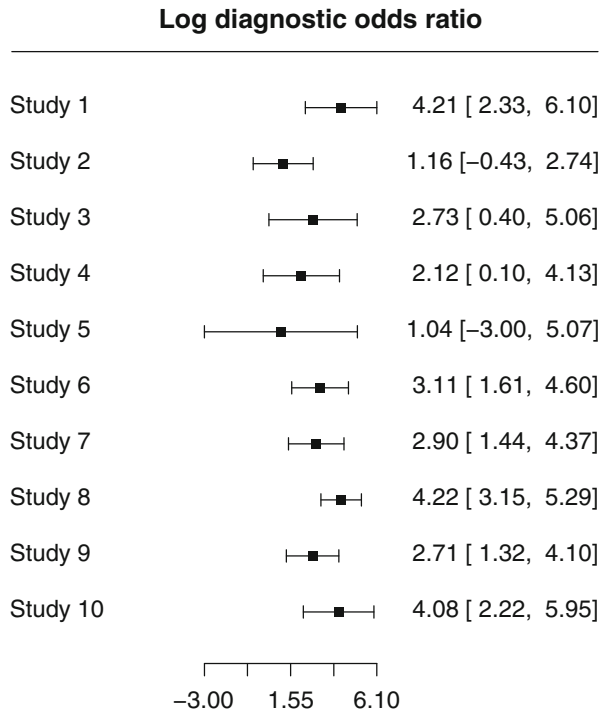
Other possible options for argument type are "DOR", "posLR", or "negLR". For these, we may prefer to use a log-transformed scale, which is achieved by using argument log=TRUE in the following command

```
> forest(md1, type="DOR", log=TRUE,
+        main="Log diagnostic odds ratio")
```

This gives the plot shown in Fig. 9.3.   □

9.2.3 Tests Based on a Continuous Marker

Diagnostic tests are often based on some continuous biomarker that is known to distinguish diseased and non-diseased individuals. For example, patients with diabetes mellitus have higher values of the biomarker HbA_{1c} than healthy people (see Example 8.4).

Let X be the continuous diagnostic marker that underlies a test, and consider two distinct probability distributions for X among diseased and non-diseased individuals, respectively. Without loss of generality, we assume that the diagnostic
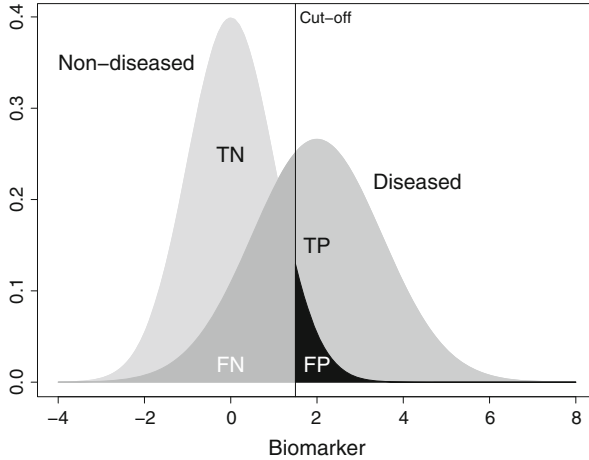
Fig. 9.4 Distributions of a continuous biomarker for diseased and non-diseased individuals with a specific cut-off. The probabilities for TN, FN, FP and TP are proportional to the shaded areas

marker tends to be higher for diseased patients than for non-diseased individuals. An idealised representation is shown in Fig. 9.4.[4]

A test is typically defined by determining a threshold or cut-off x_0 and giving individuals with $X \leq x_0$ a negative test result and individuals with $X > x_0$ a positive test result. This dichotomisation yields a two-by-two table as given in Table 9.1.

A test defined in this way is usually not perfect (unless the two distributions do not overlap). A consequence of overlapping distributions is that the threshold is not unambiguously defined. If, e.g., the cut-off point moves to the left, the total number of positive test results and thus both TPR (i.e. sensitivity) and FPR increase, whereas both TNR (i.e. specificity) and FNR decrease. Conversely, if the cut-off point moves to the right, the specificity increases, but sensitivity decreases.

It follows that, for a given biomarker, for each possible cut-off there is a different two-by-two table and therefore a distinct (sensitivity, specificity) pair. There is also a trade-off between sensitivity and specificity, as they are negatively correlated within a study. The variation in (sensitivity, specificity) as the cut-off, x_0, varies is typically shown by the receiver operating characteristic (ROC) curve (a term which comes from information theory) [13]. This is obtained by plotting the TPR (sensitivity) on the vertical axis (*y*-axis) against the FPR (one minus specificity) on the horizontal axis (*x*-axis). A typical ROC curve is shown in Fig. 9.5.[5] The dot represents the pair of sensitivity and specificity for a specific cut-off value. In Fig. 9.5 the dot refers to the cut-off given in Fig. 9.4. The better the test, the closer to the top left corner the curve lies, and the closer the area under the curve is to 1.

---

[4]R code to generate the figure is given in the web-appendix.

[5]R code to generate the ROC curve is given in the web-appendix.
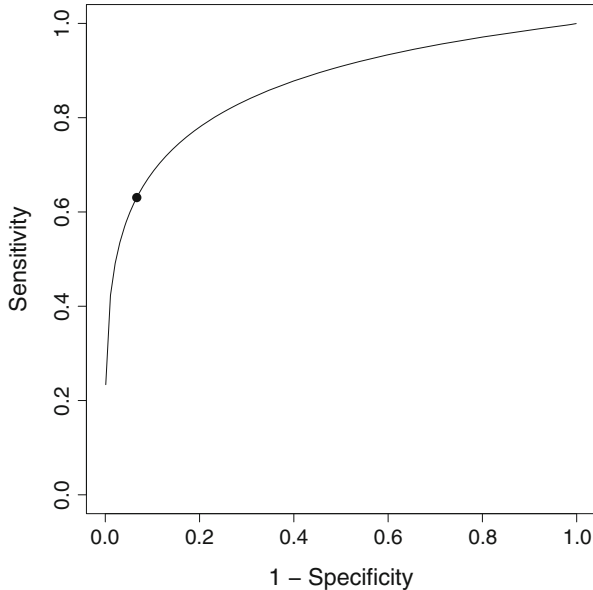
Fig. 9.5   A ROC curve for distributions shown in Fig. 9.4

In published DTA studies, often only a single (sensitivity, specificity) pair is reported, without information on the underlying cut-off value. In other words, information on other (sensitivity, specificity) pairs which are needed to plot the whole ROC curve for that study is typically missing. Thus, usually the available data look like those shown in Table 9.1. The mutual dependency between sensitivity and specificity—and the importance of this correlation for the properties of the test—is the reason why they need to be analysed as a bivariate outcome.

We note that an implicit ROC curve may even underlie diagnostic tests in radiology that are based on imaging methods such as magnetic resonance or computed tomography. This is because, when judging an image, raters may take a more sensitive or a more specific point of view. In this sense, rating an image is comparable to measuring a biomarker.

For the remainder of this section we assume that each study reports only the numbers TP, FN, TN and FP, as in Table 9.1, that is, there is only one pair of sensitivity and specificity per study. This means that the study specific ROC curves are unknown.

9.3   Scatterplot of Sensitivity and Specificity

So far, sensitivity and specificity have been summarised and plotted separately. However, statistical models for DTA studies model the bivariate distribution of sensitivity and specificity. Before fitting such models, it is a good idea to examine

a scatterplot of sensitivities and specificities. We can use base R to produce a scatterplot of FPR and sensitivity as follows:

```
> par(pty="s") # use a square plotting region
> plot(1-md1$spec$spec, md1$sens$sens,
+      xlim=c(0,1), ylim=c(0,1),
+      xlab="False positive rate (1-Specificity)",
+      ylab="Sensitivity", pch=16)
```

The resulting plot is shown in Fig. 9.6. The large heterogeneity of sensitivities as compared to the small heterogeneity in specificities is clearly visible. The `par` command is used to produce a square plot, which is natural because the *x*- and *y*-axis have the same limits.

As previously noted, heterogeneity between DTA studies is to be expected, and has two principal causes [17]. First, there may well be variation between DTA studies in the cut-off value used to dichotomize the underlying measure into a test result. If there is no other source of heterogeneity, this leads us to observing a number of points from a single ROC curve common to all the studies. However, secondly, accuracy may vary between studies due to clinical heterogeneity in patient populations and/or differences in the implementation of the diagnostic test. For this reason, the observed points need not lie on a common ROC curve at all.

In Fig. 9.6 variation in cut-off points seems to be larger than variation in accuracy. However, the heterogeneity in the precision of point estimates (i.e. pairs of sensitivity and specificity) is not shown. To show this uncertainty graphically, the `ROCellipse` function from R package **mada** can be used. It allows the plotting
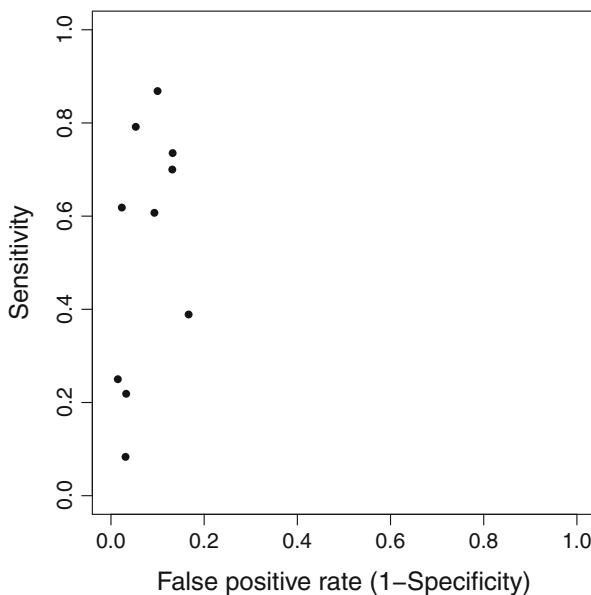
Fig. 9.6 Scatter plot of (1—specificity) and sensitivity, i.e. in the ROC space, for the MR data [18]
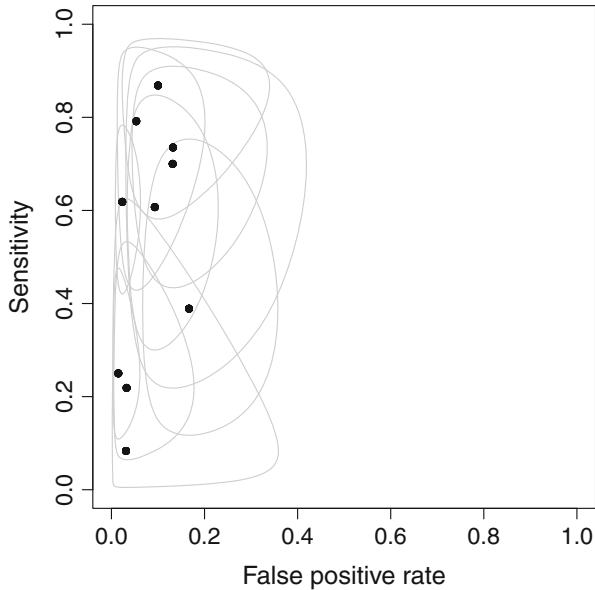
Fig. 9.7 Estimates from primary studies [18] with confidence regions in ROC space

of individual confidence regions for the estimates from the primary studies in the ROC space (these regions are assumed to be ellipsoids in the logit space, and back-transformed to the ROC space). Figure 9.7 was generated using the following R commands:

```
> par(pty="s") # use a square plotting region
> ROCellipse(data16, pch=16)
```

The `ROCellipse` function automatically uses the `TP`, `FP`, `TN`, `FN` variables from the dataset `data16`. The uncertainty is clearly visible. In particular, the smaller studies, e.g. Ho 1992 at the lower left end, have very wide confidence ellipsoids.

9.4 Models for Meta-Analysis of Diagnostic Test Accuracy Studies

The information shown in Fig. 9.6 is the basis for statistical analysis of DTA studies. Such meta-analyses may have several aims. First, we may want to estimate an average (sensitivity, specificity) pair along with a joint confidence region. Second, we may wish to estimate a prediction region where future pairs are expected to be found. Third, we may be interested in a summary ROC (SROC) curve across the observed studies.

For these goals statistical modelling is necessary. Broadly speaking, two models have become established during the last decade, a hierarchical model [17] and a bivariate model [14]. However, as two groups of researchers independently showed [1, 8], hierarchical and bivariate models are equivalent in the special (and most common) case of absence of covariates. Therefore, in the R package **mada**, the bivariate model is fitted with R function reitsma; however, the output also gives estimates of the parameters of the hierarchical model.

9.4.1 Hierarchical Model

The first authors who developed a model accounting for heterogeneity and study size were Rutter and Gatsonis [17] who proposed a hierarchical model, with an empirical Bayes version added by Macaskill [11]. The model included random effects for the cut-off and the test accuracy and focussed on estimating the SROC curve.

At the study level, it is assumed that within study k, $k = 1, \ldots, K$, the true positives (TP) and the false positives (FP) follow binomial distributions:

$$\text{TP} \sim \text{Binomial}(n_{k1}, \text{Se}_k)$$

$$\text{FP} \sim \text{Binomial}(n_{k2}, 1 - \text{Sp}_k),$$

where index 1 denotes diseased individuals and index 2 denotes non-diseased.

The authors parameterised the sensitivities and specificities as follows:

$$\text{logit}(\text{Se}_k) = (\theta_k + \alpha_k/2)e^{-\beta/2}$$

$$\text{logit}(1 - \text{Sp}_k) = (\theta_k - \alpha_k/2)e^{\beta/2},$$

where θ_k is the random threshold in study k, α_k is the random accuracy in study k, and β is a shape (asymmetry) parameter. Normal distributions are used to model variation in the study-specific parameters across studies:

$$\theta_k \sim N(\theta, \tau_\theta^2), \quad \alpha_k \sim N(\lambda, \tau_\alpha^2).$$

In total, the model has five parameters:

- mean and variance of the cut-off (θ, τ_θ^2),
- mean and variance of accuracy (λ, τ_α^2),
- shape parameter β.

A value of $\beta = 0$ would represent symmetry about the antidiagonal of the ROC space. The SROC curve is calculated by applying the inverse function of the logit (sometimes called "expit") to a function that is linear in $\text{logit}(1 - \text{Sp})$:

$$\text{Se} = \text{logit}^{-1}\left\{e^{-\beta}\text{logit}(1 - \text{Sp}) + \lambda e^{-\frac{\beta}{2}}\right\}.$$

This is equivalent to

$$\text{Se} = \left\{ 1 + \exp\left(-e^{-\beta} \log \frac{1 - \text{Sp}}{\text{Sp}} - \lambda e^{-\frac{\beta}{2}} \right) \right\}^{-1},$$

see [11, 17].

9.4.2 Bivariate Model

By contrast, Reitsma et al. [14] proposed a bivariate model of the joint distribution of sensitivity and specificity, allowing for across-study correlation. This model followed an approach developed for meta-analysis of binary outcomes [10]. It was then refined by others [1, 3].

At the study level, this model assumed that the true positives and false positives within study k, $k = 1, \ldots, K$, follow binomial distributions (as in Sect. 9.4.1). At the between-studies level, a bivariate random effects model is assumed for $\text{logit}(\text{Se}_k)$ and $\text{logit}(1 - \text{Sp}_k)$, where normal priors are assumed for the study-specific parameters:

$$\begin{pmatrix} \text{logit}(\text{Se}_k) \\ \text{logit}(1 - \text{Sp}_k) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right).$$

Alternatively, the covariance can be parametrised using the correlation coefficient ρ and the standard errors such that $\sigma_{12} = \rho \sigma_1 \sigma_2$. Either way, this model has also five parameters:

- means μ_1, μ_2,
- variances σ_1^2, σ_2^2,
- covariance σ_{12} (or, alternatively, correlation coefficient ρ).

Example 9.3 As mentioned at the beginning of this section, only the bivariate model can be fitted in R package **mada**, using the reitsma function. We now illustrate this with the MR dataset.

```
> mrfit <- reitsma(data16)
> print(summary(mrfit), digits=2)
Call:  reitsma.default(data = data16)

Bivariate diagnostic random-effects meta-analysis
Estimation method: REML

Fixed-effects coefficients
                 Estimate Std. Error      z Pr(>|z|) 95%ci.lb
tsens.(Intercept)    0.22       0.36   0.62     0.54    -0.49
tfpr.(Intercept)    -2.70       0.32  -8.39     0.00    -3.34
sensitivity          0.56          -      -        -     0.38
false pos. rate      0.06          -      -        -     0.03
```

```
                   95%ci.ub
tsens.(Intercept)     0.94
tfpr.(Intercept)     -2.07 ***
sensitivity           0.72
false pos. rate       0.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Variance components: between-studies Std. Dev and correlation
     matrix
     Std. Dev tsens tfpr
tsens    0.94  1.00    .
tfpr     0.80  0.62 1.00

logLik    AIC    BIC
 18.16 -26.32 -21.34

AUC:  0.844
Partial AUC (restricted to observed FPRs and normalized):  0.571

HSROC parameters
     Theta       Lambda        beta sigma2theta sigma2alpha
     -1.36         3.13       -0.16        0.61        0.57
```

The output first shows the estimation method (here REML, i.e. restricted maximum likelihood), then the estimates for the bivariate mean (on the logit scale) for sensitivity (tsens.(Intercept)) and FPR (tfpr.(Intercept)) with the corresponding standard errors, *z*-statistics *p*-values, and 95 % confidence limits. Next, the average sensitivity (0.56) and FPR (0.06) are presented. The correlation matrix is provided in the paragraph headed Variance components; the estimated correlation across studies is 0.62.

The program also provides estimates of the log-likelihood and the information criteria AIC (Akaike information criterion) and BIC (Bayesian information criterion) [2], which may be used to guide model choice.

In many cases, we will wish to summarize the results in an estimate of the area under the SROC curve (AUC). For a useful diagnostic test, the AUC should be markedly greater than 0.5 (the maximum possible value is 1). The software gives two estimates of the AUC, one for the whole ROC curve (here 0.844) and another using only the region where FPRs of studies were actually observed, and then normalised to the whole space. In our example the latter, partial AUC is 0.571, which is much smaller. The difference alerts us to the fact that the region in which the observed data lie is rather narrow, so we have limited direct knowledge about the shape of the overall ROC curve, and hence the AUC, from the data.

Finally, because of the equivalence of the bivariate model and the hierarchical model in the case of no covariates, the software is able to give the parameter estimates for the corresponding the hierarchical model (HSROC parameters): Theta is the estimated mean $\hat{\theta}$ of the cut-off (with variance sigma2theta), Lambda is the estimated mean accuracy parameter $\hat{\lambda}$ (with variance sigma2alpha), and beta is the estimated shape parameter $\hat{\beta}$.   □

9.5 Methods for Estimating a Summary ROC Curve

We have already looked at the scatterplot of pairs of FPRs (i.e. 1–specificity) versus the TPRs (i.e. sensitivity). Now we are interested in estimating the SROC curve across the studies. The simplest, somewhat naïve idea is to regress the logit-transformed TPRs against the logit-transformed FPRs and then back-transform to the ROC space. However, this curve addresses only one very specific question, i.e. "What is the sensitivity, given the specificity?" [1, Eq. (13)]. Further, as usual in regression, exchanging sensitivity and specificity gives a different curve [1, Eq. (14)]. Neither curve is symmetric with respect to sensitivity and specificity and both can be misleading [15].

The first attempt to defining a SROC curve in a different way was made by Moses et al. [12]. This approach is based on a regression of the difference of the logit-transformed positive rates (that is, the logarithm of the DOR) against their sum (which is a proxy for the threshold, see [1, Eq. (15)]). It is symmetric with respect to sensitivity and specificity, but does not account for potential heterogeneity or for the different precision of the estimates from different studies. The proposal by Rutter and Gatsonis [11, 17] leads to a different solution. The slope of this line in the logit space is the geometric mean of the slopes of the two regression lines, logit(Se) on logit$(1 - Sp)$ and vice versa. [1, Eq. (16)].

Arends et al. [1] pointed out that the SROC curve is in principle unidentifiable if only one (Se, Sp) pair per study is known. Rücker and Schumacher [16] proposed an SROC curve based on the assumption that investigators of the primary studies, after considering the whole empirical ROC curve, selected the cut-off that maximised some optimality criterion for reporting, e.g., the Youden index. Holling et al. [9] defined a parametric SROC curve based on the assumption of a constant ratio of log(TPR) and log(FPR). Doebler et al. [6] considered a more general family of transformations (t_α transformations) which includes the logit transformation as the special case $\alpha = 1$. The generalisations by Holling et al. and Doebler et al. are available in the R package **mada**, but will not be discussed here.

As we have already noted, none of the SROC curves proposed in the literature can be interpreted as an average ROC curve over all studies without making further assumptions. An example of a particularly challenging situation with a positive correlation between sensitivity and specificity is the lymphangiography data presented by Scheidler et al. [18], which has been also analysed by many others [4, 8, 11, 14, 16, 17, 19] and which we consider in the next example.

Example 9.4 The R object `mrfit` generated with the `reitsma` function can be used to plot both the mean sensitivity and specificity as well as the SROC curve proposed by Rutter and Gatsonis [17] using the `plot` function from R package **mada**:

```
> par(pty="s") # use a square plotting region
> plot(mrfit, predict=TRUE, cex=2)
> points(1-md1$spec$spec, md1$sens$sens, pch=16)
```
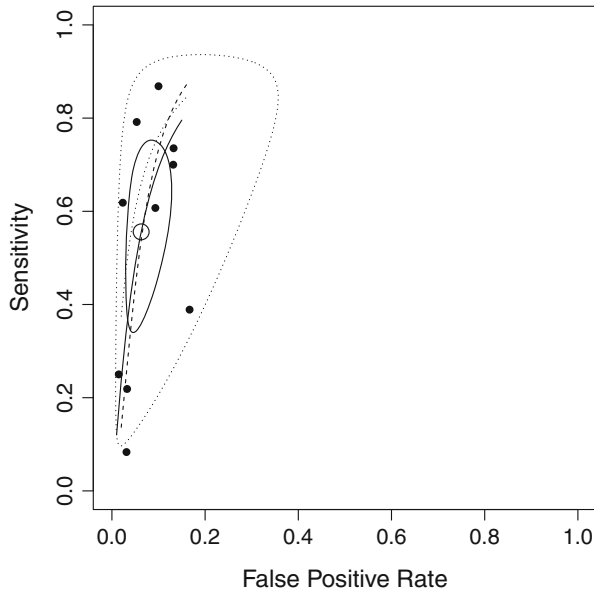
Fig. 9.8 Joint estimate of FPR and sensitivity for the MR data with 95 % confidence and prediction regions. *Solid closed curve*: 95 % confidence region; *dotted closed curve*: 95 % prediction region. In addition, three SROC curves are seen. *Short solid line*: curve proposed by Rutter and Gatsonis [17]; *dashed line*: curve proposed by Moses et al. [12]; *dotted line*: curve proposed by Rücker and Schumacher [16]

These commands give Fig. 9.8. The estimated mean (FPR, TPR) pair is shown by the open circle. It is surrounded by its 95 % confidence region (solid closed curve) and a 95 % prediction region (dotted closed curve). The SROC curve proposed by Rutter and Gatsonis [17] is shown by the black solid curve through the estimated mean (FPR, TPR) pair.

The curve proposed by Moses et al. [12] was added to Fig. 9.8 using a dashed line with the `mslSROC` function:

```
> # Argument lty=2 gives the dashed line
> mslSROC(data16, lty=2, add=TRUE)
```

Finally, the curve proposed by Rücker and Schumacher [16] was added as a dotted line, using the `rsSROC` function:

```
> # Argument lty=3 gives the dotted line
> rsSROC(data16, lty=3, add=TRUE)
```

The Rücker and Schumacher curve is based on the assumption that for each primary study, the investigator has chosen the cut-off such that a weighted mean of sensitivity and specificity (λSe $+ (1 - \lambda)$Sp) with a common parameter λ ($0 < \lambda < 1$) was maximised. The method also gives an estimate of λ, which represents the weight which is attributed to the sensitivity, whereas $1 - \lambda$ represents the weight

attributed to the specificity.[6] We can request the estimate of this parameter using the commands[7]

```
> rs <- rsSROC(data16)
> lambda <- rs$lambda
> c(lambda, 1-lambda)
[1] 0.2041087 0.7958913
```

This indicates that investigators attributed about 80 % of the weight to the specificity and only about 20 % to the sensitivity. This is reflected by the fact that the points tend to lie near the left-hand border of the ROC space, with some sensitivities being quite poor, see Fig. 9.5.

Obviously, the three curves do not completely agree in this example. Moreover, the differences may be more striking with other examples. The reason is that the distinct curves are based on different assumptions. Our example illustrates the fact that the construct of a "SROC curve" is not unambiguously defined.   □

9.6 Summary

In this chapter, we have given an overview of the issues raised in meta-analysis of DTA studies, which may be seen as a special case of multivariate meta-analysis. After introducing the basic concepts of diagnostic tests, particularly those based on a continuous biomarker, we presented the two principal models for meta-analysing DTA study data, noting they are equivalent when (as will typically be the case) we have no covariates. We then showed how the R package **mada** can be used to perform meta-analysis of DTA studies using the bivariate model. It also provides estimates of the parameters from the hierarchical model and in addition some univariate measures of accuracy such as the DOR and positive and negative likelihood ratios. The package also allows a number of useful plots to be easily created. These include confidence regions for the study-level estimates of sensitivity and specificity. To this we can add a pooled estimate of the (sensitivity, specificity) pair, a prediction region, and a number of SROC curves. The package has a number of other capabilities, which can be found in the documentation and particularly in the vignette provided at http://cran.r-project.org/web/packages/mada/vignettes/mada.pdf.

---

[6]Note that this parameter should not be confused with the parameter λ of the hierarchical model.

[7]Even though we are only interested in an estimate of parameter λ, the `rsSROC` command generates a plot.

References

1. Arends, L.R., Hamza, T.H., van Houwelingen, J., Heijenbrok-Kal, M., Hunink, M., Stijnen, T.: Bivariate random effects meta-analysis of ROC curves. Med. Decis. Making **28**(5), 621–638 (2008)
2. Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York (2002)
3. Chu, H., Cole, S.R.: Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed approach. J. Clin. Epidemiol. **59**, 1331–1333 (2006)
4. Chu, H., Guo, H.: Letter to the editor. Biostatistics **10**(1), 201–203 (2009)
5. Doebler, P.: mada: Meta-analysis of diagnostic accuracy (2014). http://www.cran.R-project.org/package=mada. R package version 0.5.6
6. Doebler, P., Holling, H., Böhning, D.: A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. Psychol. Methods **17**(3), 418–436 (2012). doi:10.1037/a0028091
7. Hamza, T.H., van Houwelingen, H.C., Stijnen, T.: Random effects meta-analysis of proportions: the binomial distribution should be used to model the within-study variability. J. Clin. Epidemiol. **61**(1), 41–51 (2007). doi:10.1016/j.jclinepi.2007.03.016
8. Harbord, R.M., Deeks, J.J., Egger, M., Whiting, P., Sterne, J.A.: A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics **8**, 239–251 (2007)
9. Holling, H., Böhning, W., Böhning, D.: Meta-analysis of diagnostic studies based upon SROC-curves: a mixed model approach using the lehmann family. Stat. Model. **12**(4), 347–375 (2012)
10. van Houwelingen, H.C., Zwinderman, K.H., Stijnen, T.: A bivariate approach to meta-analysis. Stat. Med. **12**, 2273–2284 (1993)
11. Macaskill, P.: Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. J. Clin. Epidemiol. **57**(9), 925–932 (2004)
12. Moses, L., Shapiro, D., Littenberg, B.: Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat. Med. **12**(14), 1293–1316 (1993)
13. Pepe, M.S.: The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, Oxford (2004)
14. Reitsma, J., Glas, A., Rutjes, A., Scholten, R., Bossuyt, P., Zwinderman, A.: Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J. Clin. Epidemiol. **58**(10), 982–990 (2005)
15. Rücker, G., Schumacher, M.: Letter to the editor. Biostatistics **10**(4), 806–807 (2009)
16. Rücker, G., Schumacher, M.: Summary ROC curve based on the weighted Youden index for selecting an optimal cutpoint in meta-analysis of diagnostic accuracy. Stat. Med. **29**, 3069–3078 (2010)
17. Rutter, C.M., Gatsonis, C.A.: A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat. Med. **20**, 2865–2884 (2001)
18. Scheidler, J., Hricak, H., Yu, K., Subak, L., Segal, M.: Radiological evaluation of lymph node metastases in patients with cervical cancer. A meta-analysis. JAMA **278**(13), 1096–1101 (1997)
19. Walter, S.D.: Properties of the summary receiver operating characteristic (ROC) curve for diagnostic test data. Stat. Med. **21**, 1237–1256 (2002)
20. Willis, B.H., Quigley, M.: Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. BMC Med. Res. Methodol. **11**, 27 (2011)

Appendix A
Further Information on R

Many valuable information and files for installation of R can be found on the Comprehensive R Archive Network (CRAN). The user is advised to choose a mirror of CRAN which is close to the current location. A list of all mirrors is available on the website http://cran.r-project.org/mirrors.html. At the time of writing about 100 mirrors from 48 countries exist.

On CRAN, a general R FAQ [3] and FAQs for Mac OS and Windows are provided.

Furthermore, the following manuals are available on CRAN.

1. An Introduction to R
2. R Data Import/Export
3. R Installation and Administration
4. Writing R Extensions
5. The R language definition (draft)
6. R Internals
7. The R Reference Index

For beginners, manuals 1 and 2 are of most interest.

A.1 Installation of R

Precompiled binary distributions of R are available on CRAN for the major operation systems Linux, Windows and Mac OS. Furthermore, R can be configured and built by the user on many Unix-like operation systems [3].

In order to install R for Windows or Mac OS, go to your favourite CRAN mirror, open the webpage *Download R for Windows* or *Download R for (Mac) OS X* and download the EXE- or PKG-file of the precompiled binary distribution, respectively.

For Debian, Ubuntu, Red Hat and SuSE Linux, instructions to install precompiled binaries are available on your favourite CRAN mirror on the webpage *Download R for Linux*.

For all other operation systems, the source code can be downloaded from any CRAN mirror.

A.2   Importing Data into R

The R manual 'R Data Import/Export' available on CRAN is obviously a good starting point for information on data import. Furthermore, a Use-R! book on data manipulation has two chapters on reading and writing data [11].

A.2.1   Import Text Files

Throughout the book we have used the `read.csv` to import data in R from a text file. This function expects that fields are separated by a comma whereas a dot is used as decimal point. Some information on this and similar R functions is given in the following listing.

R function	Separator (argument `sep`)	Decimal point (argument `dec`)
`read.csv`	Comma	Dot
`read.csv2`	Semicolon	Comma
`read.delim`	Tabulator	Dot
`read.delim2`	Tabulator	Comma
`read.table`	White Space	Dot

Note, these functions only differ in default values for `sep` and `dec` as well as other arguments (see shared help page for these commands).

The following function calls all result in the same dataset.

```
> rd1 <- read.csv("dataset01.csv", as.is=TRUE)
> rd2 <- read.csv2("dataset01.csv", as.is=TRUE,
+                  sep=",", dec=".")
> rd3 <- read.delim("dataset01.csv", as.is=TRUE,
+                   sep=",", dec=".")
> rd4 <- read.delim2("dataset01.csv", as.is=TRUE,
+                    sep=",", dec=".")
> rd5 <- read.table("dataset01.csv", as.is=TRUE,
+                   sep=",", dec=".", header=TRUE)
```

which can be checked using the following commands

```
> all.equal(rd1, rd2)
[1] TRUE
```

```
> all.equal(rd1, rd2)
[1] TRUE
> all.equal(rd1, rd3)
[1] TRUE
> all.equal(rd1, rd4)
[1] TRUE
> all.equal(rd1, rd5)
[1] TRUE
```

A.2.2   Import Data from RevMan 5

It is possible to export the analyses data from RevMan 5 [13] to a CSV-file (Comma Separated Values). Whereas it is in principle possible to import this CSV-file in R using any of the above-mentioned R functions, a much more convenient way to import these data is using the read.rm5 function from R package **meta**.

The following steps are necessary in RevMan 5 to create the CSV-file:

1. Open the Menu *File → Export → Data and analyses*.
2. Click 'Next' at the bottom of the dialog box *Which analyses would you like to export?*.
3. Select all items but 'Risk of bias tables' in the dialog box *Which fields do you want to include?* and click 'Next'.
4. Click 'Finish' and save the CSV-file.

The resulting CSV-file can be imported in R using the read.rm5 function.

```
> examples <- read.rm5("Examples from Meta-Analysis with R.csv",
+                       numbers.in.labels=FALSE)
> dim(examples)
[1] 31 54
> class(examples)
[1] "rm5"        "data.frame"
```

The read.rm5 function creates a data frame with the additional class rm5. This data frame contains 31 observations, i.e. studies which are included in meta-analyses, and 54 variables with detailed information to conduct the meta-analyses. The following command prints some information for this data frame.

```
> examples[, 1:6]
   comp.no outcome.no group.no    studlab year event.e
1        1          1        1  Boner 1988 1988       0
2        1          1        1  Boner 1989 1989       0
*** Output truncated ***
17       1          1        1 Todaro 1993 1993       0
18       2          1        1    De Souza   NA      14
*** Output truncated ***
31       2          1        1      Vitolo   NA      35
```

The first 17 observations contain data from the bronchoconstriction meta-analysis (see Fig. 1.2) and the following 14 observations from the high-dose chemotherapy meta-analysis (see Fig. 3.1).

The `metacr` function can be used to conduct a meta-analysis for the RevMan 5 data.

```
> args(metacr)
function (x, comp.no = 1, outcome.no = 1, method, sm,
    level = .settings$level,  level.comb = .settings$level.comb,
    comb.fixed, comb.random,
    hakn = FALSE, method.tau = "DL", tau.common = FALSE,
    prediction = .settings$prediction,
    level.predict = .settings$level.predict, swap.events,
    logscale, backtransf = .settings$backtransf,
    title, complab, outclab, keepdata = .settings$keepdata,
    warn = FALSE)
NULL
```

Argument `x` is a RevMan 5 object and arguments `comp.no` and `outcome.no` define which meta-analysis to conduct. The other arguments can be used to conduct a meta-analysis with different setting than in RevMan 5.

By default, the `metacr` function conducts a meta-analysis for the first comparison and first outcome, i.e. data from the bronchoconstriction meta-analysis in our example file. In order to do a meta-analysis for the high-dose chemotherapy data, we have to specify arguments `comp.no` and `outcome.no`.

```
> mc1.cr <- metacr(examples)
> mb1.cr <- metacr(examples, 2, 1)
```

Based on the CSV-files for each meta-analysis separately, we could conduct these meta-analyses in the following way.

```
> # 1. Read in the data
> data1 <- read.csv("dataset01.csv", as.is=TRUE)
> # 2. Conduct meta-analysis
> mc1.md <- metacont(Ne, Me, Se, Nc, Mc, Sc,
+                    data=data1, studlab=paste(author, year),
+                    comb.random=FALSE)
> # 3. Read in the data
> data7 <- read.csv("dataset07.csv")
> # 4. Conduct meta-analysis
> mb1.rr <- metabin(Ee, Ne, Ec, Nc, data=data7, studlab=study,
+                    comb.random=FALSE)
```

These meta-analyses will give the same results as those based on the CSV-file from RevMan 5 and the `metacr` function. We only show results for Example 2.1.

The class of the R objects are identical

```
> class(mc1.cr)
[1] "metacont" "meta"
> class(mc1.md)
[1] "metacont" "meta"
```

as well as the meta-analytical results

```
> print(summary(mc1.cr), digits=2)
Review:     Examples from Meta-Analysis with R
Comparison: Chapter 2 - Meta-Analysis with Continuous Outcomes
Outcome:    Example 2.1 - Nedocromil sodium for bronchoconstriction

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

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
> print(summary(mc1.md), 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

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
```

Only difference in the printout is a header giving details on the review, comparison, and outcome which is only printed for the meta-analysis conducted using the `metacr` function. However, we could add the header using arguments `title`, `complab` and `outclab` in the `metacont` function.

Another nice thing about using the `read.rm5` function is that we can conduct tests for small-study effects for all meta-analyses using a single command.

```
> metabias(examples)
Review:     Examples from Meta-Analysis with R
Comparison: Chapter 2 - Meta-Analysis with Continuous Outcomes
Outcome:    Example 2.1 - Nedocromil sodium for bronchoconstriction


Linear regression test of funnel plot asymmetry

data:  m1
```

```
t = −1.1828, df = 15, p-value = 0.2553
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias        slope
 −0.9571707   0.8092752 −11.0548678


*****

Review:     Examples from Meta-Analysis with R
Comparison: Chapter 3 - Meta-Analysis of Binary Outcomes
Outcome:    Example 3.1 - Stem Cell Transplantation


Linear regression test of funnel plot asymmetry

data:  m1
t = 0.185, df = 12, p-value = 0.8564
alternative hypothesis: asymmetry in funnel plot
sample estimates:
     bias      se.bias       slope
0.16233218 0.87767199 0.09201333
```

A similar function is available to print a summary of all meta-analyses, i.e. the `summary.rm5` function.

A.3   R Packages for Meta-Analysis

In our book we primarily used the R packages **meta**, **metasens**, **netmeta**, **mvmeta** and **mada**. In this section we will give some information on other R packages for meta-analysis which are also available on CRAN. Furthermore, we would like to refer the reader to the CRAN 'Task View on meta-analysis' with short descriptions on available R packages for meta-analysis: http://cran.r-project.org/web/views/MetaAnalysis.html. We now briefly review some of the general meta-analysis packages and some recent network meta-analysis software developments.

A.3.1   General Purpose R Packages for Meta-Analysis

Besides the R package **meta** [8, 9], **metafor** [15] and **rmeta** [5] are two other general purpose R packages for meta-analysis providing the inverse variance and Mantel–Haenszel method. In addition, **metafor** provides the Peto method.

The following R code shows how to conduct a meta-analysis with a binary outcome using the `metabin` function from R package **meta**; by default risk ratio (see Sect. 3.1.2) and Mantel–Haenszel method (see Sect. 3.3.2) are used.

```
> # Make R package meta available
> library(meta)
```

```
Loading 'meta' package (version 4.0-2).
> # Do meta-analysis
> mb1.meta <- metabin(Ee, Ne, Ec, Nc, data=data7, studlab=study,
+                        comb.random=FALSE)
```

R code to conduct this meta-analysis with binary outcomes using the `rma.mh` function from R package **metafor** is given as follows.[1]

```
> # Make R package metafor available
> library(metafor)
Loading 'metafor' package (version 1.9-5). For an overview
and introduction to the package please type: help(metafor).

Attaching package: 'metafor'

The following objects are masked from 'package:meta':

   baujat, forest, funnel, labbe, radial, trimfill

> # Do meta-analysis using default settings (numbers of events)
> mb1.metafor <- rma.mh(Ee, Ne-Ee, Ec, Nc-Ec,
+                        data=data7, measure="RR")
> # Do meta-analysis using same input as R function metabin
> mb1.metafor <- rma.mh(ai=Ee, n1i=Ne, ci=Ec, n2i=Nc,
+                        data=data7, measure="RR")
```

You can safely ignore the warnings regarding R objects masked from R package **meta** as both R packages have been designed to work together. All masked functions act as wrapper functions for other functions in R packages **meta** and **metafor**, respectively. Note, we have to use argument `measure="RR"` as the default effect measure in the `rma.mh` function is the odds ratio.

The following R code can be used to conduct a meta-analysis using the Mantel–Haenszel method using the `meta.MH` function from R package **rmeta**.[2]

```
> # Make R package rmeta available
> library(rmeta)
Loading required package: grid
> mb1.rmeta <- meta.MH(Ne, Nc, Ee, Ec, data=data7,
+                        statistic="RR")
```

Again, we have to use an argument, here `statistic="RR"`, to specific the risk ratio as the default effect measure in the `meta.MH` function is the odds ratio.

First, we print the meta-analysis result of R object `mb1.meta` created with the R package **meta** using the `print.meta` function.

```
> print(summary(mb1.meta), digits=2)
```

---

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

[2]In order to do these analyses you have to install R package **rmeta** using the command `install.packages("rmeta")`.

```
Number of studies combined: k=14

                      RR       95%-CI   z  p.value
Fixed effect model 1.11 [1.04; 1.18] 3.2   0.0014

Quantifying heterogeneity:
tau^2 = 0.0079; H = 1.25 [1; 1.71]; I^2 = 35.6% [0%; 65.9%]

Test of heterogeneity:
    Q d.f.  p.value
 20.19   13   0.0906

Details on meta-analytical method:
- Mantel-Haenszel method
```

   Next, we print the meta-analysis result of R object mb1.metafor created with the R package **metafor** using the print.rma.mh function.

```
> print(mb1.metafor, digits=2)

Fixed-Effects Model (k = 14)

Test for Heterogeneity:
Q(df = 13) = 20.19, p-val = 0.09

Model Results (log scale):

estimate       se     zval     pval   ci.lb     ci.ub
    0.10     0.03     3.20     <.01     0.04      0.16

Model Results (RR scale):

estimate    ci.lb     ci.ub
    1.11     1.04      1.18
```

   Last, we print the meta-analysis result of R object mb1.rmeta created with the R package **rmeta** using the print.meta.MH function.

```
> print(mb1.rmeta)
Fixed effects ( Mantel-Haenszel ) Meta-Analysis
Call: meta.MH(ntrt = Ne, nctrl = Nc, ptrt = Ee, pctrl = Ec,
    data = data7, statistic = "RR")
Mantel-Haenszel RR =1.11    95% CI ( 1.04, 1.18 )
Test for heterogeneity: X^2( 13 ) = 20.19 ( p-value 0.0906 )
```

   As we can see in the printouts, all R packages provide comparable outputs using different printing formats.

   The rma.uni function from R package **metafor** can be used to do a meta-regression. Actually, the metareg function from R package **meta** calls this function internally. R package **metafor** also provides several functions for model diagnostics [15] which can be used with an R object generated with the rma.uni or metareg function.

A.3.2   R Packages to Conduct Network Meta-Analysis

Van Valkenhoef et al. [14] have published an R package **gemtc** that conducts network meta-analysis using a Bayesian hierarchical model.

A widely applied analysis method is the Bayesian approach, more specifically a Markov Chain Monte Carlo (MCMC) method [1, 2, 4, 6, 7]. It is implemented in special software such as `WinBUGS` [10]. We will not explain the Bayesian approach here. It is only noted that there exist a number of interfaces from R to `WinBUGS` and similar MCMC software, such as the R package **R2WinBUGS** [12]. These packages enable R users to manage their data in R such that they can be processed by the Bayesian software, using calls from the R environment.

References

1. Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E.: Checking consistency in mixed treatment comparison meta-analysis. Stat. Med. **29**(7–8), 932–944 (2010)
2. Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E.: Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. J. R. Stat. Soc. A. Stat. Soc. **173**, 613–629 (2010)
3. Hornik, K.: The R FAQ (2014). URL http://CRAN.R-project.org/doc/FAQ/R-FAQ.html
4. Lu, G., Ades, A.E.: Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics **10**(4), 792–805 (2009)
5. Lumley, T.: rmeta: Meta-analysis (2012). URL http://CRAN.R-project.org/package=rmeta. R package version 2.16
6. Salanti, G.: Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res. Synth. Methods **3**(2), 80–97 (2012). URL doi:10.1002/jrsm.1037
7. Salanti, G., Higgins, J.P., Ades, A.E., Ioannidis, J.P.: Evaluation of networks of randomized trials. Stat. Methods Med. Res. **17**(3), 279–301 (2008)
8. Schwarzer, G.: meta: An R package for meta-analysis. R News **7**(3), 40–45 (2007). URL http://cran.r-project.org/doc/Rnews/Rnews_2007-3.pdf
9. Schwarzer, G.: meta: Meta-Analysis with R (2014). URL http://cran.R-project.org/package=meta. R package version 4.0-2
10. Smith, T.C., Spiegelhalter, D.J., Thomas, A.: Bayesian approaches to random-effects meta-analysis: A comparative study. Stat. Med. **14**, 2685–2699 (1995)
11. Spector, P.: Data Manipulation with R. Springer, New York (2008)
12. Sturtz, S., Ligges, U., Gelman, A.: R2WinBUGS: A package for running WinBUGS from R. J. Stat. Softw. **12**(3), 1–16 (2005)
13. The Cochrane Collaboration: Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre (2014)
14. van Valkenhoef, G., Lu, G., de Brock, B., Hillege, H., Ades, A.E., Welton, N.J.: Automating network meta-analysis. Res. Synth. Methods **3**(4), 285–299 (2012)
15. Viechtbauer, W.: Conducting meta-analyses in R with the metafor package. J. Stat. Softw. **36**(3), 1–48 (2010)

Index