Meta-analysis (day 2)

Advanced modelling with R

Juan R Gonzalez
juanr.gonzalez@isglobal.org

BRGE - Bioinformatics Research Group in Epidemiology
ISGlobal - Barcelona Institute for Global Health
http://brge.isglobal.org

Sub-group analysis

- Between-study heterogeneity is such an important issue in interpreting the results of our meta-analysis (make effect estimate less precise)
- We can explore sources of heterogeneity using influence analyses or detecting outliers.
- Another source of between-study heterogeneity could be that there are slight differences in the study design or intervention components between the studies.
- For example, differences in inclusion or exclusion criteria, or in study design. Many other differences of this sort are possible, and it seems plausible that such study differences may also be associated with differences in the overall effect.
- In **subgroup analyses**, we therefore have a look at different subgroups within the studies of our meta-analysis and try to determine of the differ between these subgroups.

Subgroup analysis

1. **Pooling the effect of each subgroup**. This point it rather straightforward, as the same criteria as the ones for a simple meta-analysis without subgroups apply here.
2. **Comparing the effects of the subgroups**. After we calculated the pooled effect for each subgroup, we can compare the size of the effects of each subgroup. However, to know if this difference is in fact singnificant and/or meaningful, we have to calculate the Standard Error of the differences between subgroup effect sizes, `SEdiff`, to calculate confidence intervals and conduct significance tests. There are two ways to calculate `SEdiff`, and both based on different assumptions.

NOTE: The capabilites of subgroup analyses to detect meaningful differences between studies is often limited. Subgroup analyses also need sufficient power, so it makes no sense to compare two or more subgroups when your entire number of studies in the meta-analysis is smaller than k=10 (Higgins and Thompson 2004).

Subgroup analysis

dat.bcg: Studies on the Effectiveness of the BCG Vaccine Against Tuberculosis

```
library(meta)
data("dat.bcg", package="metafor")
head(dat.bcg)
```

```
  trial               author year tpos  tneg cpos  cneg ablat      alloc
1     1              Aronson 1948    4   119   11   128    44     random
2     2    Ferguson & Simes 1949    6   300   29   274    55     random
3     3     Rosenthal et al 1960    3   228   11   209    42     random
4     4   Hart & Sutherland 1977   62 13536  248 12619    52     random
5     5 Frimodt-Moller et al 1973   33  5036   47  5761    13 alternate
6     6     Stein & Aronson 1953  180  1361  372  1079    44 alternate
```

```
table(dat.bcg$alloc)
```

```
 alternate     random systematic
         2          7          4
```

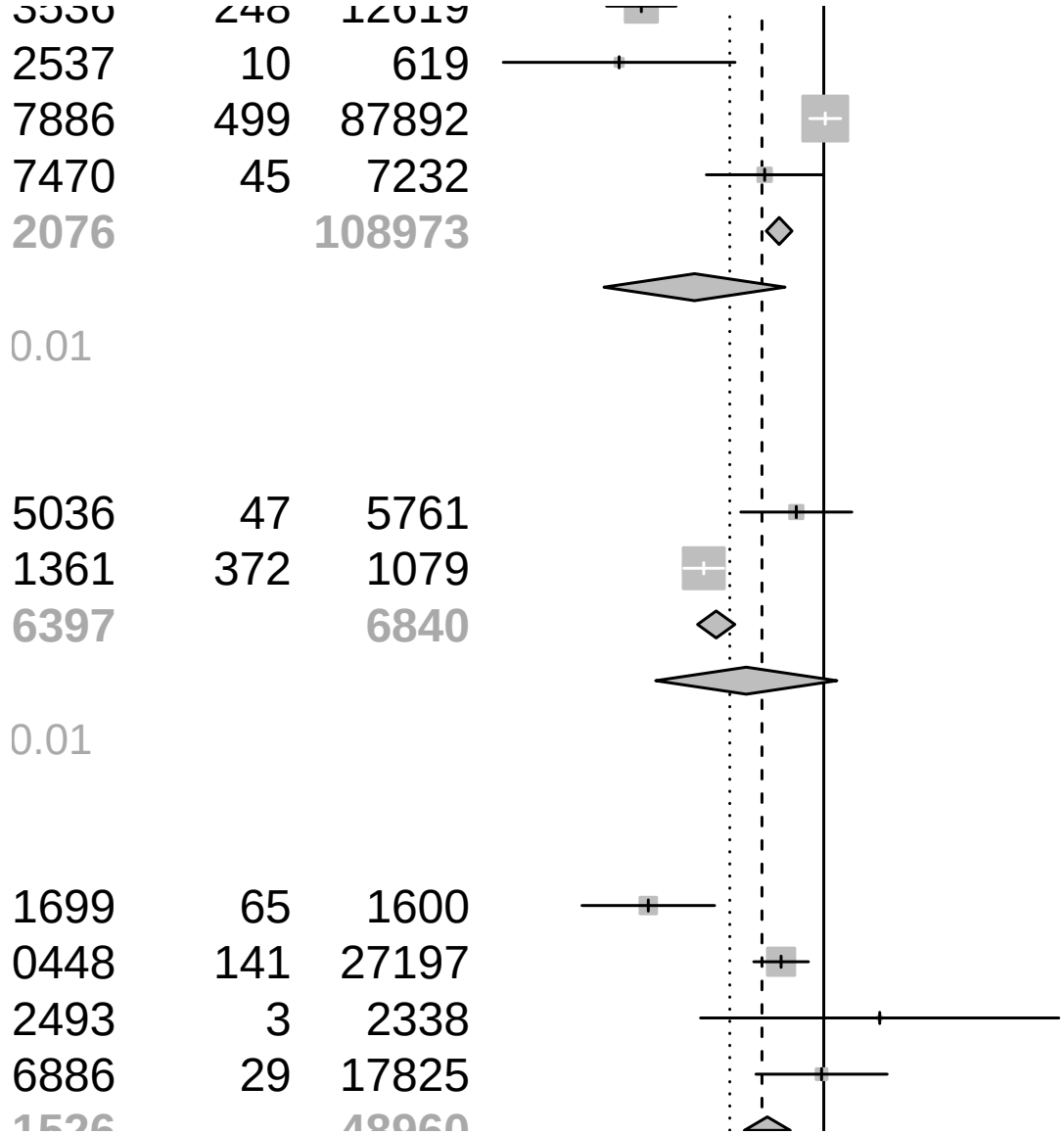
Subgroup analysis

```
res.bcg <- metabin(tpos, tneg, cpos, cneg,
                   data = dat.bcg, byvar= alloc,
                   studlab = paste(author, year, sep=","))
summary(res.bcg)
```

```
Number of studies combined: k = 13

                         RR           95%-CI      z  p-value
Fixed effect model   0.6105 [0.5656; 0.6590] -12.66 < 0.0001
Random effects model 0.4722 [0.3249; 0.6862]  -3.93 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.3623; H = 3.86 [3.20; 4.65]; I^2 = 93.3% [90.2%; 95.4%]

Quantifying residual heterogeneity:
H = 3.76 [3.06; 4.64]; I^2 = 92.9% [89.3%; 95.3%]

Test of heterogeneity:
      Q d.f.  p-value
 178.57   12 < 0.0001

Results for subgroups (fixed effect model):
                     k     RR           95%-CI      Q  tau^2   I^2
alloc = random       7 0.7004 [0.6321; 0.7761] 114.72 0.7958 94.8%
alloc = alternate    2 0.4237 [0.3651; 0.4917]   9.49 0.2453 89.5%
alloc = systematic   4 0.6372 [0.5306; 0.7653]  17.49 0.3038 82.8%

Test for subgroup differences (fixed effect model):
                     Q d.f.  p-value
Between groups   30.14    2 < 0.0001
```

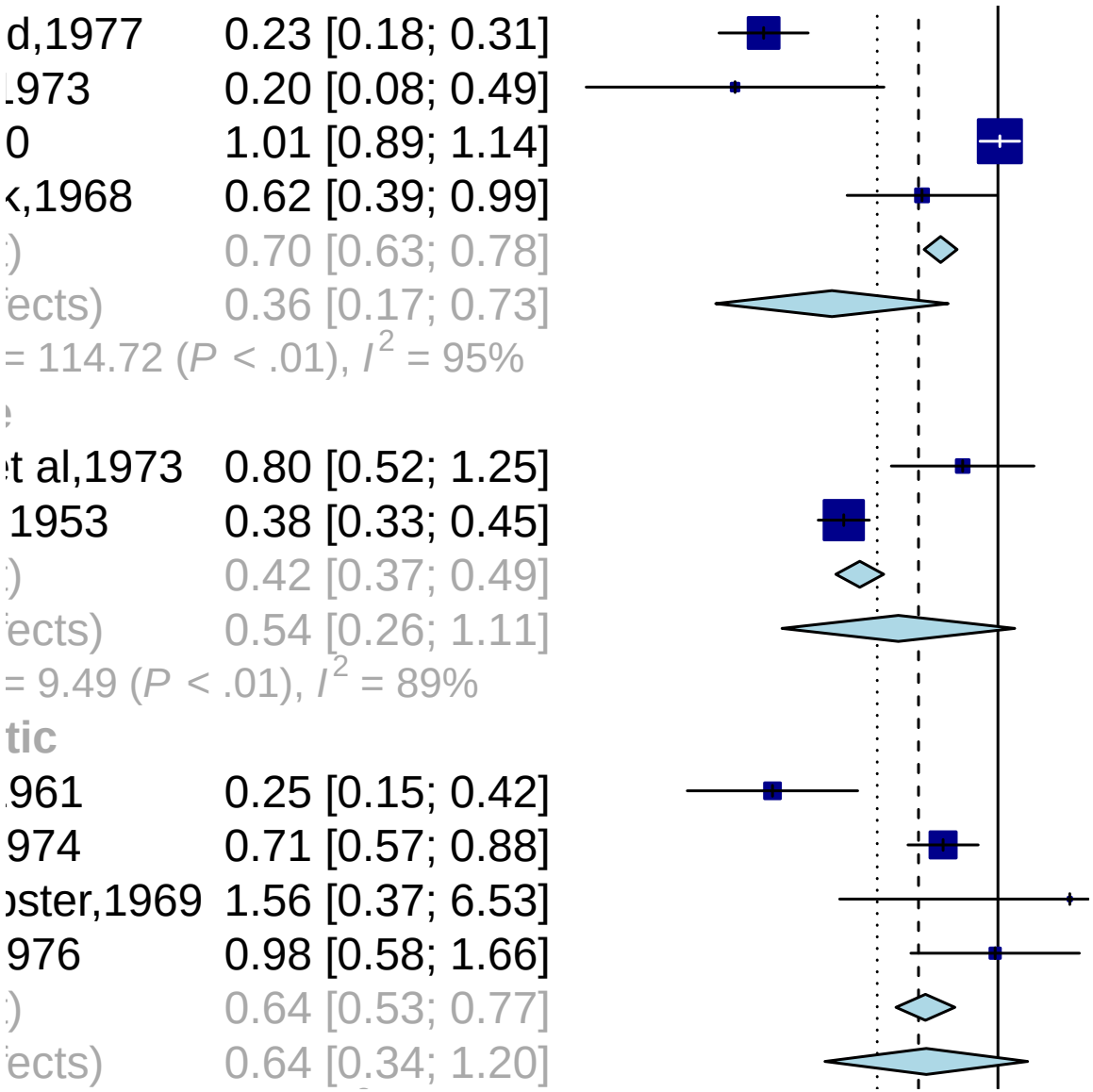
Subgroup analysis

```
forest(res.bcg)
```



Subgroup analysis

```
forest(res.bcg, layout = "JAMA")
```



Meta regression: help to explain heterogeneity

- The inclusion of covariates in the analysis may help to control for heterogeneity
- **Meta-Regression* does not differ much from a** subgroup analysis**.
- Actually, subgroup analyses with more than two groups are nothing more than a meta-regression with categorial covariates.
- Meta-regression does also allow us to use continuous data as covariates and check weather values of this variable are associated with effect size.
- Subgroup analyses make no sense when k$<$10.
- For meta-regression, Borenstein and colleages (2011) recommend that each covariate should at least contain ten studies, although this should not be seen as clear rule.

Meta regression: help to explain heterogeneity

Study-level regression for mean ES

$$Y_i = \boldsymbol{\beta}' \mathbf{x}_i + \theta_i + e_i,$$

$$\theta_i \sim N(0, \tau^2),$$

$$e_i \sim N(0, V_i).$$

$\mathbf{x}_i$ = Study-level covariates

Figure 1:

For meta-regression we compute

Meta regression: help to explain heterogeneity

`R^2 (amount of heterogeneity accounted for):` to see what the variable explains `Test of Moderators:` Global test of the covariate `Model results:` Sinificance of each category

```
res.reg <- metareg(res.bcg, alloc)
summary(res.reg)
```

```
Mixed-Effects Model (k = 13; tau^2 estimator: DL)

  logLik  deviance       AIC       BIC      AICc
-13.1531   38.1464   34.3062   36.5660   39.3062

tau^2 (estimated amount of residual heterogeneity):     0.5959 (SE = 0.4297)
tau (square root of estimated tau^2 value):             0.7719
I^2 (residual heterogeneity / unaccounted variability): 92.88%
H^2 (unaccounted variability / sampling variability):   14.05
R^2 (amount of heterogeneity accounted for):            0.00%

Test for Residual Heterogeneity:
QE(df = 10) = 140.4538, p-val < .0001

Test of Moderators (coefficient(s) 2:3):
QM(df = 2) = 1.3714, p-val = 0.5037

Model Results:

                estimate      se     zval    pval    ci.lb   ci.ub
intrcpt          -0.6018  0.5586  -1.0775  0.2813  -1.6966  0.4929
```

Meta regression: help to explain heterogeneity

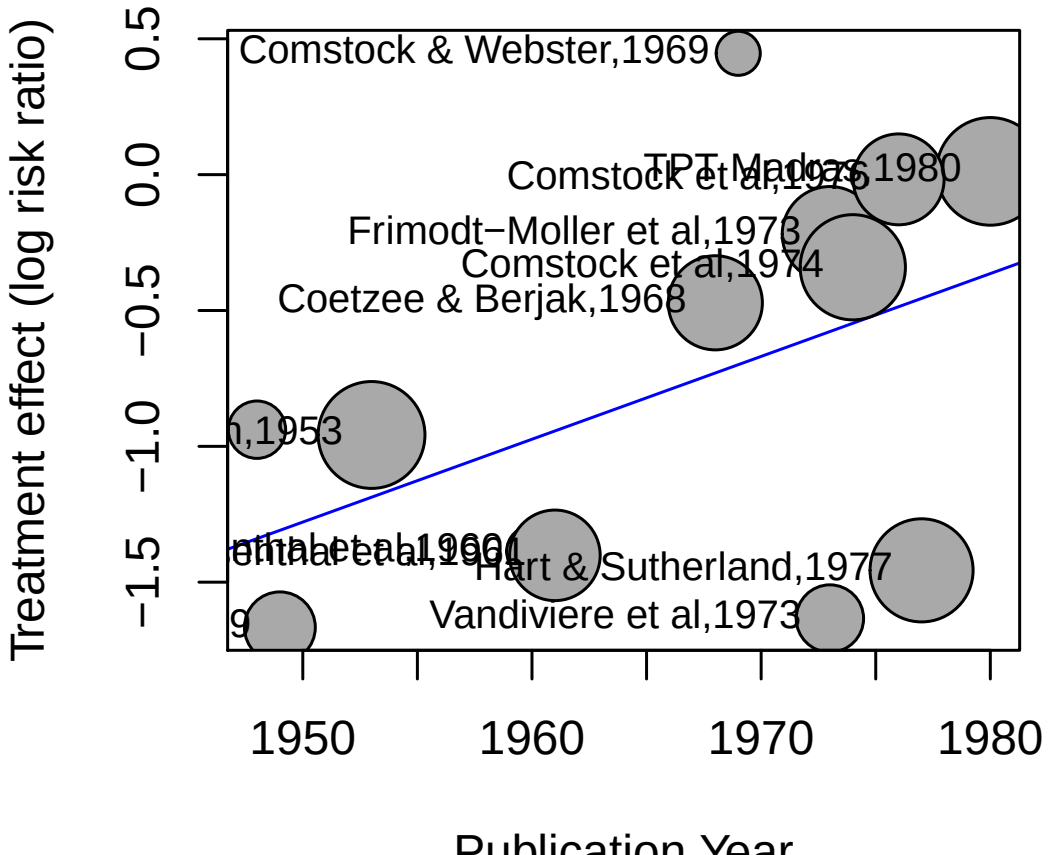
Let's assume we want to check if the *publication year* is associated with effect size.

```
res.reg2 <- metareg(res.bcg, year)
summary(res.reg2)
```

```
Mixed-Effects Model (k = 13; tau^2 estimator: DL)

  logLik  deviance       AIC       BIC      AICc
-11.4306   34.7015   28.8613   30.5561   31.5280

tau^2 (estimated amount of residual heterogeneity):     0.2959 (SE = 0.2181)
tau (square root of estimated tau^2 value):             0.5439
I^2 (residual heterogeneity / unaccounted variability): 88.79%
H^2 (unaccounted variability / sampling variability):   8.92
R^2 (amount of heterogeneity accounted for):            18.29%

Test for Residual Heterogeneity:
QE(df = 11) = 98.1297, p-val < .0001

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 3.2439, p-val = 0.0717

Model Results:

         estimate       se     zval    pval      ci.lb   ci.ub
intrcpt -60.7337  33.3078  -1.8234  0.0682  -126.0158  4.5484  .
year      0.0305   0.0169   1.8011  0.0717    -0.0027  0.0637  .
```

Meta regression: bubble plots

```
bubble(res.reg2,
       xlab = "Publication Year",
       col.line = "blue",
       studlab = TRUE)
```



Exercises

Exercises (**Using R Markdown**)

`dat.hackshaw1998`: Results from 37 studies on the risk of lung cancer from environmental tobacco smoke (ETS) exposure.

These are observational studies, so that, we only have beta effect (i.e. log odds ratio) encoded in the variable `yi` and sampling variance in `vi`.

1. Load the data into R by typing:

```
data(dat.hackshaw1998, package="metafor")
```

2. Run a fixed and random effect meta-analysis
3. Is there heterogeneity?
4. Run a sub-group analysis of studies performed at each country. Is there any difference? And by the type of design?
5. Is there any influence of the year of publication with regard the pooled effect?

Publication Bias

The File-Drawer Problem

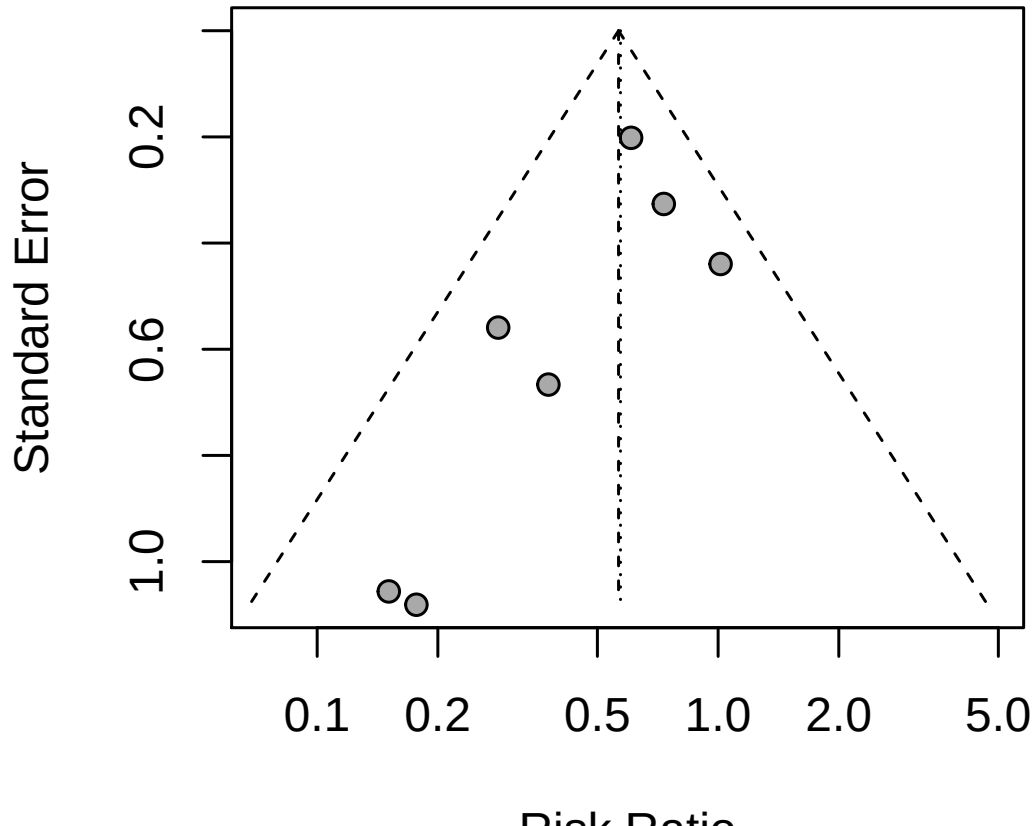
- It is possible that studies showing a significant intervention effect are more often published than studies with null results.
- When a meta-analysis is based only on studies reported in the literature, null studies relegated to the file-drawer could bias the summary intervention effect in the direction of efficacy.

Detecting publication bias: The Funnel plot

- A funnel plot is a scatter plot of the intervention effect estimates against a measure of study precision.
- Asymmetry (gaps) in the funnel may be indicative of publication bias.
- Some authors argue that judging asymmetry is too subjective to be useful.
- Spurious asymmetry can result from heterogeneity or when ESs are correlated with precision.

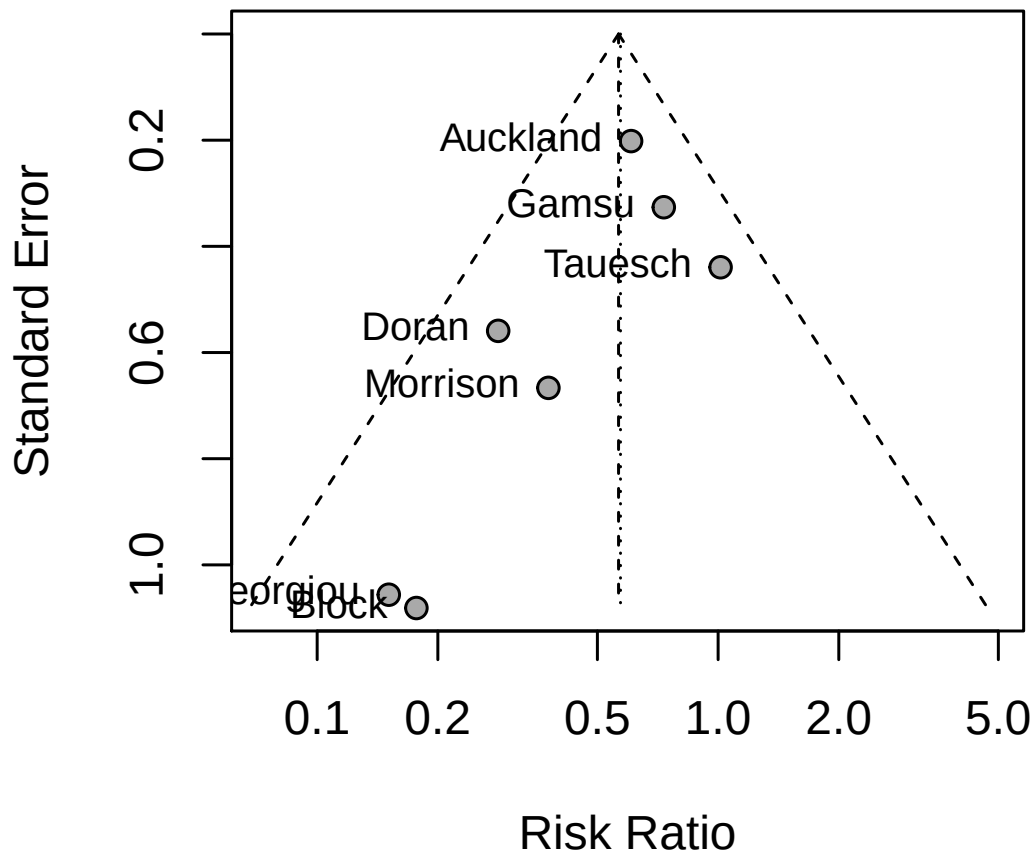
Funnel plots

```
data(cochrane, package = "rmeta")
res <- metabin(ev.trt, n.trt, ev.ctrl, n.ctrl,
               data=cochrane, studlab = name)
funnel(res)
```



Funnel plots

```
funnel(res, studlab = TRUE)
```

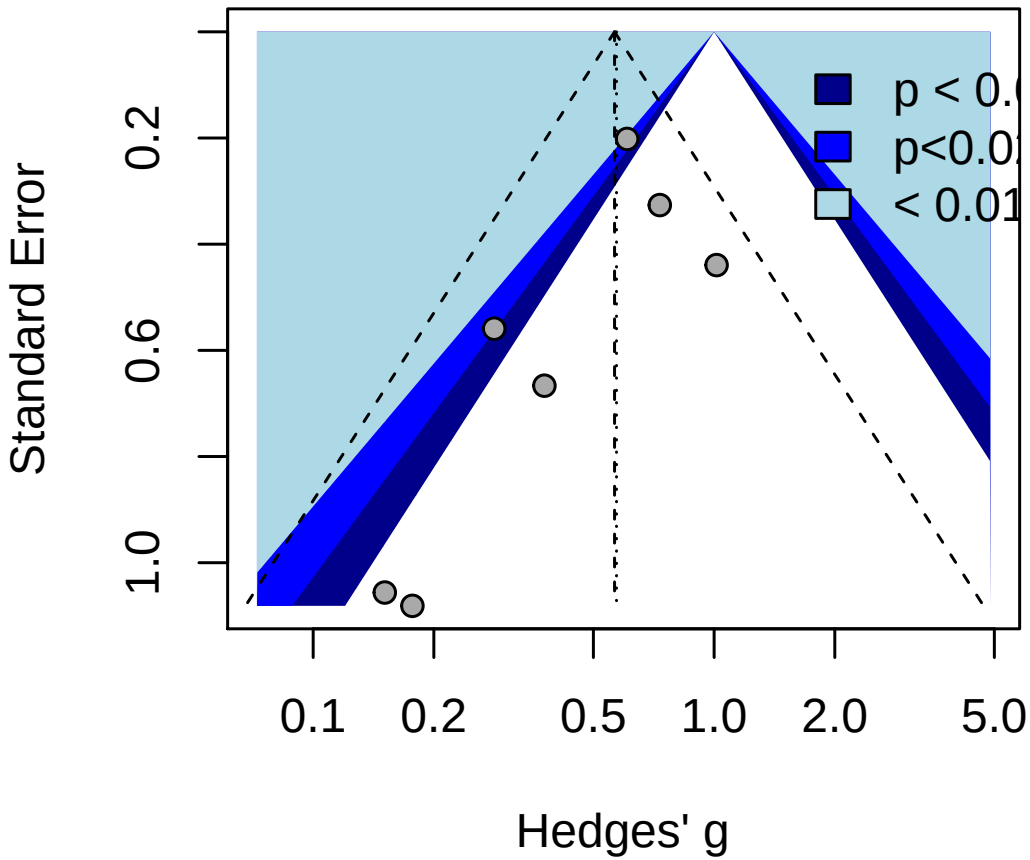


Funnel plots

An even better way to inspect the funnel plot is through contour-enhanced funnel plots, which help to distinguish publication bias from other forms of asymmetry (Peters et al. 2008). Contour-enhanced funnels include colors signifying the significance level into which the effects size of each study falls. We can plot such funnels using this code:

```
funnel(res, xlab="Hedges' g",
       contour = c(.95,.975,.99),
       col.contour=c("darkblue","blue","lightblue"))+
legend(1.4, 0, c("p < 0.05", "p<0.025", "< 0.01"),bty = "n",
       fill=c("darkblue","blue","lightblue"))
```

Funnel plots



Asymmetry: Egger's test

- Egger et al. (1997) proposed a test for asymmetry of the funnel plot. This is a test for the Y intercept $= 0$ from a linear regression of normalized effect estimate (estimate divided by its standard error) against precision (reciprocal of the standard error of the estimate).
- Harbord (2005) developed a test that maintains the power of the Egger test whilst reducing the false positive rate, which is a problem with the Egger test when there are large treatment effects, few events per trial or all trials are of similar sizes. The original Egger test should be used instead of the Harbord method if there is a large imbalance between the sizes of treatment and control groups – the same is true for the Peto odds ratio, to which this test is mathematically related.

Asymmetry: Egger's test

```
metabias(res, k.min=7) # default is 10
```

```
    Linear regression test of funnel plot asymmetry

data:  res
t = -1.8872, df = 5, p-value = 0.1178
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias    se.bias      slope
-1.2623815  0.6689173 -0.1084619
```

Asymmetry: Egger's test

- Thrombolytic Therapy after Acute Myocardial Infarction
- H_0: No asymmetry

```
data("Olkin95", package="meta")
res.olkin <- metabin(event.e, n.e, event.c, n.c,
                     data=Olkin95)
metabias(res.olkin, method.bias = "linreg") # Egger
```

```
    Linear regression test of funnel plot asymmetry

data:  res.olkin
t = -1.7704, df = 68, p-value = 0.08115
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias    se.bias      slope
-0.2891100  0.1633045 -0.2089214
```

Asymmetry: Egger's test

```
metabias(res.olkin, method.bias = "score") # Harbord
```

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

data:  res.olkin
t = -1.7333, df = 68, p-value = 0.08758
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias    se.bias      slope
-0.3044535  0.1756508 -0.2435568
```

Publication bias

- Judging asymmetry in the funnel plot can be difficult. So you will usually want to consider some additional ways of assessing the threat of publication bias.
- Sensitivity Analyses:
  - Trim-and-Fill
  - Fail Safe N

Trim-and-fill method

- The trim-and-fill method estimates the number of missing NULL studies from the meta-analysis.
- The function `trimfill` augments the observed data and returns the fitted object with the missing studies included.
- These points can be added to the funnel plot.

Trim-and-fill method

The trim-and-fill procedure includes the following five steps (Schwarzer, Carpenter, and Rücker 2015):

- Estimating the number of studies in the outlying (right) part of the funnel plot.
- Removing (trimming) these effect sizes and pooling the results with the remaining effect sizes.
- This pooled effect is then taken as the center of all effect sizes.
- For each trimmed/removed study, a additional study is imputed, mirroring the effect of the study on the left side of the funnel plot.
- Pooling the results with the imputed studies and the trimmed studies included.

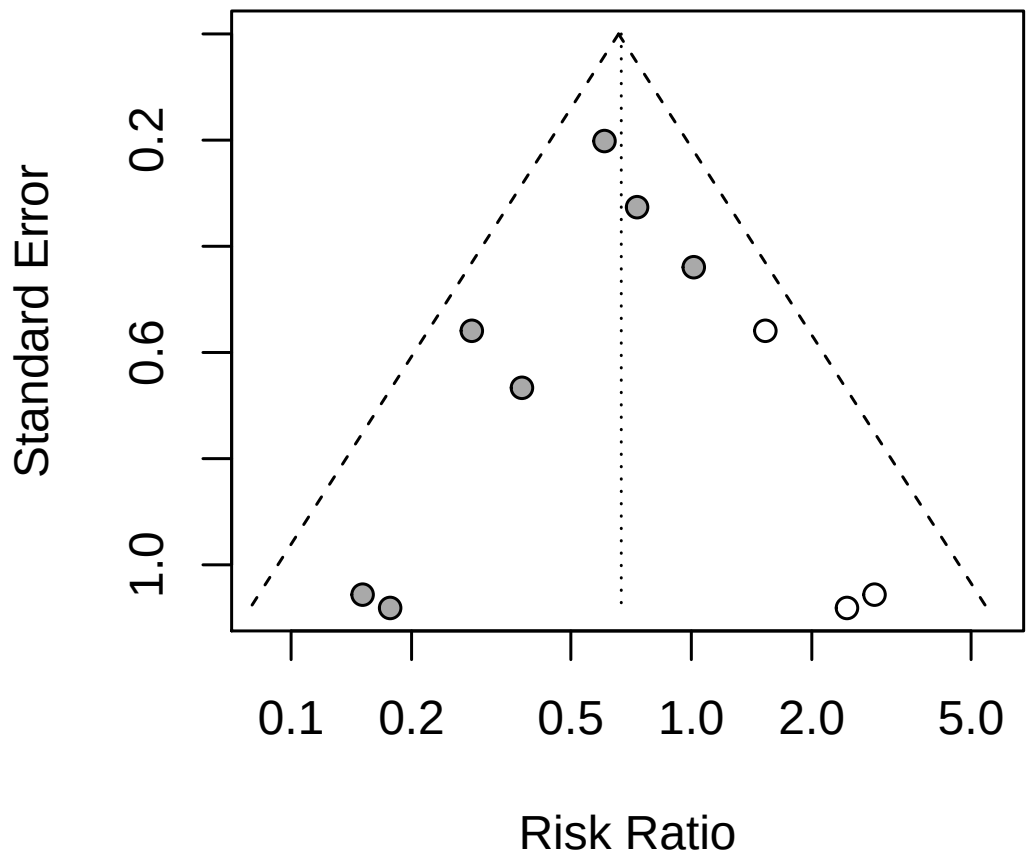
Trim-and-fill method

```
res.trim <- trimfill(res)
res.trim
```

```
                          RR            95%-CI %W(random)
Auckland              0.6068 [0.4086;  0.9011]       26.7
Block                 0.1768 [0.0212;  1.4719]        3.3
Doran                 0.2828 [0.0945;  0.8463]        9.9
Gamsu                 0.7321 [0.3862;  1.3878]       18.9
Morrison              0.3774 [0.1022;  1.3938]        7.5
Papageorgiou          0.1509 [0.0190;  1.1959]        3.4
Tauesch               1.0143 [0.4287;  2.3997]       13.6
Filled: Doran         1.5309 [0.5116;  4.5806]        9.9
Filled: Block         2.4488 [0.2942; 20.3853]        3.3
Filled: Papageorgiou  2.8692 [0.3620; 22.7378]        3.4

Number of studies combined: k = 10 (with 3 added studies)

                          RR           95%-CI     z p-value
Random effects model 0.6683 [0.4463; 1.0008] -1.96  0.0505

Quantifying heterogeneity:
tau^2 = 0.1181; H = 1.22 [1.00; 1.76]; I^2 = 32.5% [0.0%; 67.8%]

Test of heterogeneity:
     Q d.f. p-value
 13.34    9  0.1480

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
- Trim-and-fill method to adjust for funnel plot asymmetry
```

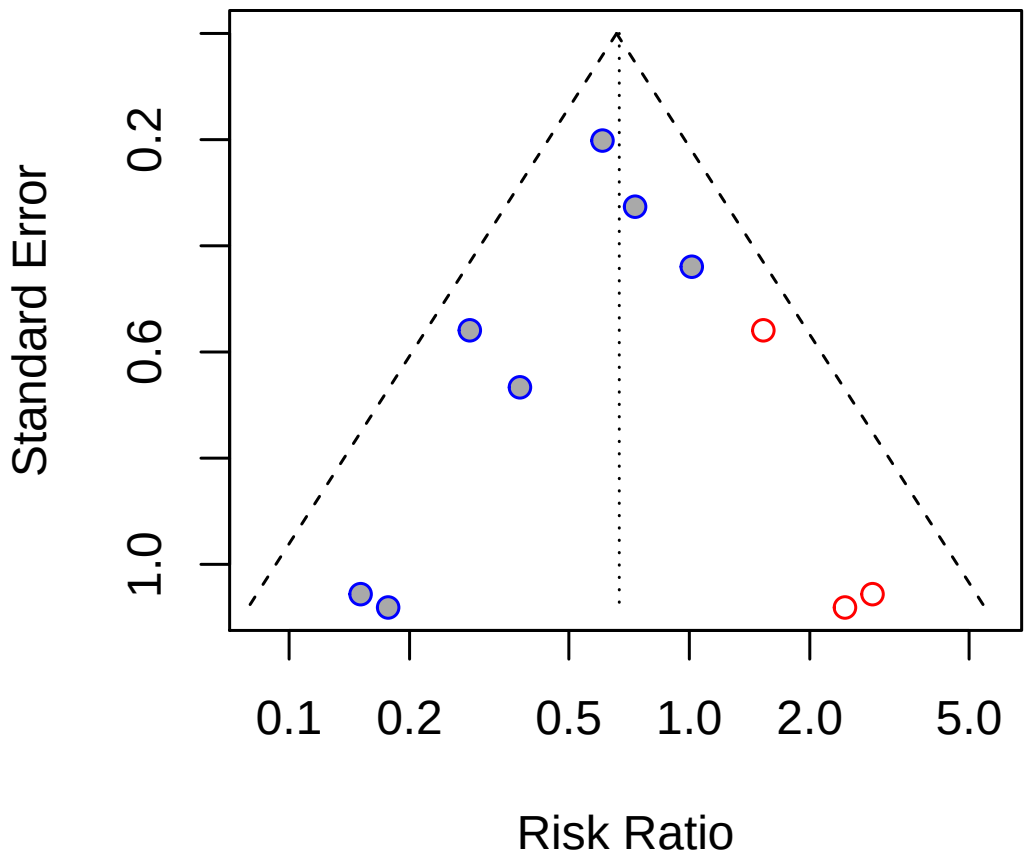
Trim-and-fill method

```
funnel(res.trim)
```



Trim-and-fill method

```
funnel(res.trim, col=ifelse(res.trim$trimfill,
                            "red", "blue"))
```



Trim-and-fill method

The new resulting estimates are:

```
summary(res)
```

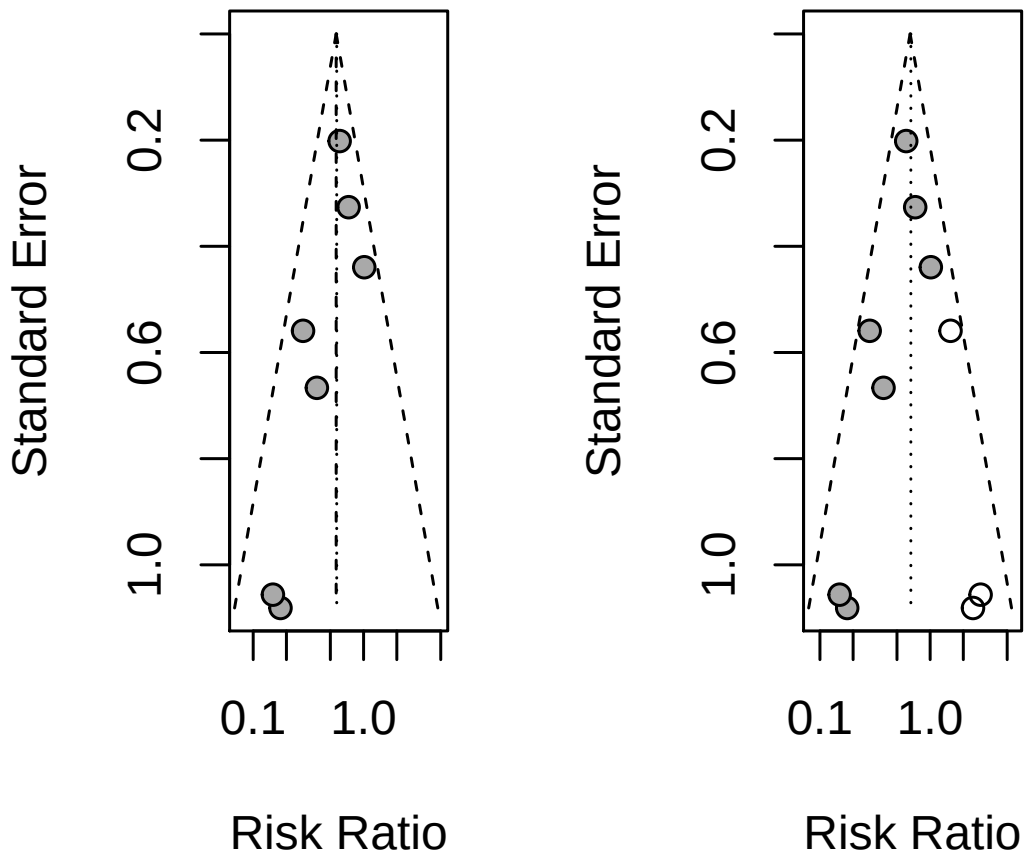
```
Number of studies combined: k = 7

                         RR            95%-CI     z  p-value
Fixed effect model   0.5646 [0.4254; 0.7493] -3.96 < 0.0001
Random effects model 0.5710 [0.4008; 0.8135] -3.10   0.0019

Quantifying heterogeneity:
tau^2 = 0.0376; H = 1.09 [1.00; 1.58]; I^2 = 16.0% [0.0%; 59.9%]

Test of heterogeneity:
    Q d.f. p-value
 7.15    6  0.3076

Details on meta-analytical method:
- Mantel-Haenszel method
- DerSimonian-Laird estimator for tau^2
```

```
summary(res.trim)
```

```
Number of studies combined: k = 10 (with 3 added studies)

                         RR            95%-CI     z p-value
Random effects model 0.6683 [0.4463; 1.0008] -1.96  0.0505

Quantifying heterogeneity:
```

Trim-and-fill method

```
par(mfrow=c(1,2))
funnel(res)
funnel(res.trim)
```



Fail-Safe N

- Rosenthal method (sometimes called a *file drawer analysis*)
- Is the number of NULL studies that have to be added to reduce the significance of the meta-analysis to (usually 0.05)
- This technique is not widely used in meta-analysis
- It is availalbe in `metafor` package

Exercises

Exercises

`dat.hackshaw1998`: Results from 37 studies on the risk of lung cancer from environmental tobacco smoke (ETS) exposure.

These are observational studies, so that, we only have beta effect (i.e. log odds ratio) encoded in the variable `yi` and sampling variance in `vi`.

1. Load the data into R by typing:

```
data(dat.hackshaw1998, package="metafor")
```

2. Assess whether there is publication bias in this meta-analysis.

P-curve

- Recent research has shown that the assumptions of the small-effect study methods (traditional) may be inaccurate in many cases. The Duval & Tweedie trim-and-fill procedure in particular has been shown to be prone to providing inaccurate effect size estimates (Simonsohn, Nelson, and Simmons 2014).
- P-curve Analysis has been proposed as an alternative way to assess publication bias and estimate the true effect behind our collected data.
- P-Curve assumes that publication bias is not primarily generated because researchers do not publish non-significant results, but because the "play" around with their data (e.g., selectively removing outliers, choosing different outcomes, controlling for different variables) until a non-significant finding becomes significant. This (bad) practice is called p-hacking, and has been shown to be extremely frequent among researchers (Head et al. 2015).

P-curve

It has been shown that P-Curve's effect estimate are not robust when the heterogeneity of a meta-analyis is high ($I^2 > 50\%$). Van Aert et al. (Aert, Wicherts, and Assen 2016) propose not to determine the 'true' effect using P-Curve when heterogeneity is high (defined as $I^2 > 50\%$).

http://p-curve.com/guide.pdf

Session info

```
sessionInfo()
```

```
R version 3.5.0 (2018-04-23)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 17134)

Matrix products: default

locale:
[1] LC_COLLATE=Spanish_Spain.1252  LC_CTYPE=Spanish_Spain.1252
[3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C
[5] LC_TIME=Spanish_Spain.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] RCurl_1.95-4.11 bitops_1.0-6    meta_4.9-4

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.1       codetools_0.2-15 digest_0.6.15    grid_3.5.0
 [5] magrittr_1.5     evaluate_0.13    stringi_1.2.2    rmarkdown_1.12
 [9] tools_3.5.0      stringr_1.3.1    xfun_0.5         yaml_2.2.0
[13] compiler_3.5.0   htmltools_0.3.6  knitr_1.22
```