Solutions Practicals: Subgroup analysis, meta-regression and bias

R packages

We will use the package **readxl** to import excel data and the package **meta** to run meta-analyses.

```
library(readxl)
library(meta)
```

Results (treatment estimates and confidence intervals) should be rounded to two digits (default is four digits).

```
settings.meta(digits = 2)
```

Dataset "bcgtrial"

Dataset bcgtrial.xls contains data from the meta-analysis of the efficacy of BCG vaccine against tuberculosis performed by Colditz et al. (JAMA 1994; 271: 698-702). Open the dataset bcgtrial.xls by using the *Import Dataset* tab or type:

```
BCG = read_excel("bcgtrial.xls")
BCG <- as.data.frame(BCG)
```

Check the data structure:

```
str(BCG)
```

```
## 'data.frame':    11 obs. of  10 variables:
##  $ trial   : num  1 2 3 5 6 7 4 8 9 10 ...
##  $ authors : chr  "Aronson" "Ferguson & Simes" "Hart & Sutherland" "Vandev
iere et al" ...
##  $ startyr : num  1935 1933 1950 1965 1968 ...
##  $ latitude: num  52 55 53 18 13 27 13 42 18 33 ...
##  $ trialnam: chr  "Northern USA" "Canada" "UK" "Haiti" ...
##  $ pop1    : num  123 306 13598 2545 88391 ...
##  $ pop0    : num  139 303 12867 629 88391 ...
##  $ cases1  : num  4 6 62 8 505 29 33 17 186 5 ...
##  $ cases0  : num  11 29 248 10 499 45 47 65 141 3 ...
##  $ alloc   : chr  "Random allocation" "Random allocation" "Random allocati
on" "Random allocation" ...
```

Use the log risk ratio and its standard error to calculate risk ratios and 95% confidence intervals for each study:

```
BCG$rr <-       (BCG$cases1/BCG$pop1) / (BCG$cases0/BCG$pop0)

BCG$logrr <-    log(BCG$rr)

BCG$selogrr <-  sqrt(1/BCG$cases1 - 1/BCG$pop1 + 1/BCG$cases0 - 1/BCG$pop0)

BCG$loglci <-   BCG$logrr-1.96*BCG$selogrr

BCG$loguci <-   BCG$logrr+1.96*BCG$selogrr

BCG$lci <-      exp(BCG$loglci)

BCG$uci <-      exp(BCG$loguci)
```

Print the results for each study:

```
BCG[,c("trial", "trialnam", "startyr", "latitude", "rr", "lci", "uci")]
```

```
##    trial       trialnam startyr latitude        rr        lci       uci
## 1      1   Northern USA    1935       52 0.4109387 0.13429881 1.2574242
## 2      2         Canada    1933       55 0.2048682 0.08629608 0.4863600
## 3      3             UK    1950       53 0.2365605 0.17927998 0.3121424
## 4      5          Haiti    1965       18 0.1977210 0.07835524 0.4989277
## 5      6         Madras    1968       13 1.0120240 0.89456995 1.1448995
## 6      7   South Africa    1965       27 0.6253663 0.39257291 0.9962049
## 7      4    Madanapalle    1950       13 0.8044895 0.51628892 1.2535683
## 8      8        Chicago    1941       42 0.2537655 0.14941949 0.4309807
## 9      9    Puerto Rico    1949       18 0.7122268 0.57251139 0.8860384
## 10    10  Georgia (Sch)    1947       33 1.5619162 0.37367929 6.5285454
## 11    11 Georgia (Comm)    1950       33 0.9828351 0.58213186 1.6593574
```

Meta-analysis of the effect of BCG on tubercolosis

Perform a meta-analysis of the effect of BCG on tuberculosis.

```
ma1 = metabin(cases1, pop1, cases0, pop0, data = BCG,
                    studlab = trialnam, sm = "RR")
summary(ma1)
```

```
## Number of studies combined: k = 11
##
##                         RR       95%-CI     z  p-value
## Fixed effect model   0.70 [0.64; 0.76] -8.02 < 0.0001
## Random effects model 0.51 [0.34; 0.77] -3.20   0.0014
##
## Quantifying heterogeneity:
## tau^2 = 0.3851; H = 3.56 [2.87; 4.41]; I^2 = 92.1% [87.9%; 94.9%]
##
```

```
## Test of heterogeneity:
##       Q d.f.  p-value
##  126.63   10 < 0.0001
##
## Details on meta-analytical method:
## - Mantel-Haenszel method
## - DerSimonian-Laird estimator for tau^2
```

Compare the fixed-effects and random-effects estimates.

Fixed effects OR (95% CI):

Random effects OR (95% CI):

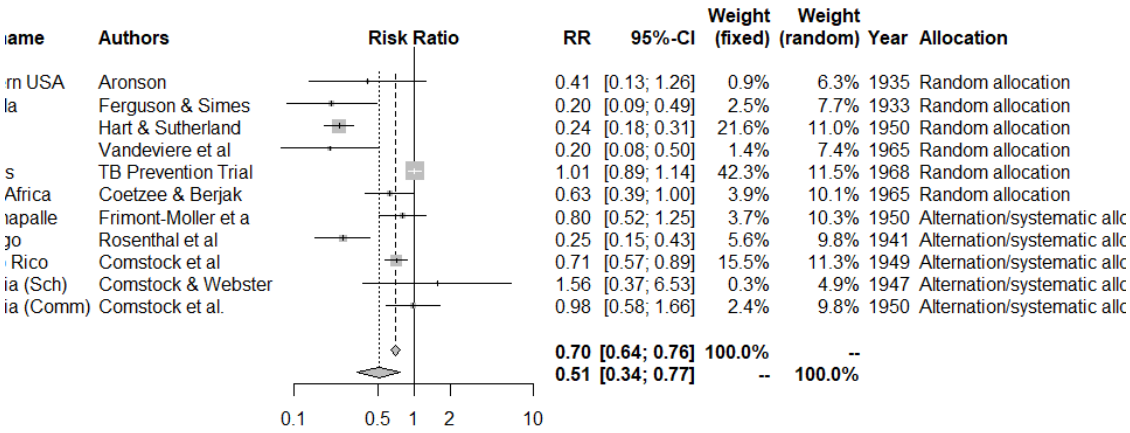
Compare the weights used in the two types of meta-analysis (run `print(ma1)` and check *%W(fixed)* and *%W(random)* in the output). Comment on the difference between the fixed- and random-effects estimates.

Answer: *While the weights vary considerably for the fixed-effects analysis, the random-effects weights are very similar. This is because of the large estimate of τ^2, the between-study variance. In consequence, the estimates from the fixed and random-effects methods vary considerably.*

Create forest plots

To obtain a forest plot we will use the function `forest`. You can use the `leftcols` and `rightcols` arguments to add columns displaying study characteristics to the graph.

```
forest(ma1, leftcols = c("trialnam","authors"), leftlabs=c("Trial name", "Authors"),
       just.addcols = "left", rightcols = c("effect", "ci", "w.fixed", "w.random", "startyr", "alloc"),
       rightlabs = c("Year", "Allocation"), digits.addcols.right = 0)
```



Subgroup analysis

Trials 1, 8, 10 and 11 were performed in the USA. Use the `byvar` argument of the `metabin` function to compare results in the USA from results elsewhere (you first need to create a variable to distinguish them).

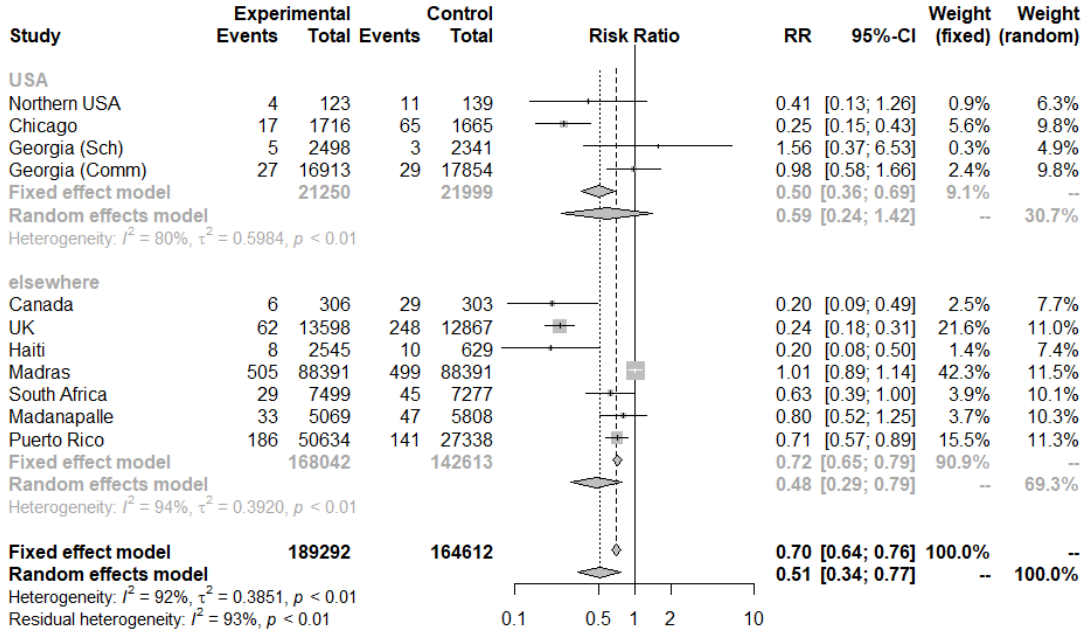
```
BCG$USA <- ifelse(BCG$trial %in% c(1,8,10,11), "USA", "elsewhere")

table(BCG$trialnam, BCG$USA)

##
##                   elsewhere USA
##   Canada                  1   0
##   Chicago                 0   1
##   Georgia (Comm)          0   1
##   Georgia (Sch)           0   1
##   Haiti                   1   0
##   Madanapalle             1   0
##   Madras                  1   0
##   Northern USA            0   1
##   Puerto Rico             1   0
##   South Africa            1   0
##   UK                      1   0

ma2 = metabin(cases1, pop1, cases0, pop0, data = BCG,
                    studlab = trialnam, sm = "RR", byvar = USA)
```

```
forest(ma2, bylab = "")
```

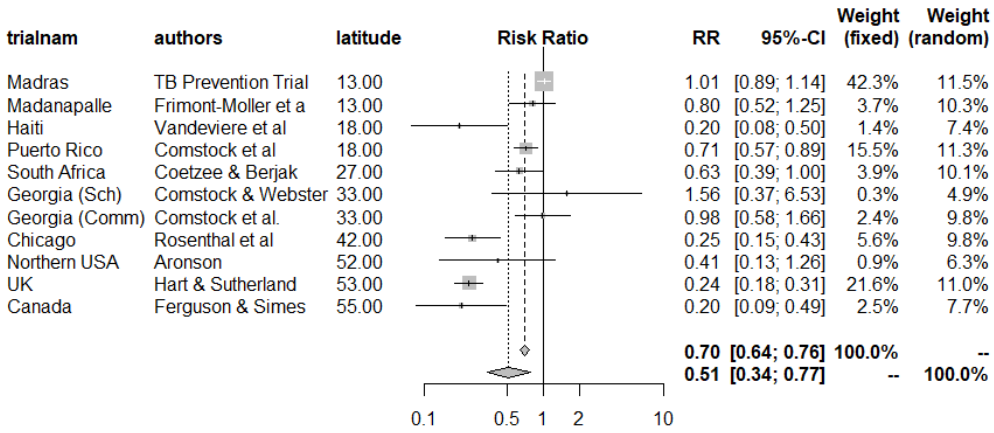


In the light of the various analyses you have performed, what vaccination policy would you recommend to the US Centers for Disease Control?

Answer: *There is clear heterogeneity, even between studies conducted in the USA. The two trials in the south of the USA show no effect, while the two trials in the north showed protection by BCG. The confidence interval for the random-effects estimate is wide. It is hard to decide whether to recommend vaccination or not without a better understanding of what is behind the heterogeneity.*

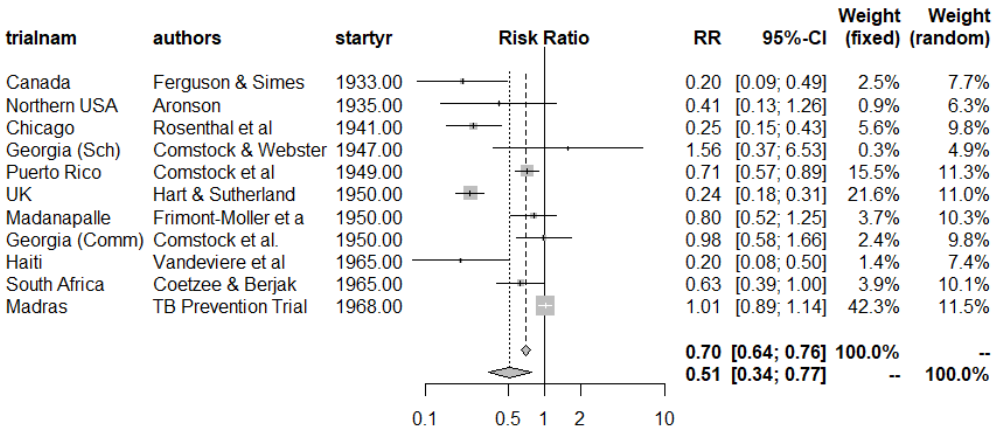
Sort the data by latitude and produce a forest plot of the effect in each study.

```
forest(ma1, leftcols = c("trialnam","authors", "latitude"), just.addcols = "l
eft", sortvar = latitude)
```



Do the same for year of start of study. What is the impression?

```
forest(ma1, leftcols = c("trialnam","authors", "startyr"), just.addcols = "left", sortvar = startyr)
```



Answer: *The visual impression is of a stronger association of risk ratio with latitude than with start year.*

Use the byvar argument of the `metabin` function to compare the results of meta-analyses of the trials that used random allocation (alloc=0) and those that used non-random allocation methods (alloc=1). Does the method of allocation explain much of the heterogeneity?

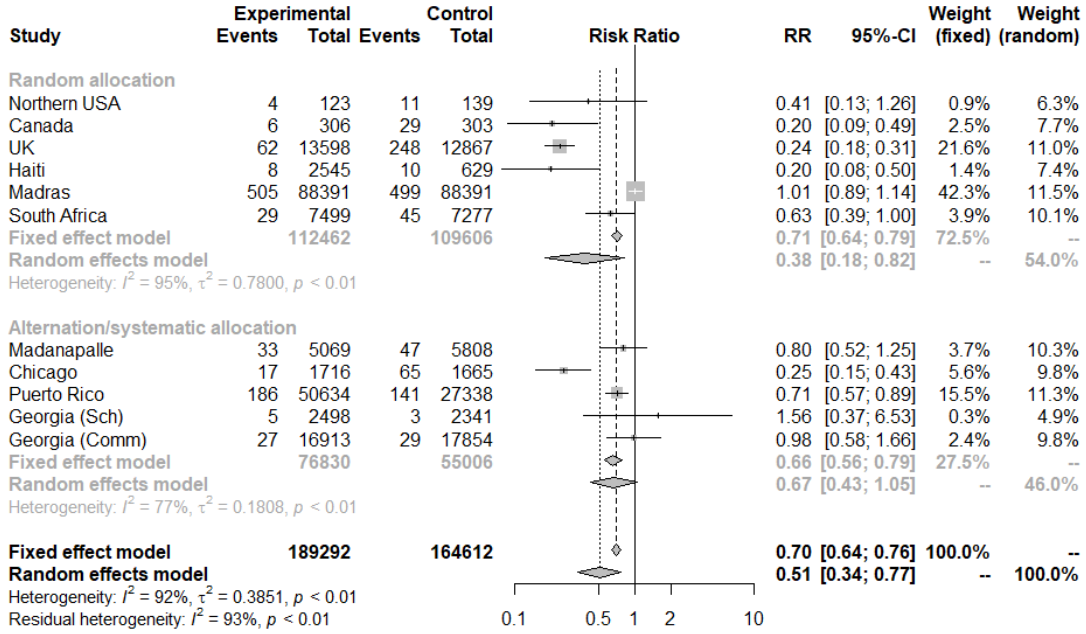
```
table(BCG$trialnam, BCG$alloc)
```

```
##
##                 Alternation/systematic allocation Random allocation
##   Canada                                        0                 1
##   Chicago                                       1                 0
##   Georgia (Comm)                                1                 0
##   Georgia (Sch)                                 1                 0
##   Haiti                                         0                 1
##   Madanapalle                                   1                 0
##   Madras                                        0                 1
##   Northern USA                                  0                 1
##   Puerto Rico                                   1                 0
##   South Africa                                  0                 1
##   UK                                            0                 1
```

```r
ma3 = metabin(cases1, pop1, cases0, pop0, data = BCG,
                    studlab = trialnam, sm = "RR", byvar = alloc)
summary(ma3)
```

```
## Number of studies combined: k = 11
##
##                         RR       95%-CI     z  p-value
## Fixed effect model   0.70 [0.64; 0.76] -8.02 < 0.0001
## Random effects model 0.51 [0.34; 0.77] -3.20   0.0014
##
## Quantifying heterogeneity:
## tau^2 = 0.3851; H = 3.56 [2.87; 4.41]; I^2 = 92.1% [87.9%; 94.9%]
##
## Quantifying residual heterogeneity:
## H = 3.74 [3.00; 4.66]; I^2 = 92.8% [88.9%; 95.4%]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  126.63   10 < 0.0001
##
## Results for subgroups (fixed effect model):
##                                              k   RR       95%-CI      Q
## alloc = Random allocation                    6 0.71 [0.64; 0.79] 108.54
## alloc = Alternation/systematic allocation    5 0.66 [0.56; 0.79]  17.32
##                                             tau^2   I^2
## alloc = Random allocation                  0.7800 95.4%
## alloc = Alternation/systematic allocation 0.1808 76.9%
##
## Test for subgroup differences (fixed effect model):
##                      Q d.f. p-value
## Between groups   0.46    1  0.4972
##
## Results for subgroups (random effects model):
##                                              k   RR       95%-CI      Q
## alloc = Random allocation                    6 0.38 [0.18; 0.82] 108.54
## alloc = Alternation/systematic allocation    5 0.67 [0.43; 1.05]  17.32
##                                             tau^2   I^2
## alloc = Random allocation                  0.7800 95.4%
## alloc = Alternation/systematic allocation 0.1808 76.9%
##
## Test for subgroup differences (random effects model):
##                      Q d.f. p-value
## Between groups   1.54    1  0.2141
##
## Details on meta-analytical method:
## - Mantel-Haenszel method
## - DerSimonian-Laird estimator for tau^2
```

```r
forest(ma3, bylab = "")
```



Answer: *The allocation method appears to explain little of the heterogeneity.*

Meta-regression

Meta-regression can be used to assess whether there is formal statistical evidence that the effect varies with a covariate or between subgroups of studies. The `metareg` function requires an object of class `meta` as argument. Conduct a meta-regression for R object *ma3* and compare the results with the subgroup analysis. Look at the results of *Test of Moderators* in the meta-regression and the *Test for subgroup differences* in the random effects subgroup meta-analysis.

As R object *ma3* was generated using argument byvar, we do not have to specify a covariate in the meta-regression command.

```
mr1 = metareg(ma3)
mr1

##
## Mixed-Effects Model (k = 11; tau^2 estimator: DL)
##
## tau^2 (estimated amount of residual heterogeneity):     0.5607 (SE = 0.422
```

```
0)
## tau (square root of estimated tau^2 value):             0.7488
## I^2 (residual heterogeneity / unaccounted variability): 92.79%
## H^2 (unaccounted variability / sampling variability):   13.86
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 9) = 124.7494, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 1.3737, p-val = 0.2412
##
## Model Results:
##
##                          estimate      se     zval    pval    ci.lb
## intrcpt                   -0.3664  0.3663  -1.0003  0.3172  -1.0843
## .byvarRandom allocation   -0.5828  0.4973  -1.1720  0.2412  -1.5574
##                           ci.ub
## intrcpt                  0.3515
## .byvarRandom allocation  0.3918
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We get the same result explicitly stating the covariate 'alloc' using *ma1*.

```
mr1 = metareg(ma1, alloc)
```

In the output, the coefficient ('estimate') for alloc is the estimate of difference in the log risk ratio between the trials with Random Allocation (alloc=1) and those with non-random allocation methods (alloc=0) (the coefficient labeled 'intrcpt' is the estimate of the log risk ratio for the alloc=0 trials and is of less interest here). The output also gives the standard error of each coefficient, a test of the null hypothesis that it is zero, and a confidence interval. Is there statistical evidence that the results vary according to the method of allocation?

Answer: *The p-value of 0.2412 for "Random Allocation" indicates that there is very little statistical evidence that the log risk ratio varies according to the allocation method in this meta-analysis.*

Note, coefficients can be extracted using R function `coef` and as results for the meta-regression are reported on the log-scale we can use `exp` for back-transformation.

Coefficients of the meta-regression:

```
coef(mr1)
```

```
##                intrcpt allocRandom allocation
##             -0.3664013             -0.5828053
```

Odds ratio for the non-random allocation subgroup:

```
logOR.nra = coef(mr1)[1]

round(exp(logOR.nra), 2)
```

```
## intrcpt
##    0.69
```

Odds ratio for the random allocation subgroup:

```
logOR.ra = sum(coef(mr1))

round(exp(logOR.ra), 2)
```

```
## [1] 0.39
```

Use the `metareg` function to examine the association of the log risk ratio with latitude and with year the trial started. With continuous covariates, such as these, the output can be interpreted as the difference in the log risk ratio associated with a one-unit increase in the covariate.

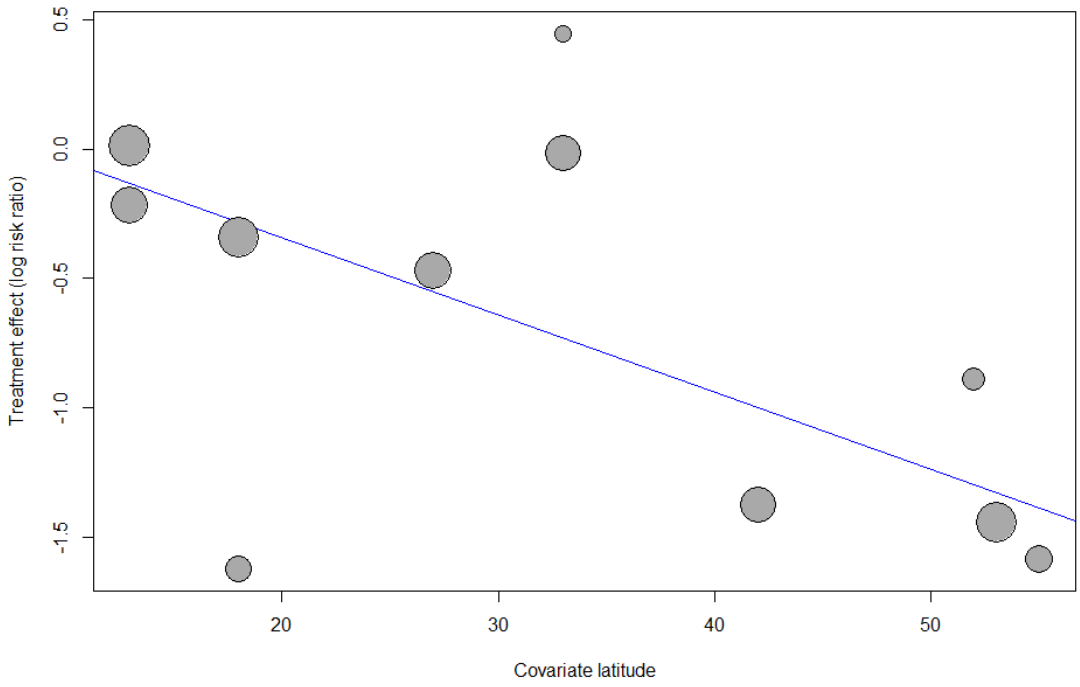
The `bubble` function can be used with a continuous covariate to plot the fitted regression line together with circles representing the estimates from each study. The size of each circle indicates the precision of each estimate (the inverse of its within-study variance). For example, to look at the association with latitude:

```
mr2 = metareg(ma1, latitude)
mr2
```

```
##
## Mixed-Effects Model (k = 11; tau^2 estimator: DL)
##
## tau^2 (estimated amount of residual heterogeneity):     0.0652 (SE = 0.065
4)
## tau (square root of estimated tau^2 value):             0.2553
## I^2 (residual heterogeneity / unaccounted variability): 61.80%
## H^2 (unaccounted variability / sampling variability):   2.62
## R^2 (amount of heterogeneity accounted for):            82.93%
##
## Test for Residual Heterogeneity:
## QE(df = 9) = 23.5614, p-val = 0.0051
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 17.7319, p-val < .0001
```

```
##
## Model Results:
##
##           estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt     0.2539  0.2333   1.0883  0.2765  -0.2034   0.7112
## latitude   -0.0298  0.0071  -4.2109  <.0001  -0.0436  -0.0159  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bubble(mr2, col.line = "blue")
```



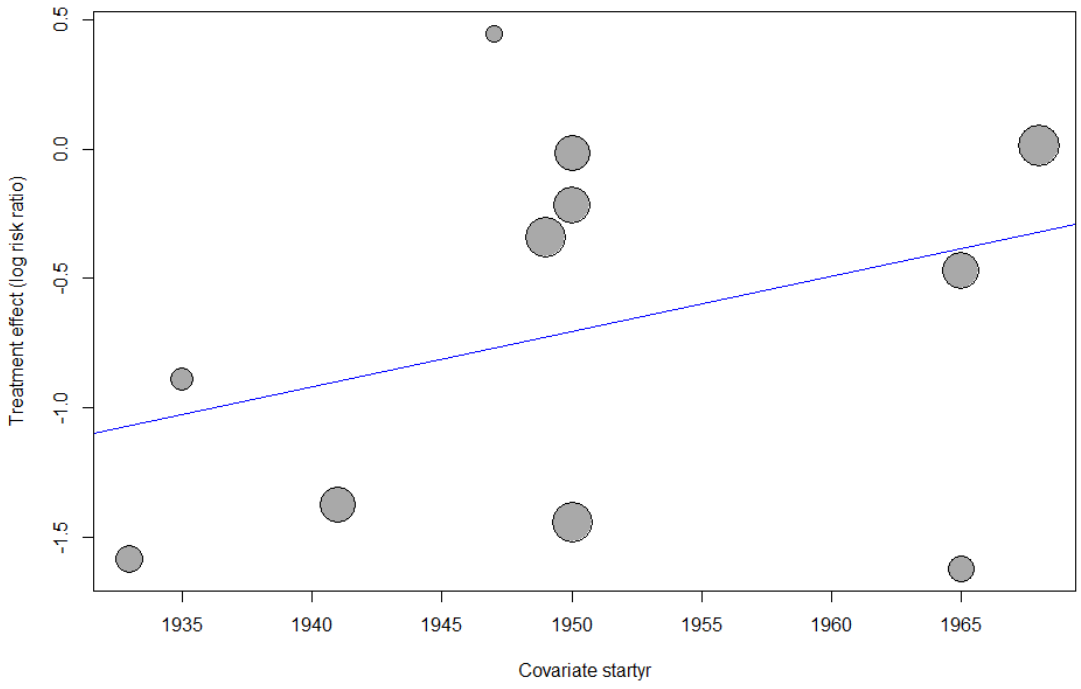
And for the association with the year the trial started:

```
mr3 = metareg(ma1, startyr)
mr3
```

```
##
## Mixed-Effects Model (k = 11; tau^2 estimator: DL)
##
## tau^2 (estimated amount of residual heterogeneity):     0.3228 (SE = 0.226
4)
## tau (square root of estimated tau^2 value):             0.5681
## I^2 (residual heterogeneity / unaccounted variability): 86.84%
## H^2 (unaccounted variability / sampling variability):   7.60
## R^2 (amount of heterogeneity accounted for):            15.46%
##
## Test for Residual Heterogeneity:
```

```
## QE(df = 9) = 68.3730, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 1.4073, p-val = 0.2355
##
## Model Results:
##
##          estimate       se     zval    pval      ci.lb    ci.ub
## intrcpt  -42.4672  35.2308  -1.2054  0.2280  -111.5182  26.5838
## startyr    0.0214   0.0181   1.1863  0.2355    -0.0140   0.0568
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bubble(mr3, col.line = "blue")
```



Comment the extent to which different study characteristics explain the between-trial heterogeneity.

Answer: *Meta-regression confirms the visual impression of a strong association of the effect of BCG with latitude, but not with study year.*

Dataset “magnes”

Dataset magnes.xls gives the effects (from all known trials) of magnesium on the prevention of mortality after acute myocardial infarction.

Open the dataset:

```
Magnesium = read_excel("magnes.xls")
```

Perform meta-analyses

a) on studies performed up to and including 1991

Summarize the study-specific odds ratios for mortality from studies using metabin. We will create an object of class meta called “OR1”. The argument ‘subset’ specifies we only include studies performed up to 1991.

```
OR1 = metabin( deaths1, pop1, deaths0, pop0, data = Magnesium,
                    studlab = trialnam, sm = "OR", subset = year<=1991)
summary(OR1)

## Number of studies combined: k = 13
##
##                         OR       95%-CI     z  p-value
## Fixed effect model   0.38 [0.26; 0.55] -5.13 < 0.0001
## Random effects model 0.41 [0.28; 0.60] -4.53 < 0.0001
##
## Quantifying heterogeneity:
## tau^2 = 0; H = 1.00 [1.00; 1.51]; I^2 = 0.0% [0.0%; 56.4%]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  11.93   12  0.4510
##
## Details on meta-analytical method:
## - Mantel-Haenszel method
## - DerSimonian-Laird estimator for tau^2
## - Continuity correction of 0.5 in studies with zero cell frequencies
```

Fixed effects OR (95% CI):

Random effects OR (95% CI):

b) excluding the ISIS-4 trial

```
OR2 <- update(OR1, subset = trialnam!="ISIS-4")
summary(OR2)

## Number of studies combined: k = 15
##
##                         OR       95%-CI     z  p-value
```

```
## Fixed effect model   0.54 [0.44; 0.68] -5.45 < 0.0001
## Random effects model 0.43 [0.29; 0.63] -4.29 < 0.0001
##
## Quantifying heterogeneity:
## tau^2 = 0.1580; H = 1.23 [1.00; 1.67]; I^2 = 33.8% [0.0%; 64.3%]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  21.15   14  0.0978
##
## Details on meta-analytical method:
## - Mantel-Haenszel method
## - DerSimonian-Laird estimator for tau^2
## - Continuity correction of 0.5 in studies with zero cell frequencies
```

Fixed effects OR (95% CI):

Random effects OR (95% CI):

c) including all trials

```
ORall = metabin(deaths1, pop1, deaths0, pop0, data = Magnesium,
                      studlab = trialnam, sm = "OR")
summary(ORall)
```

```
## Number of studies combined: k = 16
##
##                         OR       95%-CI     z p-value
## Fixed effect model   1.01 [0.95; 1.07]  0.20  0.8414
## Random effects model 0.48 [0.33; 0.71] -3.71  0.0002
##
## Quantifying heterogeneity:
## tau^2 = 0.2244; H = 1.77 [1.37; 2.30]; I^2 = 68.2% [46.5%; 81.1%]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  47.14   15 < 0.0001
##
## Details on meta-analytical method:
## - Mantel-Haenszel method
## - DerSimonian-Laird estimator for tau^2
## - Continuity correction of 0.5 in studies with zero cell frequencies
```

Fixed effects OR (95% CI):

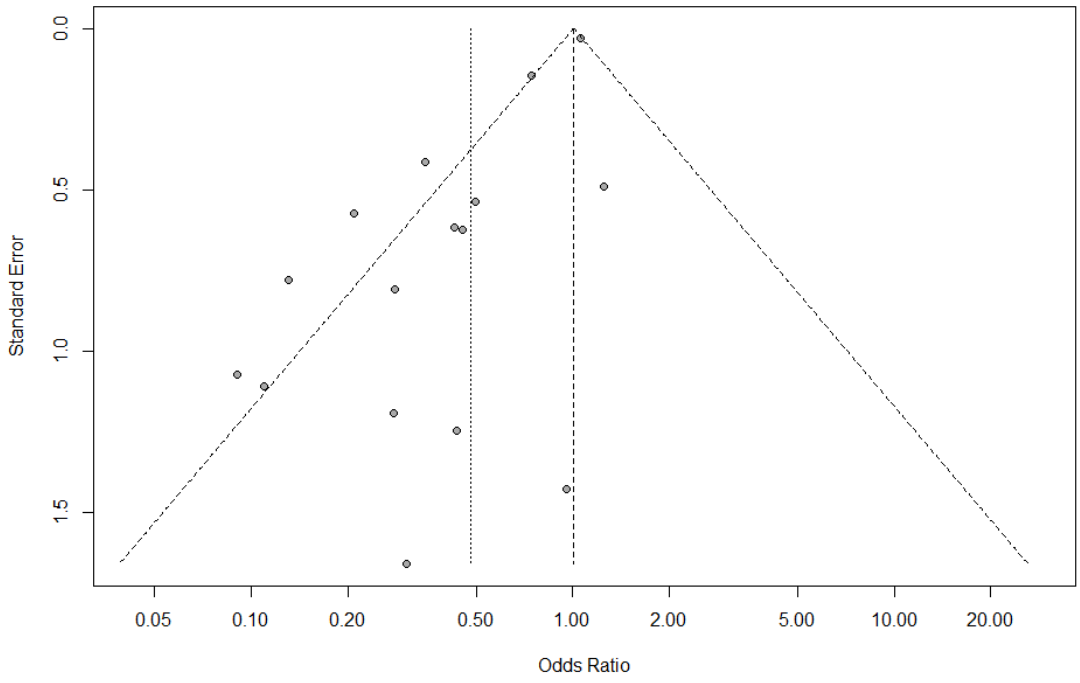
Random effects OR (95% CI):

Answer: *There is a clear protective effect of magnesium from the trials until 1991.*
When considering all trials except for ISIS=4,the fixed and random effects estimates differ considerably, because the larger LIMIT-2 trial showed a much smaller effect.
The results of the large ISIS-4 trial mean there is no effect from the fixed-effect analysis. The fixed and random effects estimates now differ considerably, because the larger LIMIT-2 trial showed a much smaller effect.However because the random-effects weights are very different, there is still a highly statistically significant effect in the random-effects analysis.

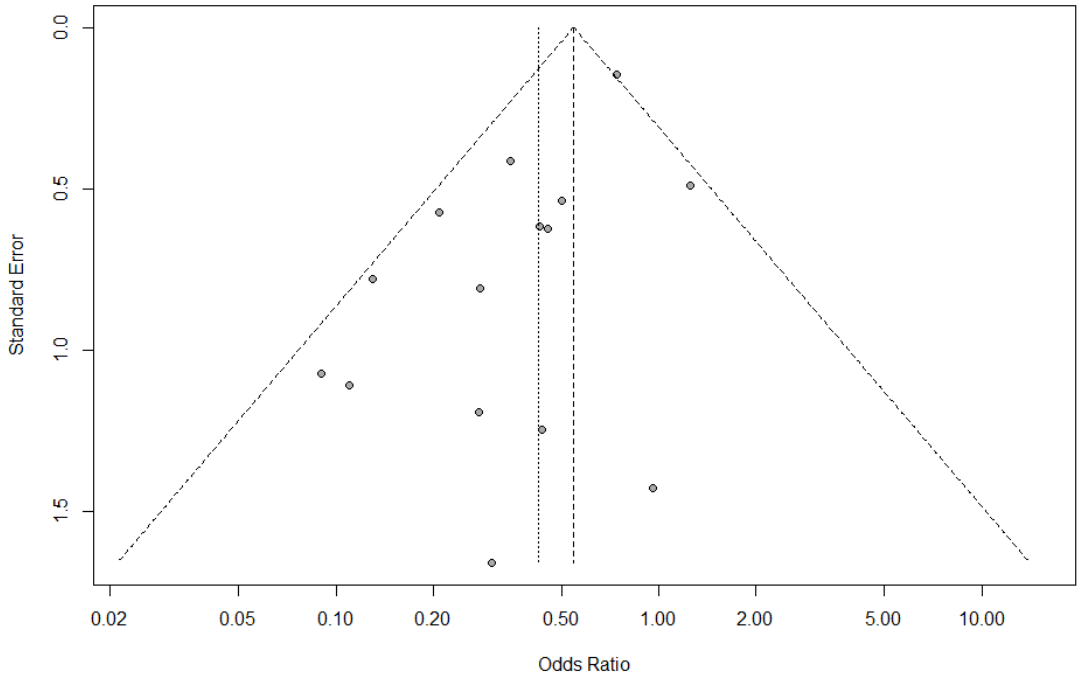
Funnel plots

The function `funnel` produces funnel plots. Examine the funnel plot of the magnesium data, with and without inclusion of the ISIS-4 trial. For example:

```
funnel(ORall)
```




```
funnel(OR2)
```



Function `metabias` produces a test for funnel plot asymmetry. The user needs to specify which test shoulds be conducted and reported in the `method.bias` argument. For example, for the Harbord's score-based test for the meta-analysis including all studies:

```
metabias(ORall, method.bias = "score")
```

```
##
##  Linear regression test of funnel plot asymmetry (efficient score)
##
## data:  ORall
## t = -5.6949, df = 14, p-value = 5.536e-05
## alternative hypothesis: asymmetry in funnel plot
## sample estimates:
##       bias    se.bias      slope
## -1.7525380  0.3077396  0.1029101
```

Using this command, use Harbord's test to examine whether there is evidence of funnel plot asymmetry in the magnesium dataset, both including and excluding the ISIS-4 mega-trial. Comment.

```
metabias(OR2, method.bias = "score")
```

```
##
##  Linear regression test of funnel plot asymmetry (efficient score)
##
```

```
## data:  OR2
## t = -2.7604, df = 13, p-value = 0.01621
## alternative hypothesis: asymmetry in funnel plot
## sample estimates:
##       bias    se.bias      slope
## -1.2070825  0.4372929 -0.1975284
```

Answer: *There is clear evidence of small-study effects, even when the very large ISIS-4 trial is excluded.*