Workshop on
Statistical Meta-Analysis

Guido Knapp and Bimal Sinha

Emails: guido.knapp@tu-dortmund.de, sinha@umbc.edu

September 6, 2019

Department of Mathematics
Clarkson University
Potsdam NY 13699 , USA

Topics of the Course

- Introduction and review of basic statistical meta-analysis methods
- More on basic statistical meta-analysis methods
- Meta-analysis of binary data
- Publication bias and small study effects
- Meta-regression and subgroup analysis

Main References for SMA and R packages for SMA

- Hartung, J., Knapp, G., Sinha, B.K. (2008). Statistical Meta-Analysis with Applications. Wiley & Sons, New York.
- **R package** `meta`
  Schwarzer, G., Carpenter, J.R., Rücker, G. (2015). Meta-Analysis with R. Springer, Cham.
  See also: http://meta-analysis-with-r.org/
- **R package** `metafor`
  Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–48.
  See also: http://www.metafor-project.org/

Two recently published review articles:

- Veroniki, A.A., Jackson, D, Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuß, O., Higgins, J.P.T., Langan, D., Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods* **7**, 55–79.
- Veroniki, A.A., Jackson, D., Bender, R., Kuss, O., Langan, D., Higgins, J.P.T., Knapp, G., Salanti, G. (2019). Methods to calculate uncertainty in the overall effect size from a random-effects meta-analysis. *Research Synthesis Methods* **10**, 23–43.

Introduction to Meta-Analysis

- *Meta-analysis*, a term coined by Glass (1976), is intended to provide
  the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.
- Pearson (1904) used data from five independent samples, and computed a pooled estimate of correlation between mortality and inoculation in order to evaluate the efficacy of the vaccine for enteric fever.
- Birge (1932) combined estimates across experiments at different laboratories to establish reference values for some fundamental constants in physics.
- Early works of Cochran (1937), Yates and Cochran (1938), Tippett (1931) and Fisher (1932) dealt with combining information across experiments in the agricultural sciences.

Areas of application

- Social sciences:
  Reliability and validity studies, teacher expectancies studies, ...
- Life sciences:
  Effectiveness of drugs, second-hand smoking, ...
- Archaeology, astronomy, chemistry,
  engineering, environmental sciences, geosciences,
  mass communication, military operations analysis, official statistics,
  physics, psychology, sports science...
- Interlaboratory trials, metrology

Introduction

Four important stages of research synthesis:

1. **Problem formulation** stage
2. **Data collection** stage
3. **Data evaluation** stage
4. **Data analysis and interpretation** stage

Data analysis: Results from **published** studies

Motivation for statistical meta-analysis (see Cochrane Handbook)

1. to increase power
2. to improve precision
3. to answer questions not posed by the individual studies
4. to settle controversies or generate new hypotheses

Examples

Four experiments about the percentage of albumin in plasma protein in human subjects

Experiment	n_i	Mean	Variance
A	12	62.3	12.986
B	15	60.3	7.840
C	7	59.5	33.433
D	16	61.5	18.513

Examples

Studies of the relationship between an observation measure of teacher indirectness and student achievement

Study	No. of teachers	Correlation coefficient r
1	15	-0.073
2	16	0.308
3	15	0.481
4	16	0.428
5	15	0.180
6	17	0.290
7	15	0.400

Examples

Number of patients and mortality rate from all causes, for six trials comparing the use of aspirin and placebo by patients following a heart attack

	Aspirin		Placebo	
Study	No. of Pat.	Mort. Rate (%)	No. of Pat.	Mort. Rate (%)
UK-1	615	7.97	624	10.74
CDPA	758	5.80	771	8.30
GAMS	317	8.52	309	10.36
UK-2	832	12.26	850	14.82
PARIS	810	10.49	406	12.81
AMIS	2267	10.85	2257	9.70

Examples

Number of cases of lung cancer in women who did not actively smoke cigarettes and estimated relative risk of lung cancer in relation exposure to environmental tobacco smoke

Study	No. of Cases	Estimated RR (95% CI)	Study	No. of Cases	Estimated RR (95% CI)
1	94	1.52 (0.88 - 2.63)	11	24	0.79 (0.25 - 2.45)
2	19	1.52 (0.39 - 5.99)	12	86	1.55 (0.90 - 2.67)
3	41	0.81 (0.34 - 1.90)	13	199	1.65 (1.16 - 2.35)
4	84	0.75 (0.43 - 1.30)	14	60	2.01 (1.09 - 3.71)
5	22	2.07 (0.82 - 5.25)	15	32	1.03 (0.41 - 2.55)
6	246	1.19 (0.82 - 1.73)	16	67	1.28 (0.76 - 2.15)
7	134	1.31 (0.87 - 1.98)	17	34	1.26 (0.57 - 2.82)
8	54	2.16 (1.08 - 4.29)	18	62	2.13 (1.19 - 3.83)
9	20	2.34 (0.81 - 6.75)	19	28	1.41 (0.54 - 3.67)
10	22	2.55 (0.74 - 8.78)			

Examples

13 trials on the prevention of tuberculosis using BCG vaccination

Trial	Vaccinated		Not vaccinated		Latitude
	Disease	No disease	Disease	No Disease	
1	4	119	11	128	44
2	6	300	29	274	55
3	3	228	11	209	42
4	62	13536	248	12619	52
5	33	5036	47	5761	13
6	180	1361	372	1079	44
7	8	2537	10	619	19
8	505	87886	499	87892	13
9	29	7470	45	7232	27
10	17	1699	65	1600	42
11	186	50448	414	27197	18
12	5	2493	3	2338	33
13	27	16886	29	17825	33

Data

Possible data available from the published study:

- Raw data (very rare) or summary statistics (quite often)
- P-value only
- Effect size estimate plus standard error
  (Data required for meta-analysis)
- Effect size estimate plus confidence interval
- Effect size estimate plus P-value
- Effect size estimate plus sample size

Examples

- Validity studies: Correlation between student ratings of instructor with student achievement

Study	Sample	n	r
Bolton et al. (1979)	General psychology	10	0.68

  Data are sufficient for meta-analysis of correlation coefficients.

- Studies of the effects of teacher expectancy on pupil IQ

Study	d	SE(d)
Rosenthal et al. (1974)	0.03	0.125

  Cohen's d is an estimate of the standardized mean difference.
  Data are in the required form for meta-analysis.

Examples

- Second-hand smoking

Study	RR (95% CI)
Akiba, Kato, and Blot (1986)	$1.52\ (0.88 - 2.63)$

- How can we extract the standard error of this estimate? Note that the interval is not symmetric about its estimate.
- For relative risk and odds ratio, the meta-analysis is done on the logarithmic scale, that is, first log-transform the estimate and the bounds of the confidence interval.
- Assume a Wald confidence interval for $\log(\text{RR})$, that is,

$$\log(\text{RR}) \pm \text{SE}(\log(\text{RR}))\ z_{1-\alpha/2}$$

  with z_γ the γ-quantile of the standard normal distribution.
- Note the length of the interval is

$$2\ \text{SE}(\log(\text{RR}))\ z_{1-\alpha/2}.$$

- The standard error of the log RR can then be calculated by

$$\text{SE}(\log(\text{RR})) = \frac{\text{length CI}}{2\ z_{1-\alpha/2}}$$

Generic Meta-Analysis Models

- **Fixed-effect model**: common effect size
- **Random-effects model**: average effect size

Let us consider k independent studies, then we have for $i = 1, 2, \ldots, k$

θ_i — true effect size in the ith study

$\hat{\theta}_i$ — estimated effect size in the ith study

$\sigma^2(\hat{\theta}_i)$ — true variance of $\hat{\theta}_i$, which may depend on θ_i

$\hat{\sigma}^2(\hat{\theta}_i)(= \hat{\sigma}_i^2)$ — estimate of $\sigma^2(\hat{\theta}_i)$

In the ith study:

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta_i, \sigma^2(\hat{\theta}_i)\right), \qquad \hat{\sigma}^2(\hat{\theta}_i) \text{ given}$$

Justification for normality assumption:

- Central limit theorem
- Asymptotic normality of maximum-likelihood estimator

Use linear mixed-effects model theory for the statistical analysis.

Fixed-Effect Model

Homogeneity assumption: $\theta_1 = \theta_2 = \cdots = \theta_k =: \theta$ (common effect size)

Fixed-effect model is given by

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta, \hat{\sigma}_i^2\right), \ i = 1, 2, \ldots, k.$$

(Conditional) ML estimate and (conditional) BLUE for θ is

$$\hat{\theta}_{FE} = \frac{\sum_{i=1}^k v_i \, \hat{\theta}_i}{\sum_{i=1}^k v_i}, \quad v_i = \frac{1}{\hat{\sigma}_i^2}, \ i = 1, 2, \ldots, k,$$

with

$$\text{Var}\left(\hat{\theta}_{FE}\right) = \frac{1}{\sum_{i=1}^k v_i}.$$

Approximate (standard) $(1 - \alpha)$-confidence interval for θ

$$\hat{\theta}_{FE} \pm \sqrt{\text{Var}\left(\hat{\theta}_{FE}\right)} \, z_{1-\alpha/2}$$

and z_γ is the γ-quantile of the standard normal distribution.

Fixed-Effect Model

Homogeneity test problem:

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \qquad \text{versus} \qquad H_1 : \exists (i,j)\ \theta_i \neq \theta_j, i \neq j.$$

Cochran's Q:

$$Q_C = \sum_{i=1}^{k} v_i \left(\hat{\theta}_i - \hat{\theta}_{FE} \right)^2$$

is approximately χ^2-distributed with $k-1$ degrees of freedom.

Reject H_0 at (approximate) level α, if $Q_C > \chi^2_{k-1;1-\alpha}$.

Choice of $\alpha = 0.1$ or $\alpha = 0.2$ not unusual.

No Homogeneity

What shall we do, if the homogeneity assumption is not fulfilled?

One possibility: Assume a normal-normal hierarchical model.

Observational model:

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta_i, \hat{\sigma}^2(\hat{\theta}_i)\right), \ i = 1, 2, \ldots, k,$$

Structural model:

$$\theta_i \sim \mathcal{N}\left(\theta, \tau^2\right), \ i = 1, 2, \ldots, k,$$

Marginally,

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta, \tau^2 + \hat{\sigma}^2(\hat{\theta}_i)\right), \ i = 1, 2, \ldots, k,$$

The main parameter of interest, θ, is then called average effect size. The nuisance parameter τ^2 stands for the between-study variability, also called heterogeneity parameter.

Random-Effects Model

Random-effects model:

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta, \tau^2 + \hat{\sigma}^2(\hat{\theta}_i)\right), \; i = 1, 2, \ldots, k.$$

(Conditional) ML estimator and (conditional) BLUE for θ for known τ^2

$$\tilde{\theta}_{RE} = \frac{\sum_{i=1}^k w_i \, \hat{\theta}_i}{\sum_{i=1}^k w_i}, \quad w_i = \frac{1}{\tau^2 + \hat{\sigma}^2(\hat{\theta}_i)}, \; i = 1, 2, \ldots, k,$$

How do we estimate τ^2?

- DerSimonian-Laird estimator

$$\hat{\tau}^2_{DSL} = \frac{Q_C - (k-1)}{\sum_{i=1}^k v_i - \sum_{i=1}^k v_i^2 / \sum_{i=1}^k v_i}$$

  with $v_i = 1/\hat{\sigma}_i^2$ and Q_C is Cochran's homogeneity test statistic.
- The estimator $\hat{\tau}^2_{DSL}$ may yield a negative estimate for the heterogeneity parameter, and hence the truncated version $\max\{0, \hat{\tau}^2_{DSL}\}$ is used.

Random-Effects Model

- Quantifying the amount of heterogeneity:

$$I^2 = \frac{Q_C - (k-1)}{Q_C}\ 100\%$$

  where Q_C is Cochran's homogeneity statistic.
- Benchmarks for I^2: 25% small, 50% moderate, 75% high levels of heterogeneity
- Borenstein, M. et al. (2017): Basic meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*. DOI: 10.1002/jrsm.1230

By plugging in an estimate of τ^2 in $\tilde{\theta}_{RE}$, we obtain the estimated average effect size

$$\hat{\theta}_{RE} = \frac{\sum_{i=1}^k \hat{w}_i\ \hat{\theta}_i}{\sum_{i=1}^k \hat{w}_i}, \quad \hat{w}_i = \frac{1}{\hat{\tau}^2 + \hat{\sigma}^2(\hat{\theta}_i)},\ i = 1, 2, \ldots, k,$$

with 'classical' variance estimate

$$\widehat{\mathsf{Var}}\left(\hat{\theta}_{RE}\right) = \frac{1}{\sum_{i=1}^k \hat{w}_i}.$$

Approximate (standard) $(1-\alpha)$-confidence interval for θ

$$\hat{\theta}_{RE} \pm \sqrt{\widehat{\mathsf{Var}}\left(\hat{\theta}_{RE}\right)}\ z_{1-\alpha/2}.$$

Example

Results of eight randomized controlled trials comparing the effectiveness of amlodipine and a placebo on work capacity

Protocol	Amlodipine 10 mg (E)			Placebo (C)		
	n_{Ei}	$\bar{y}_{Ei}$	s^2_{Ei}	n_{Ci}	$\bar{y}_{Ci}$	s^2_{Ci}
154	46	0.2316	0.2254	48	-0.0027	0.0007
156	30	0.2811	0.1441	26	0.0270	0.1139
157	75	0.1894	0.1981	72	0.0443	0.4972
162A	12	0.0930	0.1389	12	0.2277	0.0488
163	32	0.1622	0.0961	34	0.0056	0.0955
166	31	0.1837	0.1246	31	0.0943	0.1734
303A	27	0.6612	0.7060	27	-0.0057	0.9891
306	46	0.1366	0.1211	47	-0.0057	0.1291

Example

Let μ_E be the expected value in the amlodipine group and μ_C in the control group. We are interested in $\delta = \mu_E - \mu_C$.
In each study, an estimator of δ_i is given by the difference of means, that is,

$$D_i = \bar{X}_{Ei} - \bar{X}_{Ci},$$

The variance of D_i can be estimated by

$$\widehat{\text{Var}(D_i)} = \frac{S_{Ei}^2}{n_{Ei}} + \frac{S_{Ci}^2}{n_{Ci}}$$

with S_{Ei}^2 und S_{Ci}^2 denoting the sample variances in the respective groups and n_{Ei} and n_{Ci} the respective sample sizes.

Given values of D_i and $\widehat{\text{Var}}(D_i)$, we can easily use the function `metagen` in the R Paket *meta* to perform a meta-analysis.

General call of the function:

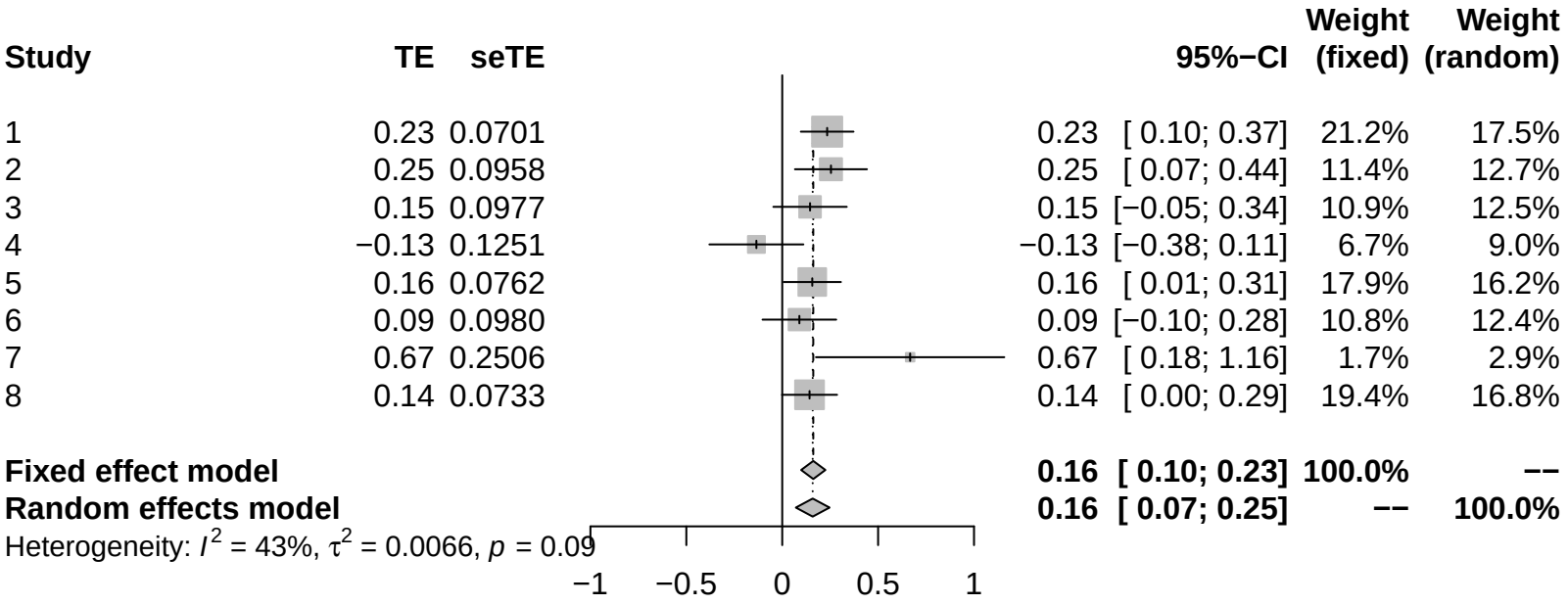
```
metagen(TE, seTE, sm=""),
```

with `TE` the vector of effect sizes, `seTE` the **vector of standard errors**, and `sm=""` a character string indicating underlying summary measure, e.g.,`"MD"` for the difference of means. The DerSimonian-Laird estimator is used as default for estimating the heterogeneity parameter.

Example

```
                     95%-CI %W(fixed) %W(random)
1  0.2343 [ 0.0969; 0.3717]      21.2       17.5
2  0.2541 [ 0.0663; 0.4419]      11.4       12.7
3  0.1451 [-0.0464; 0.3366]      10.9       12.5
4 -0.1347 [-0.3798; 0.1104]       6.7        9.0
5  0.1566 [ 0.0072; 0.3060]      17.9       16.2
6  0.0894 [-0.1028; 0.2816]      10.8       12.4
7  0.6669 [ 0.1758; 1.1580]       1.7        2.9
8  0.1423 [-0.0015; 0.2861]      19.4       16.8

                                        95%-CI    z  p-value
Fixed effect model   0.1619 [0.0986; 0.2252] 5.01 < 0.0001
Random effects model 0.1589 [0.0710; 0.2467] 3.54   0.0004

Quantifying heterogeneity:
tau^2 = 0.0066; H = 1.33 [1.00; 2.00]; I^2 = 43.2% [0.0%; 74.9%]
Test of heterogeneity:
     Q d.f. p-value
 12.33    7  0.0902
Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
```

Forest Plot




```
forest(meta-object)
```

Random-Effects Model

Estimators of the heterogeneity parameter τ^2 available in R packages `meta` and `metafor`:

- DerSimonian-Laird
- Paule-Mandel
- Maximum-likelihood
- Restricted maximum-likelihood
- Hunter-Schmidt
- Sidik-Jonkman (note this estimator is always positive)
- Hedges
- Empirical Bayes

Random-Effects Model

- The solution of

$$Q(\tau^2) = \sum_{i=1}^{k} w_i \left[\hat{\theta}_i - \tilde{\theta}_{RE}\right]^2 \stackrel{!}{=} k-1, \; w_i = (\tau^2 + \hat{\sigma}^2(\hat{\theta}_i))^{-1}$$

  say $\hat{\tau}^2_{PM}$, is called the Paule-Mandel estimator for τ^2.
  The solution is unique and exists provided that $Q(0) > k-1$.
  If $Q(0) < k-1$, the Paule-Mandel estimator is set to zero.
- Paule-Mandel estimator is identical to the Empirical Bayes estimator.

Which heterogeneity estimator should we use?

- Veroniki, A.A. et al. (2016) recommend to use Paule-Mandel and restricted maximum-likelihood estimator of τ^2.
- Does the choice of the heterogeneity estimate affect the statistical inference on the fixed effect(s)?

Random-Effects Model

Recall: Approximate (standard) $(1-\alpha)$-confidence interval for θ

$$\hat{\theta}_{RE} \pm \sqrt{\widehat{\mathsf{Var}}\left(\hat{\theta}_{RE}\right)}\, z_{1-\alpha/2}$$

with

$$\widehat{\mathsf{Var}}\left(\hat{\theta}_{RE}\right) = \frac{1}{\sum_{i=1}^{k} \hat{w}_i}$$

Hartung-Knapp-Sidik-Jonkman-interval (Hartung, Knapp (2001, Statist. Med.), Sidik, Jonkman (2002, Statist. Med.):

$$\hat{\theta}_{RE} \pm \sqrt{\widehat{\mathsf{Var}}_m\left(\hat{\theta}_{RE}\right)}\, t_{k-1;1-\alpha/2}$$

with

$$\widehat{\mathsf{Var}}_m\left(\hat{\theta}_{RE}\right) = \frac{1}{\sum_{i=1}^{k} \hat{w}_i} \frac{1}{k-1} \sum_{i=1}^{k} \hat{w}_i \left(\hat{\theta}_i - \hat{\theta}_{RE}\right)^2$$

and $t_{\nu,\gamma}$ the γ-quantile of the t-distribution with ν degrees of freedom.

Some Remarks

Two competing confidence intervals with lengths

$$\mathsf{L}_{\mathsf{Stand}} = 2 \sqrt{\frac{1}{\sum_{i=1}^{k} \hat{w}_i}}\, z_{1-\alpha/2}$$

and

$$\mathsf{L}_{\mathsf{HKSJ}} = 2 \sqrt{\frac{1}{\sum_{i=1}^{k} \hat{w}_i} \frac{1}{k-1} \sum_{i=1}^{k} \hat{w}_i \left(\hat{\theta}_i - \hat{\theta}_{RE}\right)^2}\, t_{k-1;1-\alpha/2}$$

- Often: $\mathsf{L}_{\mathsf{HKSJ}} > \mathsf{L}_{\mathsf{Stand}}$
- But possible: $\mathsf{L}_{\mathsf{HKSJ}} < \mathsf{L}_{\mathsf{Stand}}$, if the number of studies is reasonably large and the study results are quite homogeneous (fixed-effect model)
- Ad-hoc modification: Use

$$\max\left\{1, \frac{1}{k-1} \sum_{i=1}^{k} \hat{w}_i \left(\hat{\theta}_i - \hat{\theta}_{RE}\right)^2\right\}$$

  in HKSJ to ensure $\mathsf{L}_{\mathsf{HKSJ}} \geq \mathsf{L}_{\mathsf{Stand}}$.
- The ad-hoc modification is already applied when using the Paule-Mandel estimator for estimating τ^2. Then

$$\frac{1}{k-1} \sum_{i=1}^{k} \hat{w}_i \left(\hat{\theta}_i - \hat{\theta}_{RE}\right)^2 = 1.$$

Comments

We use the meta-analysis model

$$\hat{\theta}_i \sim \mathcal{N}\left(\theta, \tau^2 + \hat{\sigma}^2(\hat{\theta}_i)\right), \; i = 1, 2, \ldots, k,$$

assuming $\tau^2 = 0$ (fixed effect) or $\tau^2 > 0$ (random effects)
and the following assumptions and interpretations (depending on the effect size of interest):

- the analysis is conditionally on the observed $\hat{\sigma}^2(\hat{\theta}_i)$;
- the statistical uncertainty of $\hat{\sigma}^2(\hat{\theta}_i)$ is ignored;
- a possible correlation between $\hat{\theta}_i$ and $\hat{\sigma}^2(\hat{\theta}_i)$ is also ignored;
- estimator $\hat{\theta}_i$ do not have to be unbiased;
- unknown parameters are θ and τ^2.

Hartung-Knapp-Sidik-Jonkman-Interval in R package `meta`

```
metagen(TE, seTE,sm=" ", hakn=TRUE)
```

R Package `meta`

Function	Comments
metabin	Meta-analysis of binary outcome data
metacont	Meta-analysis of continuous outcome data
metagen	Generic inverse variance meta-analysis
metacor	Meta-analysis of correlations
metamean	Meta-analysis of single means
metaprop	Meta-analysis of single proportions
metarate	Meta-analysis of single incidence rates
metareg	Meta-regression
metacr	Meta-analysis of outcome data from Cochrane review
metacum	Cumulative meta-analysis
metainf	Influence analysis in meta-analysis using leave-one-out method
forest	Forest plot
funnel	Plot to assess funnel plot asymmetry
metabias	Test for funnel plot asymmetry
...	

R Package `metafor`

Function	Comments
escalc	Effect size calculation (compute outcomes) plus sampling variances (square of standard errors)
rma.uni	Meta-analysis via linear (mixed-effects) models
rma.mh	Meta-analysis via the Mantel-Haenszel method (for binary outcome)
rma.peto	Meta-analysis via Peto's method (for odds ratio)
rma.glmm	Meta-analysis via generalized linear (mixed effects) models
rma.mv	Meta-analysis via multivariate/multilevel linear (mixed effects) models
hc	Meta-analysis based on the method by Henmi and Copas
forest	Forest plots
funnel	Funnels plots
...	

R function *escalc*

A lot of outcome measures can be calculated, for instance, when combining comparative trials (treatment vs. control)

- dichotomous variables: log relative risk, log odds ratio, risk difference
- event counts: log incidence rate ratio, incidence rate difference
- quantitative variables: (raw) mean difference, standardized mean difference

or measures for variable association

- correlation coefficients (quantitative), Φ coefficient (qualitative)

See the documentation of the escalc function for further outcome measures.

Effect Sizes based on Means

- Denote the population means of the two groups (experimental and control) by μ_1 and μ_2, and their variances by σ_1^2 and σ_2^2, respectively.
- Standardized difference between μ_1 and μ_2

$$\theta = \frac{\mu_1 - \mu_2}{\sigma},$$

where σ denotes either the standard deviation σ_2 of the population control group, or an average population standard deviation (namely, an average of σ_1 and σ_2).

- Suppose we have a random sample of size n_1 from the first population with the sample mean $\bar{X}_1$ and sample variance S_1^2, and also a random sample of size n_2 from the second population with the sample mean $\bar{X}_2$ and sample variance S_2^2.
- Cohen's d:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S},$$

where

$$S^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2}.$$

Effect Sizes based on Means

- Hedges's g:

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S^*},$$

where

$$S^{*2} = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}.$$

- It can be shown that

$$\begin{aligned} \mathsf{E}(g) &\approx \theta + \frac{3\,\theta}{4(n_1 + n_2) - 9} \\ \mathsf{Var}(g) &\approx \frac{n_1 + n_2}{n_1\, n_2} + \frac{\theta^2}{2(n_1 + n_2 - 3.94)} \end{aligned}$$

- In case $\sigma_1^2 = \sigma_2^2$ and under the assumption of normality of the data, it holds that $\sqrt{\tilde{n}}\; g$ follows a non-central t-distribution with non-centrality parameter $\sqrt{\tilde{n}}\; \theta$ and $(n_1 + n_2 - 2)$ degrees of freedom, $\tilde{n} = n_1\, n_2/(n_1 + n_2)$.

Effect Sizes based on Means

- With $N = n_1 + n_2$, the mean and variance of Hedges's g are given by

$$\begin{aligned} \mathsf{E}(g) &= \sqrt{\frac{N-2}{2}}\,\frac{\Gamma\left(\frac{N-3}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)}\,\theta \\ \mathsf{Var}(g) &= \frac{N-2}{N-4}\left(\frac{1}{\tilde{n}} + \theta^2\right) - \theta^2\,\frac{N-2}{2}\,\frac{\left(\Gamma\left(\frac{N-3}{2}\right)\right)^2}{\left(\Gamma\left(\frac{N-2}{2}\right)\right)^2} \end{aligned}$$

  and $\Gamma(\cdot)$ denotes the gamma function.

- Approximately *unbiased* estimator g^* of the standardized mean difference is given as

$$g^* = \left(1 - \frac{3}{4(n_1+n_2)-9}\right)\,g$$

- Estimate of the variance

$$\widehat{\mathsf{Var}}(g^*) = \frac{n_1+n_2}{n_1\,n_2} + \frac{g^2}{2\,(n_1+n_2-2)}$$

- Large-sample confidence interval

$$g^* \pm \sqrt{\widehat{\mathsf{Var}}(g)}\;z_{1-\alpha/2}$$

Effect Sizes based on Means

Typical data situation for meta-analysis

Studies of two anaesthetic agents relating to recovery time

Study	Total sample size	Standardized mean difference (g)	Unbiased standardized mean difference (g^*)
1	76	0.72	0.71
2	6167	0.06	0.06
3	355	0.59	0.59
4	1050	0.43	0.43
5	136	0.27	0.27
6	2925	0.89	0.89
7	45222	0.35	0.35

Example

Results of eight randomized controlled trials comparing the effectiveness of amlodipine and a placebo on work capacity

Protocol	Amlodipine 10 mg (E)			Placebo (C)		
	n_{Ei}	$\bar{y}_{Ei}$	s^2_{Ei}	n_{Ci}	$\bar{y}_{Ci}$	s^2_{Ci}
154	46	0.2316	0.2254	48	-0.0027	0.0007
156	30	0.2811	0.1441	26	0.0270	0.1139
157	75	0.1894	0.1981	72	0.0443	0.4972
162A	12	0.0930	0.1389	12	0.2277	0.0488
163	32	0.1622	0.0961	34	0.0056	0.0955
166	31	0.1837	0.1246	31	0.0943	0.1734
303A	27	0.6612	0.7060	27	-0.0057	0.9891
306	46	0.1366	0.1211	47	-0.0057	0.1291

Example: R Code

```
# Data in the first group (sample size, mean, and variance)
n1   <- c(    46,     30,     75,     12,     32,     31,     27,     46)
x1   <- c(0.2316, 0.2811, 0.1894, 0.0930, 0.1622, 0.1837, 0.6612, 0.1366)
var1 <- c(0.2254, 0.1441, 0.1981, 0.1389, 0.0961, 0.1246, 0.7060, 0.1211)
#
# Data in the second group (sample size, mean, and variance)
n2   <- c(     48,     26,     72,     12,     34,     31,      27,      47)
x2   <- c(-0.0027, 0.0270, 0.0443, 0.2277, 0.0056, 0.0943, -0.0057, -0.0057)
var2 <- c( 0.0007, 0.1139, 0.4972, 0.0488, 0.0955, 0.1734,  0.9891,  0.1291)
#
# load package
library(meta)
# use of metacont function
meta.1 <- metacont(n1, x1, sqrt(var1), n2, x2, sqrt(var2), sm="SMD")
meta.2 <- metacont(n1, x1, sqrt(var1), n2, x2, sqrt(var2), sm="SMD", hakn=T)
```

Example: R Output (slightly modified)

```
> summary(meta.1)

                          SMD            95%-CI      z  p-value
Fixed effect model   0.4202 [0.2570; 0.5834] 5.0465 < 0.0001
Random effects model 0.4237 [0.2259; 0.6215] 4.1988 < 0.0001

Quantifying heterogeneity:
tau^2 = 0.0225; H = 1.18 [1; 1.76]; I^2 = 28.1% [0%; 67.8%]

Test of heterogeneity:
    Q d.f.  p-value
 9.74    7   0.2037

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
- Hedges' g (bias corrected standardised mean difference)
```

Example: R Output (slightly modified)

```
> summary(meta.2)

                          SMD            95%-CI    z|t  p-value
Fixed effect model   0.4202 [0.2570; 0.5834] 5.0465 < 0.0001
Random effects model 0.4237 [0.1759; 0.6716] 4.0428   0.0049

*** Heterogeneity statistics erased ***

Details on meta-analytical method:
- Inverse variance method
- DerSimonian-Laird estimator for tau^2
- Hartung-Knapp adjustment for random effects model
- Hedges' g (bias corrected standardised mean difference)
```

Example

Random effects results for **standardized mean difference**

Method			Standard	HKSJ
τ^2	$\hat{\tau}^2$	$\hat{\theta}$	95% CI_1 for θ	95% CI_2 for θ
DerSimonian-Laird	0.0227	0.4238	[0.2259 ; 0.6218]	[0.1757 ; 0.6720]
Hedges	0.0738	0.4192	[0.1632 ; 0.6751]	[0.1516 ; 0.6868]
ML	0	0.4204	[0.2574 ; 0.5835]	[0.1880 ; 0.6529]
REML	0.0076	0.4228	[0.2468 ; 0.5987]	[0.1841 ; 0.6614]
Sid./Jonk.	0.0737	0.4192	[0.1634 ; 0.6750]	[0.1517 ; 0.6867]

Controlled Trials with Binary Outcome

- Let us consider k independent trials comparing treatment (T) versus control (C) and the outcome is binary, e.g. success/failure.
- Let p_{Ti} be the success probability in the treatment group and p_{Ci} in the control group in the i th study.
    - Observed frequencies on two binary characteristics in study i

	Success	Failure	Total
Treatment	n_{T1i}	n_{T0i}	$n_{T.i}$
Control	n_{C1i}	n_{C0i}	$n_{C.i}$
Total	$n_{.1i}$	$n_{.0i}$	$n_{..i}$

Effect Sizes

- Risk difference: $\theta_{1i} := p_{Ti} - p_{Ci}$
- Relative risk: $\theta_{2i} := p_{Ti}/p_{Ci}$
- Odds ratio:

$$\theta_{3i} := \frac{p_{Ti}\,(1 - p_{Ci})}{(1 - p_{Ti})\,p_{Ci}}$$

- Variance stabilizing transformation

$$\theta_{4i} := \sin^{-1}(\sqrt{p_{Ti}}) - \sin^{-1}(\sqrt{p_{Ci}})$$

- Appropriate estimates of effect sizes plus standard errors
  $\Rightarrow$ pooling of results with generic inverse variance method
  either in fixed effect model or random effects model

Effect Sizes

- Choice of the effect size may introduce heterogeneity

Study	Treatment	Control	Effect sizes		
	p_T	p_C	θ_1	θ_2	θ_3
1	0.6	0.2	0.4	3	6
2	0.6	0.2	0.4	3	6

Study	Treatment	Control	Effect sizes		
	p_T	p_C	θ_1	θ_2	θ_3
1	0.6	0.2	0.4	3	6
2	0.7	0.3	0.4	2.333	5.444

- Bivariate approach with observed relative frequencies $(\hat{p}_{Ti}, \hat{p}_{Ci})$, $i = 1, \ldots, k$?

Risk Difference

- Estimate

$$\hat{\theta}_{1i} = \frac{n_{T1i}}{n_{T.i}} - \frac{n_{C1i}}{n_{C.i}} = \hat{p}_{Ti} - \hat{p}_{Ci}$$

- Variance estimate

$$\widehat{\mathsf{Var}}(\hat{\theta}_{1i}) = \frac{n_{T1i}\ n_{T0i}}{n_{T.i}^2\ (n_{T.i} - 1)} + \frac{n_{C1i}\ n_{C0i}}{n_{C.i}^2\ (n_{C.i} - 1)}$$

- Critical data situations, i.e. inverses of variances do not exist, occur only in extreme cases, namely when $n_{T1i} = n_{C1i} = 0$ or $n_{T0i} = n_{C0i} = 0$ or $n_{T1i} = n_{C0i} = 0$ or $n_{T0i} = n_{C1i} = 0$.

Risk Difference

Alternative method in **fixed effect model**: Mantel-Haenszel approach (Greenland, Robins, 1985)

- Overall estimate

$$\hat{\theta}_{1,MH} = \frac{\sum_{i=1}^{k} (n_{T1i}\ n_{C.i}/n_{..i} - n_{T.i}\ n_{C1i}/n_{..i})}{\sum_{i=1}^{k} n_{T.i}\ n_{C.i}/n_{..i}} = \frac{\sum_{i=1}^{k} \frac{n_{T.i}\ n_{C.i}}{n_{..i}}\ \hat{\theta}_{1i}}{\sum_{i=1}^{k} \frac{n_{T.i}\ n_{C.i}}{n_{..i}}}$$

- The weight $\frac{n_{T.i}\ n_{C.i}}{n_{..i}}$ is the inverse of the variance of the RD estimator in a study when there is no difference between treatment and control and the underlying probabilities are the same in all studies.
- Asymptotic variance of $\hat{\theta}_{1,MH}$, see Greenland and Robins (1985)
- Confidence interval:

$$\hat{\theta}_{1,MH} \pm \sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_{1,MH})}\ z_{1-\alpha/2}$$

Relative Risk

- Estimate

$$\hat{\theta}_{2i} = \frac{n_{T1i} + 0.5}{n_{T.i} + 0.5} \Big/ \frac{n_{C1i} + 0.5}{n_{C.i} + 0.5}$$

- Variance estimate

$$\widehat{\text{Var}}(\ln \hat{\theta}_{2i}) = \frac{1}{n_{T1i} + 0.5} - \frac{1}{n_{T.i} + 0.5} + \frac{1}{n_{C1i} + 0.5} - \frac{1}{n_{C.i} + 0.5}.$$

- Combine results on the log scale and then back transform the meta-analysis result to the original scale.
- Note: the variance estimate changes when the role of success and failure changes, see discussion later.
- Often 0.5 is only added if they are zero events in a group. Otherwise

$$\hat{\theta}_{2i} = \frac{\hat{p}_{Ti}}{\hat{p}_{Ci}}.$$

Relative Risk

Overall Mantel-Haenszel-type estimator in the **fixed effect model**

- Estimate

$$\hat{\theta}_{2,MH} = \frac{\sum_{i=1}^{k} n_{T1i}\ n_{C.i}/n_{..i}}{\sum_{i=1}^{k} n_{C1i}\ n_{T.i}/n_{..i}} = \frac{\sum_{i=1}^{k} \frac{n_{C1i}\ n_{T.i}}{n_{..i}}\ \hat{\theta}_{2i}}{\sum_{i=1}^{k} \frac{n_{C1i}\ n_{T.i}}{n_{..i}}}$$

- Variance estimate

$$\widehat{\mathrm{Var}}(\ln \hat{\theta}_{2,MH}) = \frac{\sum_{i=1}^{k}(n_{T.i}\ n_{C.i}\ n_{.1i} - n_{T1i}\ n_{C1i}\ n_{..i})/n_{..i}^2}{(\sum_{i=1}^{k} n_{T1i}\ n_{C.i}/n_{..i})\ (\sum_{i=1}^{k} n_{C1i}\ n_{T.i}/n_{..i})}$$

- Confidence interval on log scale

$$\ln \hat{\theta}_{2,MH} \pm \sqrt{\widehat{\mathrm{Var}}(\ln \hat{\theta}_{2,MH})}\ z_{1-\alpha/2}$$

- Result is back transformed to the relative risk scale.

Odds Ratio

- Estimate

$$\hat{\theta}_{3i} = \frac{(n_{T1i} + 0.5)(n_{C0i} + 0.5)}{(n_{T0i} + 0.5)(n_{C1i} + 0.5)}$$

- Variance estimate

$$\widehat{\text{Var}}(\ln \hat{\theta}_{3i}) = \frac{1}{n_{T1i} + 0.5} + \frac{1}{n_{T0i} + 0.5} + \frac{1}{n_{C1i} + 0.5} + \frac{1}{n_{C0i} + 0.5}$$

- Combine results on the log scale and then back transform the meta-analysis result to the original scale.
- Note: the variance estimate does not change when the role of success and failure changes, see discussion later.
- Often 0.5 is only added if they are zero events in a group. Otherwise

$$\hat{\theta}_{3i} = \frac{\hat{p}_{Ti}\ (1 - p_{Ci})}{(1 - \hat{p}_{Ti})\ \hat{p}_{Ci}}$$

Odds Ratio

Overall Mantel-Haenszel-type estimator in the **fixed effect model**

- Estimate

$$\hat{\theta}_{3,MH} = \frac{\sum_{i=1}^{k} n_{T1i}\ n_{C0i}/n_{..i}}{\sum_{i=1}^{k} n_{T0i}\ n_{C1i}/n_{..i}} = \frac{\sum_{i=1}^{k} \frac{n_{T0i}\ n_{C1i}}{n_{..i}}\ \hat{\theta}_{3i}}{\sum_{i=1}^{k} \frac{n_{T0i}\ n_{C1i}}{n_{..i}}}$$

- The Mantel-Haenszel weights are based on the assumption that the odds ratios are one in all studies and have the advantage of guaranteeing that the summary odds ratio will be finite even if some of the individual odds rations are infinite because of one or more zero entries in the 2×2-tables.
- Variance estimate, see Robins, Breslow, Greenland (1986),

$$\widehat{\text{Var}}(\ln \hat{\theta}_{3,MH}) = \ldots$$

- Confidence interval on log scale

$$\ln \hat{\theta}_{3,MH} \pm \sqrt{\widehat{\text{Var}}(\ln \hat{\theta}_{3,MH})}\ z_{1-\alpha/2}$$

Odds Ratio

Peto-Method in the **fixed effect model**

- Estimate

$$\ln \hat{\theta}_{3i,Peto} = \frac{O_i - E_i}{V_i} = \frac{n_{T1i} - n_{T.i}\ n_{.1i}/n_{..i}}{V_i}$$

- Variance estimate

$$\widehat{\mathrm{Var}}(\ln \hat{\theta}_{3i,Peto}) = \frac{1}{V_i} = \left(\frac{n_{T.i}\ n_{C.i}\ n_{.1i}\ n_{.0i}}{(n_{..i} - 1)\ n_{..i}^2} \right)^{-1}$$

- Apply inverse variance method in the fixed effect model for pooling the results
- The approximation used in the computation of the log odds ratio works well when intervention effects are small (odds ratios are close to one), events are not particularly common and the studies have similar numbers in experimental and control groups. In other situations it has been shown to give biased answers. As these criteria are not always fulfilled, Peto's method is not recommended as a default approach for meta-analysis.

Function metabin in R package `meta`

Function with essential arguments

```
metabin(event.e,n.e,event.c, n.c,method="Inverse",sm="OR", ... )
```

Arguments	Description
event.e	Number of events in experimental group.
n.e	Number of observations in experimental group.
event.c	Number of events in control group.
n.c	Number of observations in control group.
method	A character string indicating which method is to be used for pooling of studies. One of `"Inverse"`, `"MH"`, or `"Peto"`, can be abbreviated.
sm	A character string indicating which summary measure (`"RR"`, `"OR"`, `"RD"`, or `"ASD"`) is to be used for pooling of studies.
incr	Numerical value (default 0.5) which is added to each cell frequency for studies with a zero cell count
allincr	A logical indicating if incr is added to each cell frequency of all studies if at least one study has a zero cell count. If FALSE (default), incr is added only to each cell frequency of studies with a zero cell count.

An Example

- Data: rare binary events, $k = 3$ trials

		Treatment		Control	
Study	Duration	Events	Total N	Events	Total N
1	6 weeks	5	2209	5	2224
2	6 weeks	2	1228	8	1229
3	9 weeks	0	1220	6	1239

- Which effect size? Additive (risk difference) or multiplicative (relative risk / odds ratio)?
- Which model? Fixed effect or random effects?
- Which method? 'Standard' interval or HKSJ interval or MH interval (if fixed model is approriate)?
  Note for HKSJ interval: $k = 3$ trials $\Rightarrow$ t-quantile with 2 df, $t_{2;0.975} = 4.303$ but $z_{0.975} = 1.96$

An Example

Risk Difference

R package *meta*, function *metabin*, Inverse Variance Method

```
       RD              95%-CI %W(fixed) %W(random)
1  0.0000  [-0.0028;  0.0028]     56.96      41.62
2 -0.0049  [-0.0099;  0.0001]     17.56      26.62
3 -0.0048  [-0.0090; -0.0007]     25.48      31.76

                         RD            95%-CI       z  p.value
Fixed effect model   -0.0021  [-0.0042; 0e+00] -1.9361   0.0529
Random effects model -0.0028  [-0.0064; 8e-04] -1.5434   0.1227

Test of heterogeneity:
    Q d.f.  p.value
 5.03    2   0.0807
```

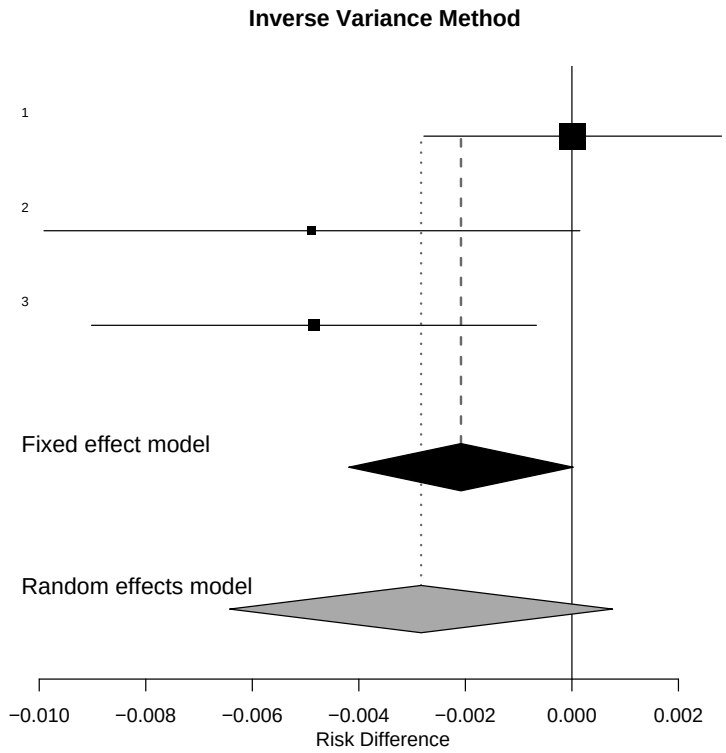
An Example

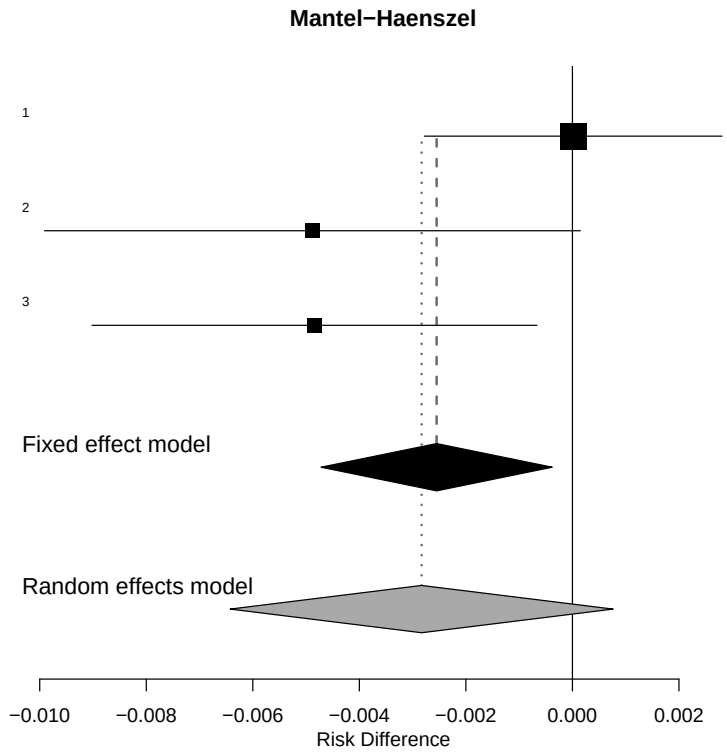
Risk Difference

R package *meta*, function *metabin*, Mantel-Haenszel Method

```
       RD              95%-CI %W(fixed) %W(random)
1  0.0000  [-0.0028;  0.0028]     47.42      41.62
2 -0.0049  [-0.0099;  0.0001]     26.28      26.62
3 -0.0048  [-0.0090; -0.0007]     26.30      31.76

                          RD             95%-CI       z  p.value
Fixed effect model   -0.0025  [-0.0047; -4e-04] -2.3021   0.0213
Random effects model -0.0028  [-0.0064;  8e-04] -1.5434   0.1227
```

An Example


Inverse Variance Method
1
2
3
Fixed effect model
Random effects model
−0.010
−0.008
−0.006
−0.004
−0.002
0.000
0.002
Risk Difference


Mantel−Haenszel
1
2
3
Fixed effect model
Random effects model
−0.010
−0.008
−0.006
−0.004
−0.002
0.000
0.002
Risk Difference

An Example

Odds Ratio

R package *meta*, function *metabin*, Inverse Variance Method

```
      OR            95%-CI %W(fixed) %W(random)
1 1.0068  [0.2911; 3.4826]     54.77      45.75
2 0.2490  [0.0528; 1.1748]     35.04      37.45
3 0.0777  [0.0044; 1.3815]     10.19      16.80

                         OR            95%-CI       z  p.value
Fixed effect model   0.4753  [0.1897; 1.1909] -1.5871   0.1125
Random effects model 0.3880  [0.1018; 1.4784] -1.3871   0.1654

Test of heterogeneity:
    Q d.f.  p.value
 3.59    2   0.1659
```

An Example

Odds Ratio

R package *meta*, function *metabin*, Mantel-Haenszel Method

```
      OR            95%-CI %W(fixed) %W(random)
1 1.0068  [0.2911; 3.4826]     25.62      45.75
2 0.2490  [0.0528; 1.1748]     41.15      37.45
3 0.0777  [0.0044; 1.3815]     33.23      16.80

                          OR           95%-CI       z  p.value
Fixed effect model    0.3863  [0.1662; 0.8975] -2.2114   0.027
Random effects model  0.3880  [0.1018; 1.4784] -1.3871   0.1654
```

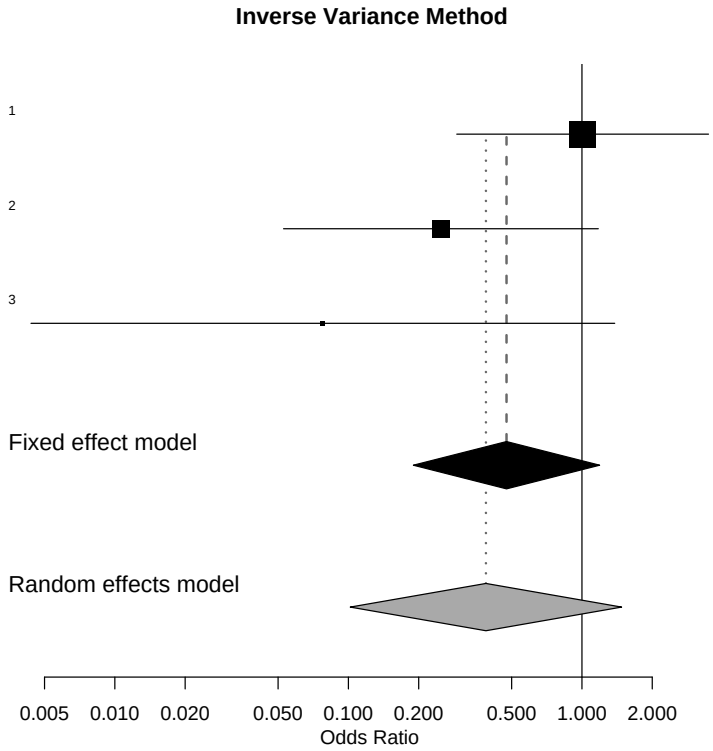
An Example

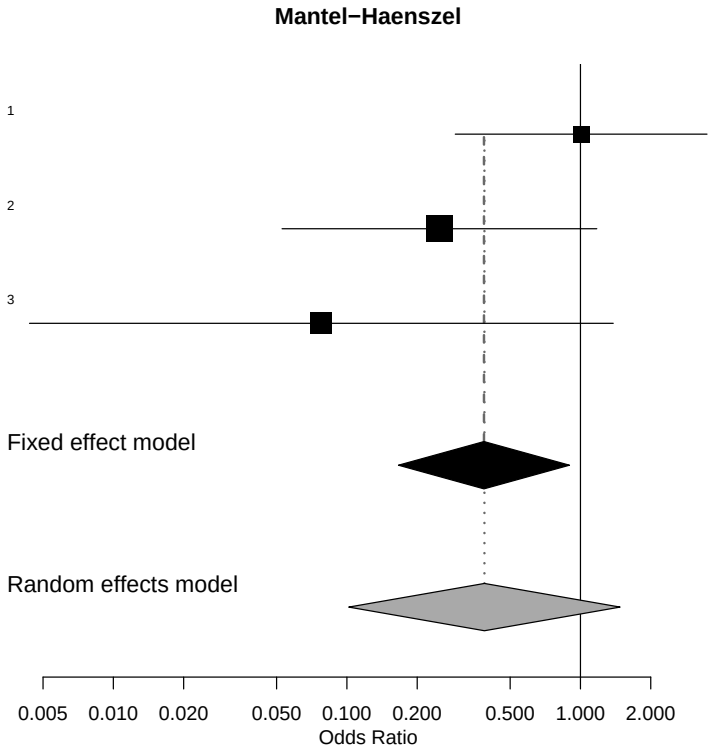
Odds Ratio

R package *meta*, function *metabin*, Peto Method

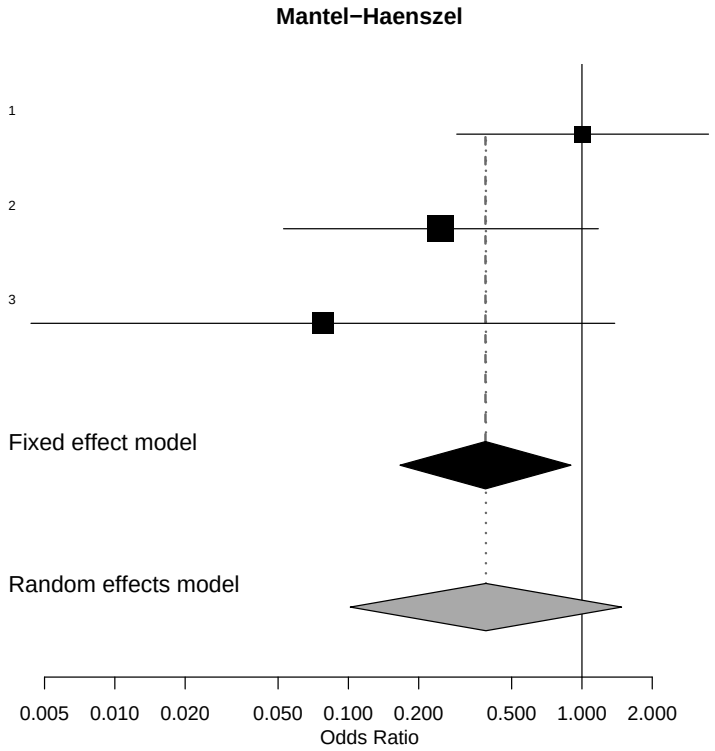
```
      OR            95%-CI %W(fixed)
1 1.0068  [0.2911; 3.4821]     38.49
2 0.3001  [0.0867; 1.0390]     38.42
3 0.1369  [0.0276; 0.6793]     23.09

                         OR            95%-CI       z  p.value
Fixed effect model   0.3989  [0.1847; 0.8613] -2.3400   0.0193
```

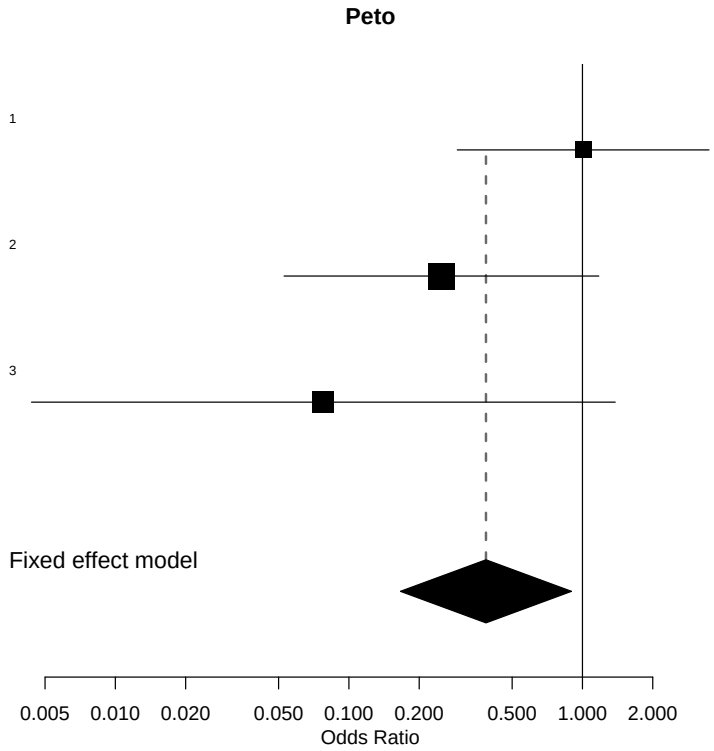
An Example


Inverse Variance Method
1
2
3
Fixed effect model
Random effects model
0.005
0.010
0.020
0.050
0.100
0.200
0.500
1.000
2.000
Odds Ratio


Mantel−Haenszel
1
2
3
Fixed effect model
Random effects model
0.005
0.010
0.020
0.050
0.100
0.200
0.500
1.000
2.000
Odds Ratio

An Example


Mantel−Haenszel
1
2
3
Fixed effect model
Random effects model
0.005
0.010
0.020
0.050
0.100
0.200
0.500
1.000
2.000
Odds Ratio


Peto
1
2
3
Fixed effect model
0.005
0.010
0.020
0.050
0.100
0.200
0.500
1.000
2.000
Odds Ratio

Summary Statistics for Binary Data

Deeks (2002): Selection of a summary statistics

- risk difference (RD)
- risk ratio (RR)
  - risk ratio of beneficial outcomes (RR(B))
  - risk ratio of harmful outcomes (RR(H))
- odds ratio (OR)

How do we select the summary statistic for meta-analysis?

Deeks (2002): Selection of a summary statistics

- The selection of the appropriate summary statistic is a subject of debate due to conflicts in the relative importance of mathematical properties and the ability of intuitively interpret results.
- The appropriate selection for a particular meta-analysis may depend on understanding reasons for variation in control group event rates.

Effect of Coding of Binary Data

Meta-analysis on thrombolytic therapy after acute myocardial infarction

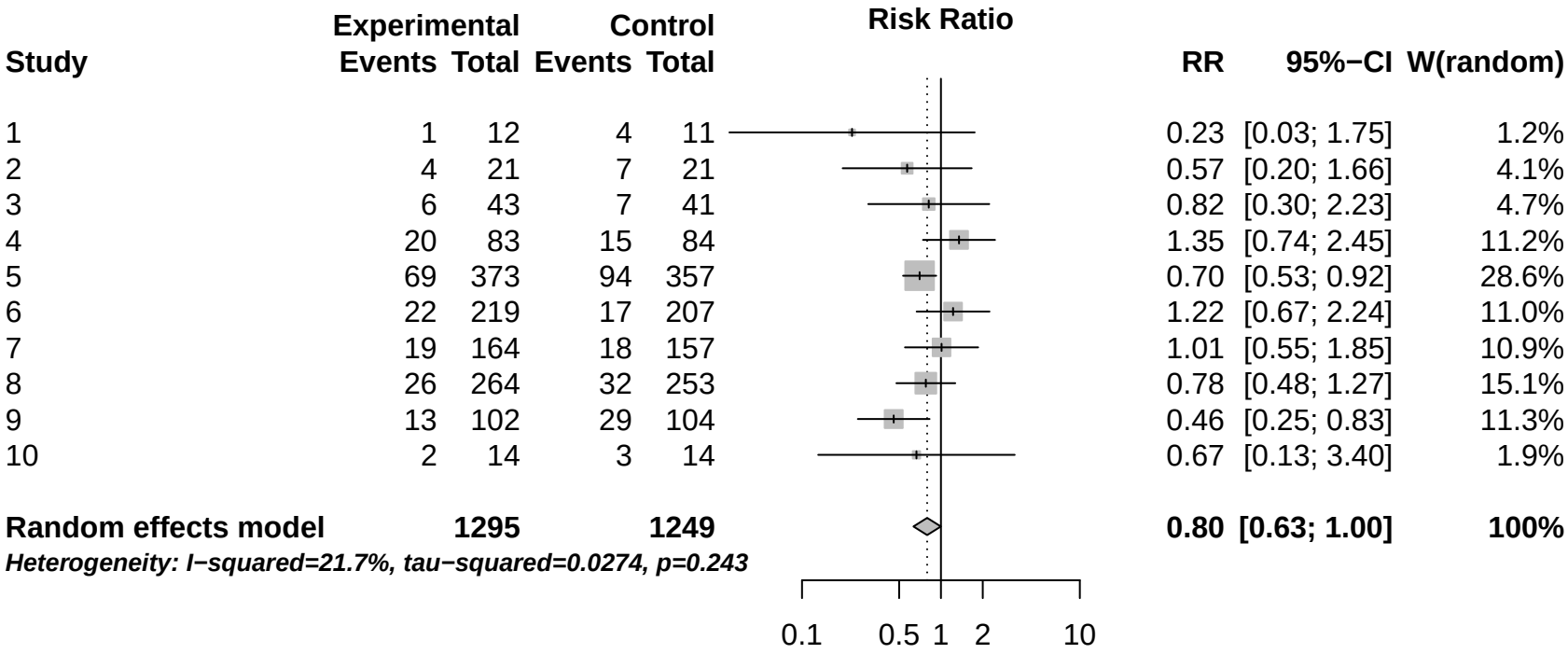
Author	Year	Event.e	n.e	Event.c	n.c
Fletcher	1959	1	12	4	11
Dewar	1963	4	21	7	21
Lippschutz	1965	6	43	7	41
European 1	1969	20	83	15	84
European 2	1971	69	373	94	357
Heikinheimo	1971	22	219	17	207
Italian	1971	19	164	18	157
Australian 1	1973	26	264	32	253
Frankfurt 2	1973	13	102	29	104
Gormsen	1973	2	14	3	14

Effect of Coding of Binary Data

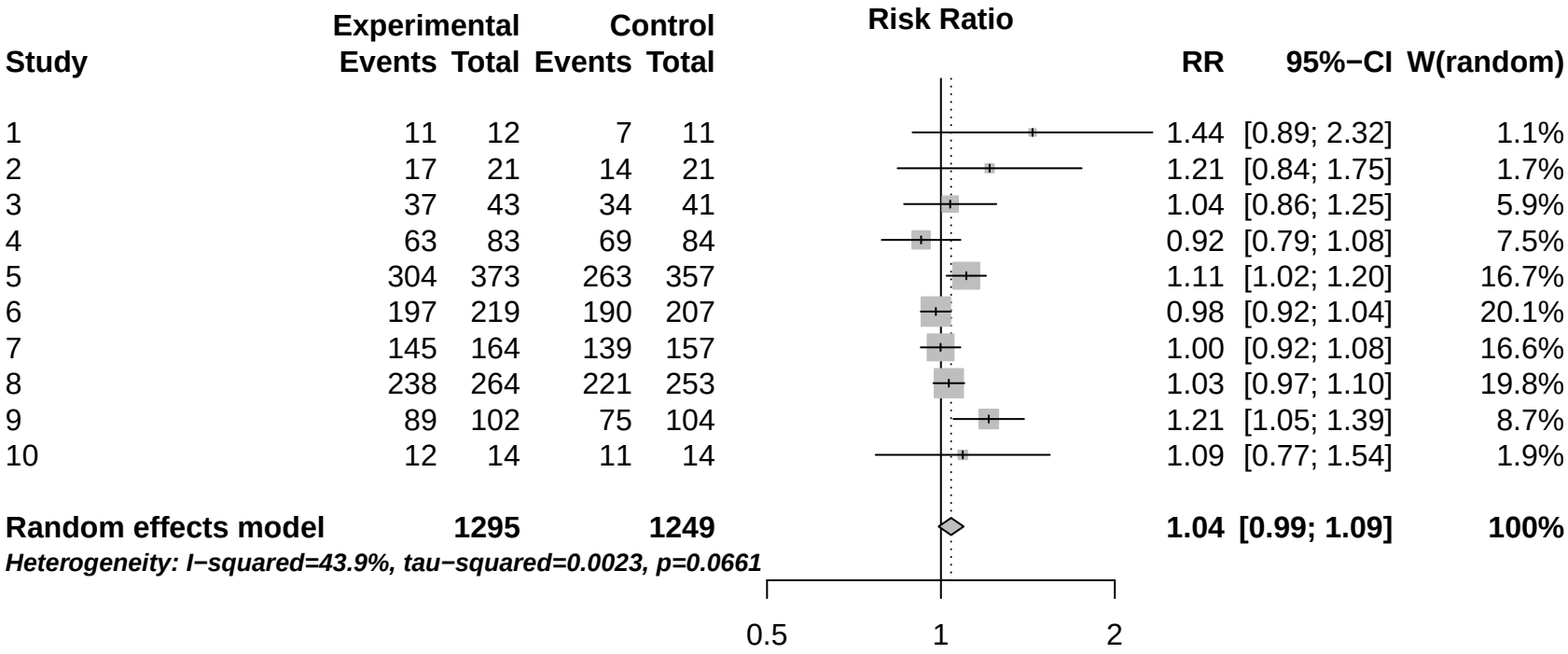
Random effects meta-analysis on thrombolytic therapy after acute myocardial infarction

Effect size	Events	Non-Events
Risk difference	-0.0361 (95% CI: -0.08 – 0.01) p=0.0934	0.0361 (95% CI: -0.01 – 0.08) p=0.0934
Odds ratio	0.7604 (95% CI: 0.57 – 1.01) p=0.0606	1.3151 (95% CI: 0.99 – 1.75) p=0.0606
Fisher's v.s.t.	-0.0502 (95% CI: -0.11 – 0.01) p=0.0789	0.0502 (95% CI: -0.01 – 0.11) p=0.0789

Effect of Coding of Binary Data



Effect of Coding of Binary Data



Effect of Coding of Binary Data

- Meta-analysis conclusions are not affected by the coding (0/1) of binary data for risk difference, odds ratio, and Fisher's v.s.t.
- For risk ratio, we must consider whether the outcome is beneficial or harmful.
- Meta-analysis of control group (prevalence estimate) should be added.
  Last example: Prob = 0.19 (95% CI: 0.13 – 0.24)

Publication Bias

- Meta-analysis is often restricted to published studies.
- Risk of biased conclusions when there may be many nonsignificant studies which are often unpublished and hence are ignored.
- It is quite possible that their combined effect, significant and nonsignificant studies together, may change the overall conclusion.
- Publication bias results from ignoring unavailable nonsignificant studies.
- Note: If the meta-analysis of k available studies leads to a nonsignificant conclusion (when the difference between the groups is important), then the issue of publication bias does not arise!
- The notion **small study effects** is now often used in the context of publication bias.

File-Drawer Method

- File-drawer method (Rosenthal, 1979) is designed to provide a simple qualification on a summary P-value from a meta-analysis.
- Assume that the meta-analysis of k available studies leads to a significant result, that is, the combination of k P-values leads to rejection of the null hypothesis H_0.
- Question: Does a set of k_0 unpublished nonsignificant studies exist, which would have made the rejection of H_0 on the basis of all the $k + k_0$ studies impossible?
- Use of Inverse Normal method for combining P-values.
- Let be $P_1, \ldots, P_k$ the P-values of the k published studies.
- Compute $Z = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Phi^{-1}(P_i) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} Z_i$ and reject H_0 if $|Z| > z_{1-\alpha}$.
- Assume that the average observed effect of the k_0 unpublished studies is 0, i.e., the sum of the Z scores corresponding to these k_0 studies is 0.
- Non-rejection of H_0, if $\frac{1}{\sqrt{k+k_0}} \left| \sum_{i=1}^{k} Z_i \right| < z_{1-\alpha}$,

$$\Longrightarrow k_0 > -k + \Big(\sum_{i=1}^{k} Z_i \Big)^2 / (z_{1-\alpha})^2.$$

Implementation in R

- The Rosenthal method (inverse normal method) is implemented in the function `fsn` (Fail-Safe N) in the package **metafor**, see R program FSN.R.

```
> ### load BCG vaccine data
> data(dat.colditz1994)
> ### calculate log relative risks and corresponding sampling variances
> dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos, di=cneg,
+  data=dat.colditz1994)
> fsn(yi, vi, data=dat)

Fail-safe N Calculation Using the Rosenthal Approach

Observed Significance Level: <.0001
Target Significance Level:   0.05

Fail-safe N: 598
```

Further Fail-safe N Methods

- The Orwin method calculates the number of studies averaging null results that would have to be added to the given set of observed outcomes to reduce the (unweighted) average effect size to a target (unweighted) average effect size.

  ```
  > fsn(yi, vi, data=dat, type="Orwin")
  ```

- The Rosenberg method calculates the number of studies averaging null results that would have to be added to the given set of observed outcomes to reduce significance level (p-value) of the (weighted) average effect size (based on a fixed-effect model) to a target alpha level (e.g., .05).

  ```
  > fsn(yi, vi, data=dat, type="Rosenberg")
  ```

Funnel Plot

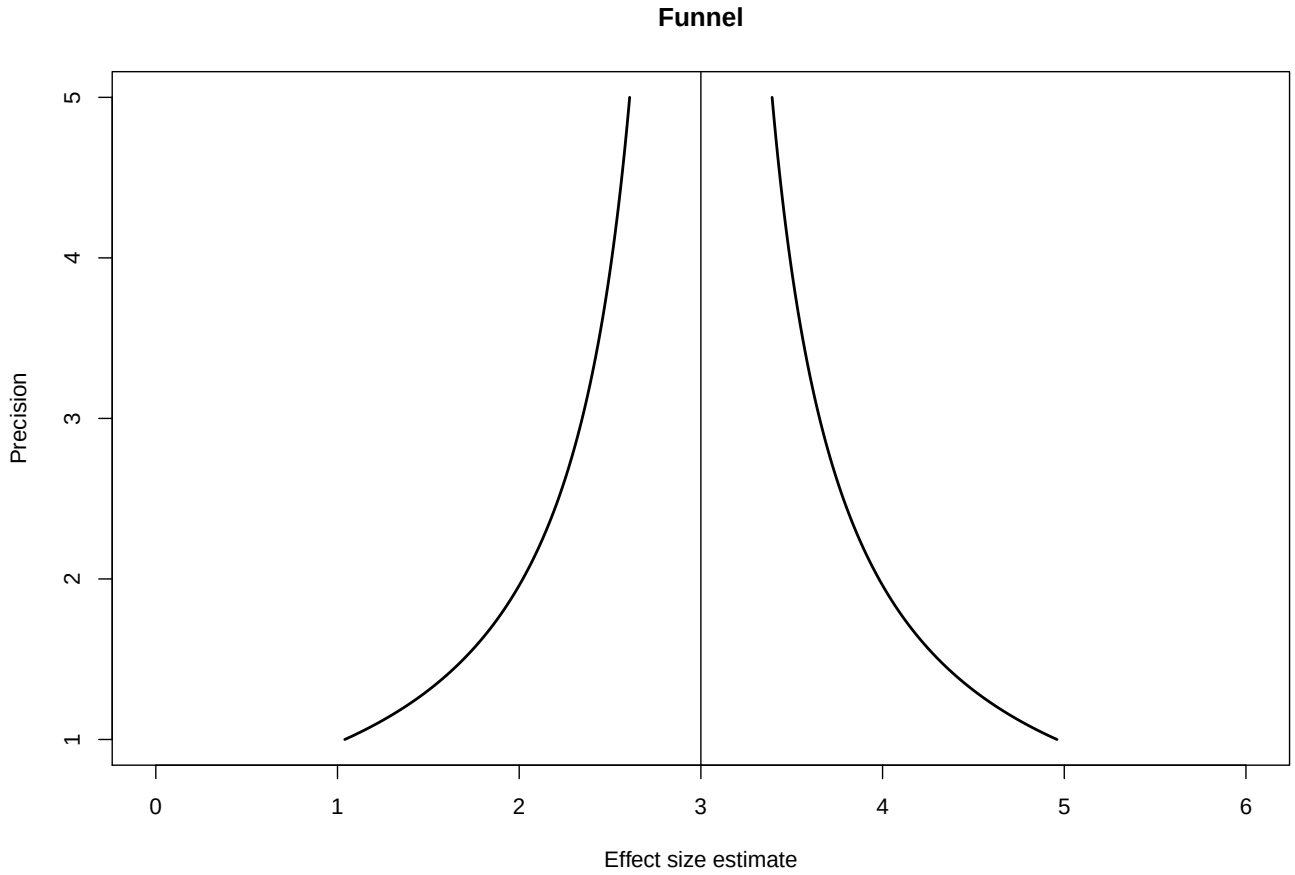
Idea: Consider independently and identically distributed random variables X_i $i = 1, \ldots, n$, $\text{Var}(X_1) = \sigma^2 \Rightarrow \bar{X}$ is (approximatively) normally distributed and an (approximative) $(1-\alpha)$-confidence interval for the expected value is given by

$$\bar{X} \pm \sqrt{\frac{\sigma^2}{n}}\, z_{1-\alpha/2}$$

Measure(s) for the precision of $\bar{X}$: $\frac{n}{\sigma^2}$, $\sqrt{\frac{n}{\sigma^2}}$, n, $\sqrt{n}$

Note: Under normality assumption, $\bar{X}$ and estimator of σ^2 are stochastically independent.

Funnel Plot

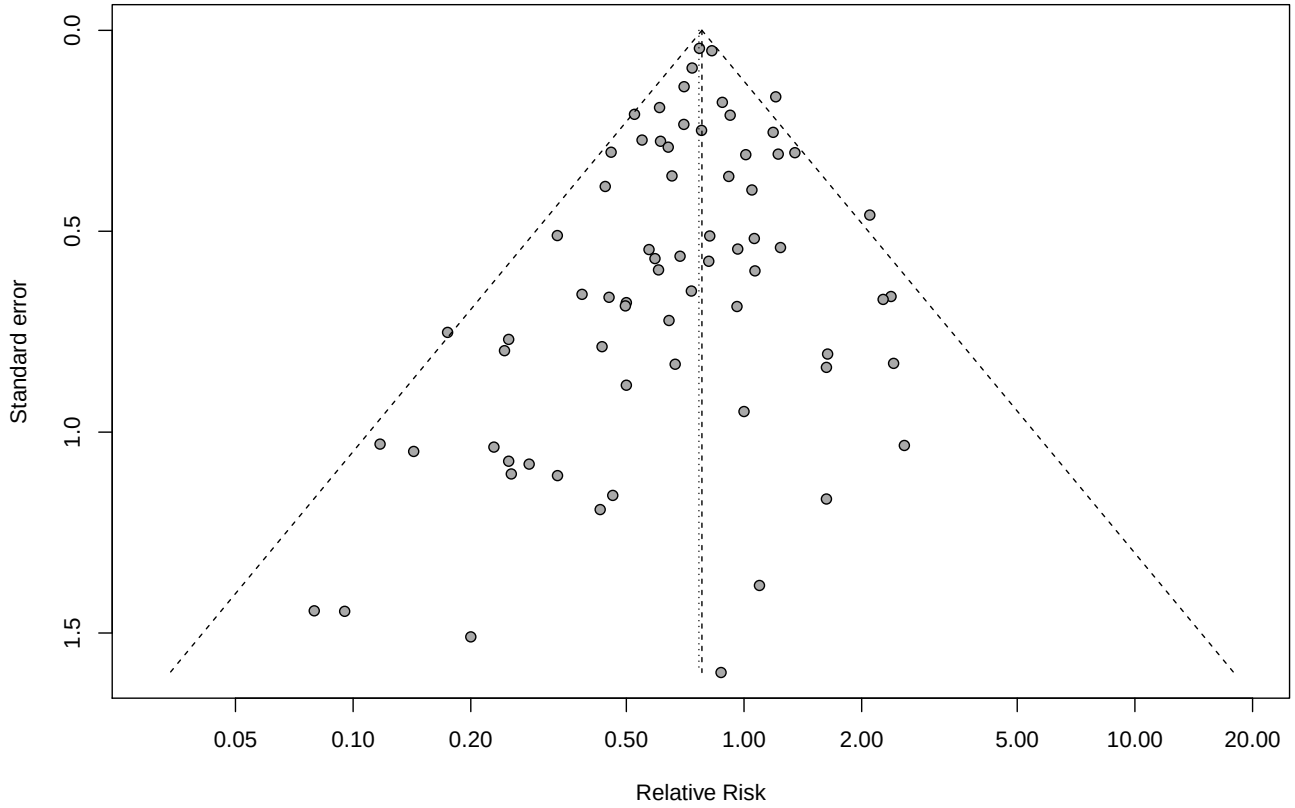


Funnel Plot

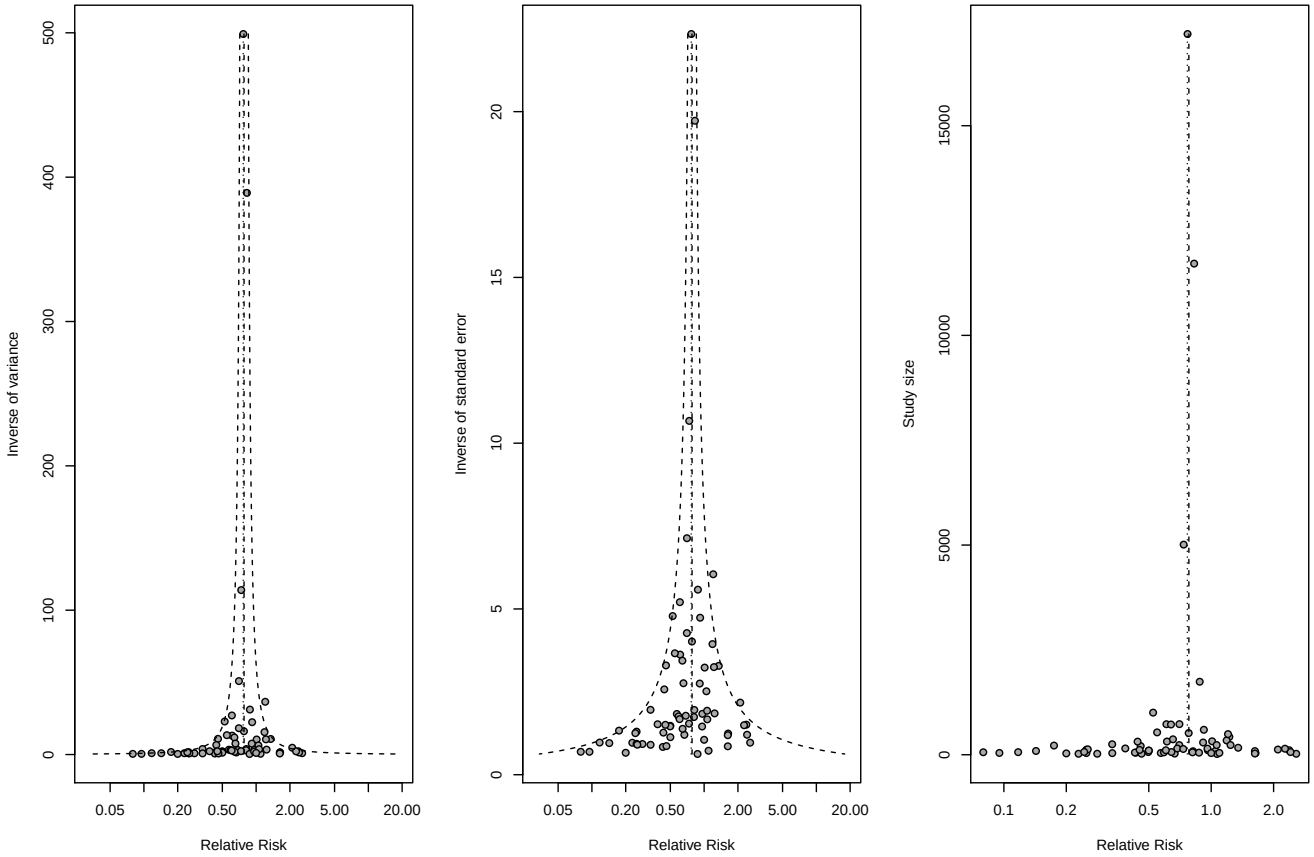
- Funnel-Plot (Light, Pillemer, 1984): Informal assessment of publication bias by a graphical method.
- If no bias is present, this plot would look like a funnel. This is because there will be a broad spread of points for the highly variable small studies (due to small sample sizes) at the bottom and decreasing spread as the sample sizes increase, with the indication that the publication bias is unlikely to be a factor of this meta-analysis.
- An asymmetric funnel plot may indicate a possible publication bias.

Choice of the measure of precision: "the standard error of the treatment estimates is plotted on the y axis which is likely to be the best choice"' (Sterne and Egger, 2001).

Funnel Plot


Standard error
0.0
0.5
1.0
1.5
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
20.00
Relative Risk

Different Choices of Measures of Precision



Tests for Publication Bias

Egger's regression test (Egger et al., 1997, BMJ):
Model 1:

$$\hat{\theta}_i = \beta_1 + \beta_2 \, \hat{\sigma}(\hat{\theta}_i) + e_i, \quad \text{Var}(e_i) = \hat{\sigma}^2(\hat{\theta}_i), \quad i = 1, \ldots, k.$$

Test for publication bias (or asymmetry of the funnel plot):

$$\text{H}_0 : \text{No Publication Bias} \iff \text{H}_0 : \beta_2 = 0$$

Reject H_0 at level α if

$$\frac{|\hat{\beta}_2|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2)}} > t_{k-2,1-\alpha/2}$$

Tests for Publication Bias

Model 2: (function metabias in the R package `meta`, method = "linreg")

$$\frac{\hat{\theta}_i}{\hat{\sigma}(\hat{\theta}_i)} = \beta_1 + \beta_2 \frac{1}{\hat{\sigma}(\hat{\theta}_i)} + \tilde{e}_i, \quad \text{Var}(\tilde{e}_i) = 1, \quad i = 1, \ldots, k.$$

Test for publication bias (or asymmetry of the funnel plot):

$$\text{H}_0 : \text{No Publication Bias} \iff \text{H}_0 : \beta_1 = 0$$

Reject H_0 at level α if

$$\frac{|\hat{\beta}_1|}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_1)}} > t_{k-2,1-\alpha/2}$$

Variants of the Regression Test

- In the above models: include $\hat{\tau}^2$ in the variance structure (`method="mm"` in function `metabias` of R package **meta**)
- $\hat{\theta}_i$ and $\hat{\sigma}(\hat{\theta}_i)$ are not stochastically independent for binary data $\Longrightarrow$ asymmetric funnel plot even if no publication bias is present
  - Peters et al. (2006): Instead of $\hat{\sigma}(\hat{\theta}_i)$ use the inverse of the total sample size $1/(n.E + n.C)$ as the explanatory variable in above model
  - Rücker et al. (2008): Use the arcsine transformed effect size in the funnel plot.
- All the methods are implemented in the function `metabias` in the R package **meta**. The regression based tests are also implemented in the function `regtest` in the R package **metafor**.

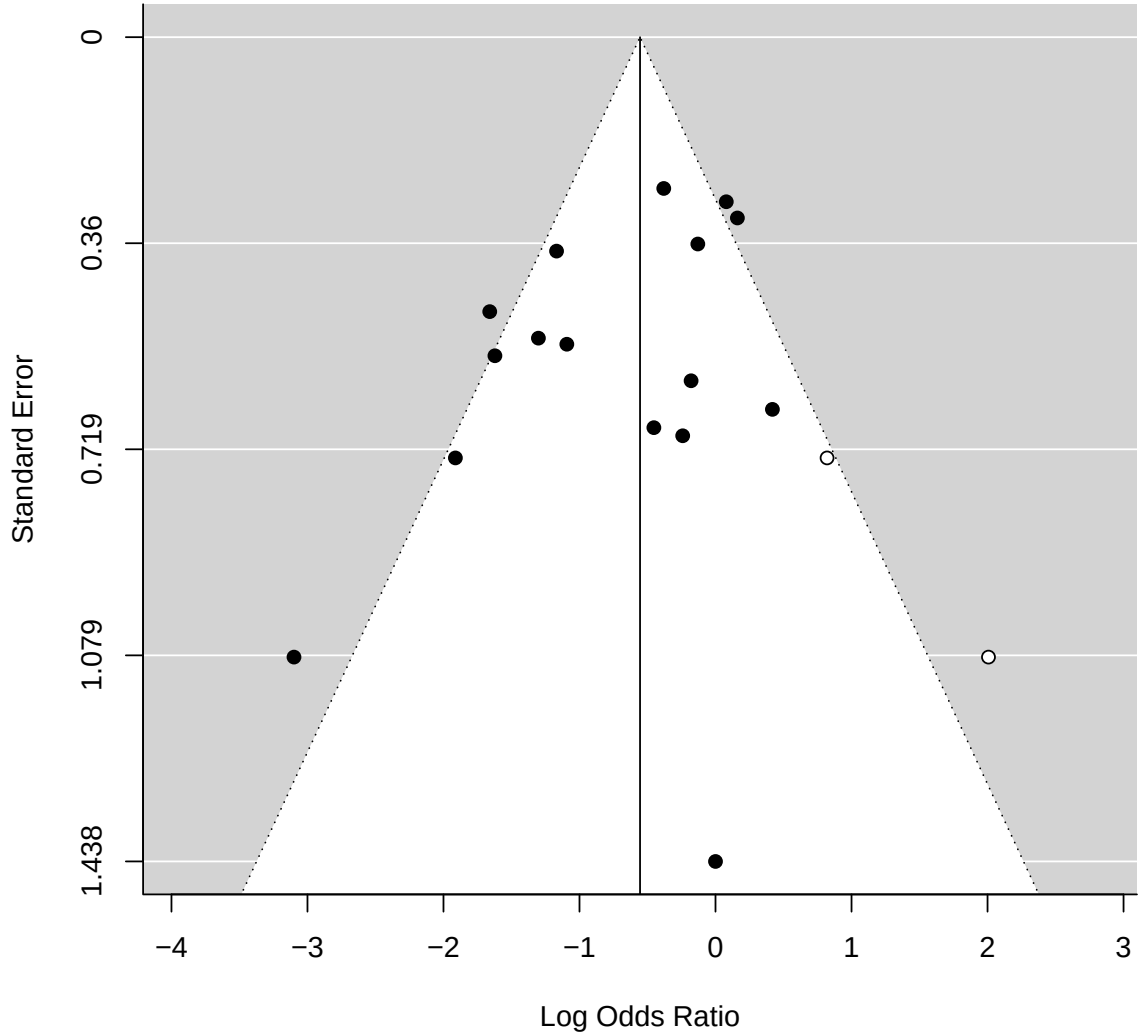
Correcting for Publication Bias: Trim-and-fill Method

- The trim-and-fill method is a nonparametric method to assess selection bias / publication bias.
- The method provides
  - the number of missing studies and
  - 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.
- Five steps:
  1. Estimate the number of studies in the outlying part using rank-based methods;
  2. remove (trim) these studies and do meta-analysis on the remaining studies;
  3. consider the estimated from the $trimmed$ meta-analysis as the true center of the funnel plot;
  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.

Correcting for Publication Bias: Trim-and-fill Method

- In steps 1-4 and 5, respectively, either a fixed effect or random effect model can be used. Recommendation: Fixed-effect model for steps 1-4 and random effects model for step 5.
- The trim-and-fill method assumes that the small-study effect is caused by selection, but requires no assumption about the mechanism leading to small-study effects.
- However, it is build on the strong assumption of a symmetric funnel plot.
- The method is known to perform poorly in the presence of substantial between-study heterogeneity.
- Estimation and inference are based on a dataset containing imputed intervention effect estimates, potentially resulting in too narrow confidence intervals for the average treatment effect.

Correcting for Publication Bias: Trim-and-fill Method



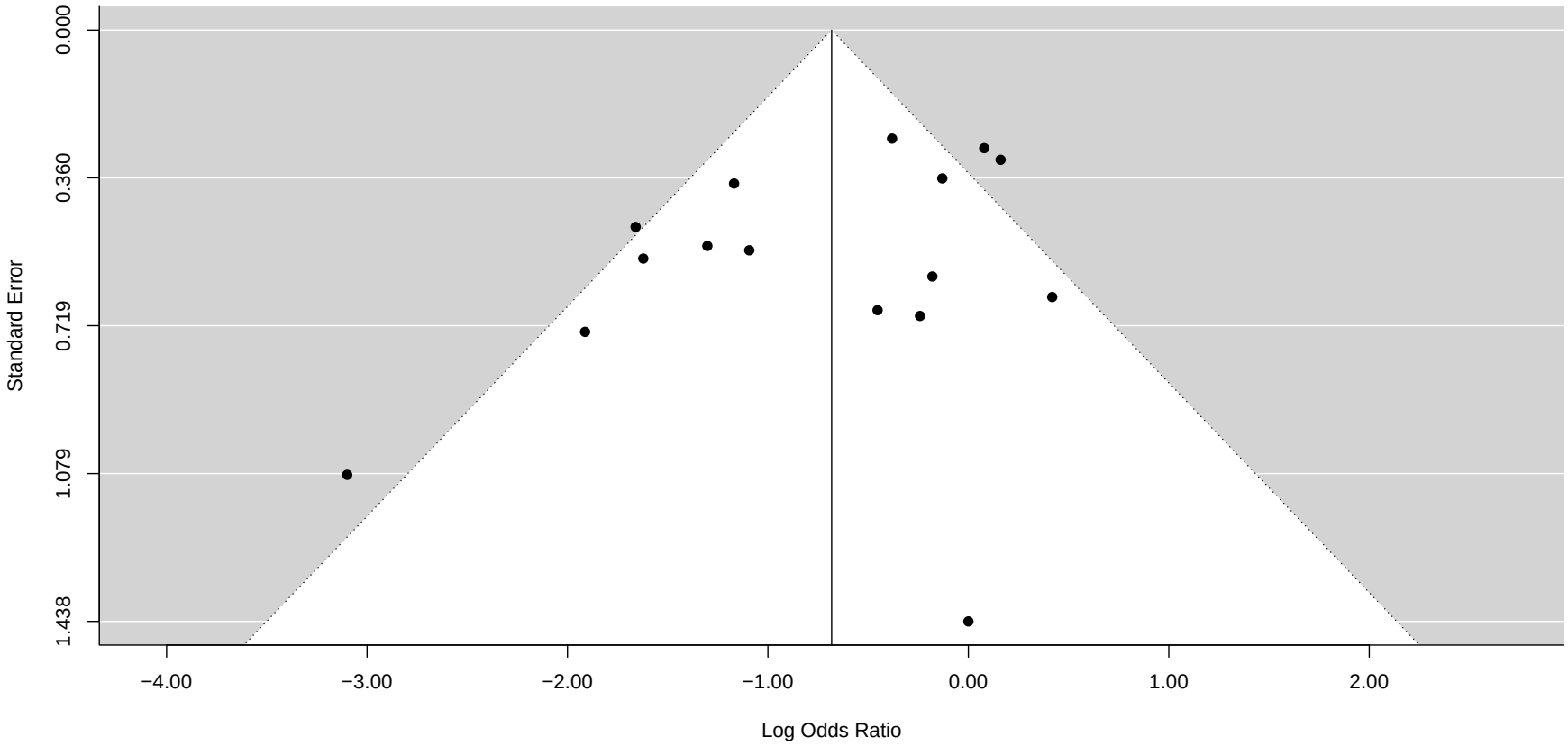
Correcting for Publication Bias: Henmi-Copas Approach

- Basic idea: Use the fixed-effect meta-analysis estimator, that is, use only the within-study variances as weights.
- Then, calculate the variance of this estimator under the random-effects meta-analysis model.
- Construct a confidence interval for the parameter of interest which should be robust to publication bias.

Henmi-Copas Approach in metafor

```
 ### load data from Lee & Done (2004)
 data(dat.lee2004)
 ### meta-analysis based on log odds ratios
 (res <- rma(measure="OR", ai=ai, n1i=n1i, ci=ci, n2i=n2i, data=dat.lee2004))

Random-Effects Model (k = 16; tau^2 estimator: REML)
tau^2 (estimated amount of total heterogeneity): 0.3526 (SE = 0.2254)

Model Results:
estimate       se     zval     pval    ci.lb    ci.ub
 -0.6820   0.2013  -3.3877   0.0007  -1.0766  -0.2874

 ### funnel plot
 funnel(res)
```

Henmi-Copas Approach in metafor



Henmi-Copas Approach in metafor

```
> ### use method by Henmi and Copas (2010) as a sensitivity analysis
> hc(res)
    method   tau2 estimate     se   ci.lb   ci.ub
rma   REML 0.3526  -0.6820 0.2013 -1.0766 -0.2874
hc      DL 0.3325  -0.5145 0.2178 -0.9994 -0.0295
>
> ### back-transform results to odds ratio scale
> hc(res, transf=exp)
    method   tau2 estimate se  ci.lb  ci.ub
rma   REML 0.3526   0.5056 NA 0.3408 0.7502
hc      DL 0.3325   0.5978 NA 0.3681 0.9709
```

Meta-Regression

- In case of substantial heterogeneity between the studies, possible causes of the heterogeneity should be explored.
- In the context of meta-analysis this can be done by either covariates on the study level that could explain the differences between the studies or by covariates on the subject level.
- However, the latter approach is only possible when individual data are available.
- Since often only information on the study level is available, explaining and investigating heterogeneity by covariates on the study level has drawn much attention in applied sciences.
- Since the number of studies in a meta-analysis is usually quite small, there is a great danger of overfitting.
- So, there is only room for a few explanatory variables in a meta-regression, whereas a lot of characteristics of the studies may be identified as potential causes of heterogeneity.
- Investigations of differences between the studies and their results are observational associations and are subject to biases (such as aggregation bias) and confounding (resulting from correlation between study characteristics).
- Consequently, there is a clear danger of misleading conclusions if P-values from multiple meta-regression analyses are interpreted naïvely.

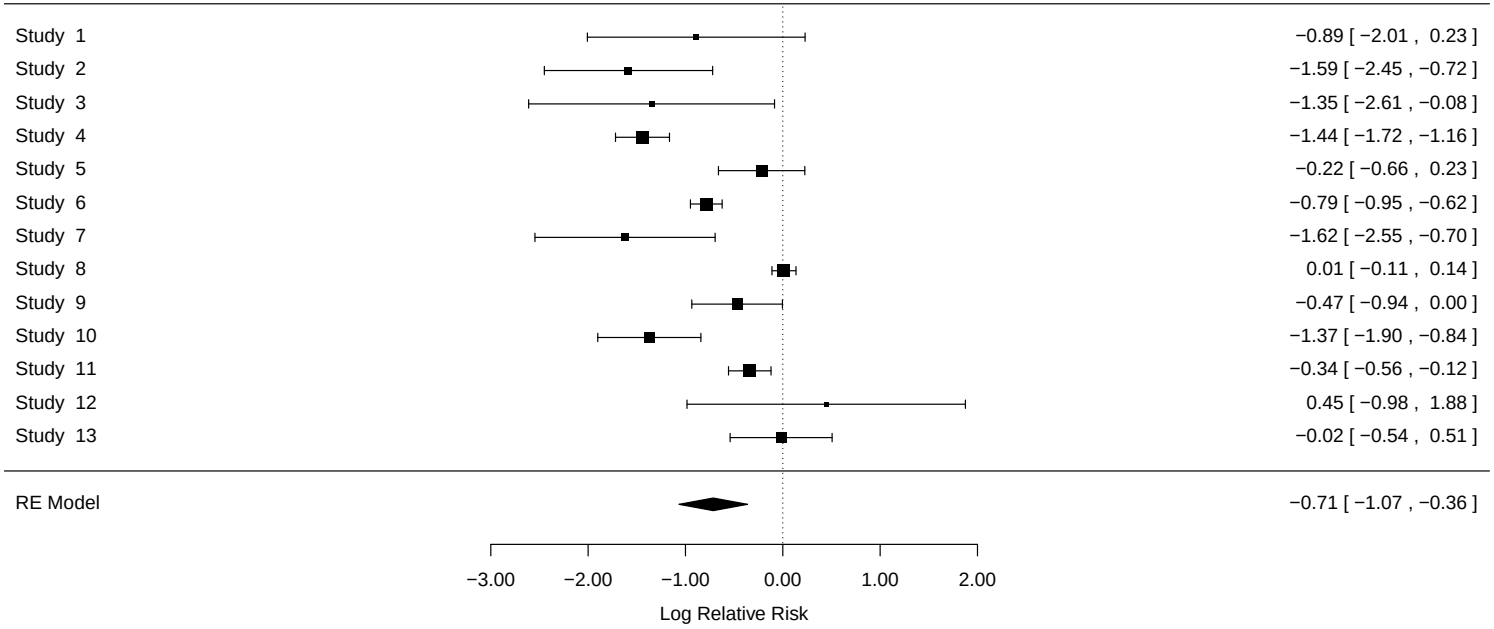
Example

Data on 13 trials on the prevention of tuberculosis using BCG vaccination

Trial	Vaccinated		Not vaccinated		Latitude
	Disease	No disease	Disease	No Disease	
1	4	119	11	128	44
2	6	300	29	274	55
3	3	228	11	209	42
4	62	13536	248	12619	52
5	33	5036	47	5761	13
6	180	1361	372	1079	44
7	8	2537	10	619	19
8	505	87886	499	87892	13
9	29	7470	45	7232	27*
10	17	1699	65	1600	42
11	186	50448	414	27197	18
12	5	2493	3	2338	33
13	27	16886	29	17825	33

Further covariates available: Year of publication, type of allocation (alternate, random, systematic)

Example


Study 1
Study 2
Study 3
Study 4
Study 5
Study 6
Study 7
Study 8
Study 9
Study 10
Study 11
Study 12
Study 13
RE Model
−0.89 [−2.01 , 0.23]
−1.59 [−2.45 , −0.72]
−1.35 [−2.61 , −0.08]
−1.44 [−1.72 , −1.16]
−0.22 [−0.66 , 0.23]
−0.79 [−0.95 , −0.62]
−1.62 [−2.55 , −0.70]
0.01 [−0.11 , 0.14]
−0.47 [−0.94 , 0.00]
−1.37 [−1.90 , −0.84]
−0.34 [−0.56 , −0.12]
0.45 [−0.98 , 1.88]
−0.02 [−0.54 , 0.51]
−0.71 [−1.07 , −0.36]
−3.00
−2.00
−1.00
0.00
1.00
2.00
Log Relative Risk

Meta-Regression (one covariate)

Let us consider k independent trials (experiments) and each trial provides an estimate, say $\hat{\theta}_i$, $i = 1, \ldots, k$, of a parameter of interest, say θ, and an estimate of the variance of $\hat{\theta}_i$, say $\hat{\sigma}_i^2$, $i = 1, \ldots, k$. Moreover, one covariate on study-level, say x_i, is known for each trial.

Normal-normal hierarchical model:

$$\hat{\theta}_i \sim N\left(\theta_i, \hat{\sigma}_i^2\right)$$

and

$$\theta_i \sim N\left(\theta_{MA}, \tau_{MA}^2\right)$$

OR

$$\theta_i \sim N\left(\theta_{MR} + \beta x_i, \tau_{MR}^2\right)$$

Meta-Regression (one covariate)

Random effects meta-analysis model

$$\hat{\theta}_i \sim N\left(\theta_{MA}, \tau^2_{MA} + \hat{\sigma}^2_i\right).$$

- θ_{MA} – average effect size
- τ^2_{MA} – between-study variability (heterogeneity parameter)

Random effects meta-regression

$$\hat{\theta}_i \sim N\left(\theta_{MR} + \beta x_i \ , \ \tau^2_{MR} + \hat{\sigma}^2_i\right).$$

- θ_{MR} – effect size given that the covariate is zero
- τ^2_{MR} – residual heterogeneity
- β – average change in the effect size when the covariate increases by one unit

Note: The covariate must have a meaningful interpretation for the issue.
Target:

$$\tau^2_{MR} << \tau^2_{MA}$$

Meta-Regression (one covariate)

Random effects meta-regression

$$\hat{\theta}_i \sim N\left(\theta_{MR} + \beta x_i \ , \ \tau^2_{MR} + \hat{\sigma}^2_i\right).$$

Objectives:

- Fixed effects or random effects meta-regression?
  Test of $H_0 : \tau^2_{MR} = 0$!
- Estimate and confidence interval for τ^2_{MR}
- Estimates and confidence intervals for θ_{MR} and β

Extend analysis methods from meta-analysis to meta-regression:

- (Conditional) Restricted maximum likelihood estimation for τ^2_{MR}
- Method of moments estimation for τ^2_{MR}
- Weighted least-squares regression for the fixed effects.

Note: The model can be easily extended to more than one covariate.

Inference on the Fixed Effects

- Let $\tilde{\theta}$ and $\tilde{\beta}$ be the weighted least-squares estimators with known variances.
- Knapp and Hartung (2003) considered the quadratic form

$$Q_2 = \frac{1}{k-2} \sum_{i=1}^{k} w_i \left(Y_i - \tilde{\theta} - \tilde{\beta}\, x_i\right)^2 , \;\; k > 2 \, .$$

  that is, a mean sum of the weighted least-squares residuals.
- Under normality of Y_i, the quadratic form Q_2 is stochastically independent of the $\tilde{\theta}$ and $\tilde{\beta}$, and $(k-2)\, Q_2$ is χ^2-distributed with $k-2$ degrees of freedom.
- Hence, unbiased and non-negative estimators of the variances of $\tilde{\theta}$ and $\tilde{\beta}$ are given by

$$Q_2(\tilde{\theta}) = \frac{1}{k-2} \sum_{i=1}^{k} g_i \left(Y_i - \tilde{\theta} - \tilde{\beta}\, x_i\right)^2$$

  with $\quad g_i = w_i \,/\, [\sum w_j - (\sum w_j\, x_j)^2 / \sum w_j\, x_j^2], \;\; i = 1, \ldots, k$, and

$$Q_2(\tilde{\beta}) = \frac{1}{k-2} \sum_{i=1}^{k} h_i \left(Y_i - \tilde{\theta} - \tilde{\beta}\, x_i\right)^2$$

  with $\quad h_i = w_i \,/\, [\sum w_j\, x_j^2 - (\sum w_j\, x_j)^2 / \sum w_j], \;\; i = 1, \ldots, k.$

Inference on the Fixed Effects

- Replacing the unknown variance components in $Q_2(\tilde{\theta})$ and $Q_2(\tilde{\beta})$ by appropriate estimates, Knapp and Hartung (2003) proposed the following approximate $(1-\alpha)$-confidence intervals on θ and β:

$$\hat{\theta} \pm \sqrt{\hat{Q}_2(\hat{\theta})}\; t_{k-2;\alpha/2}$$

and

$$\hat{\beta} \pm \sqrt{\hat{Q}_2(\hat{\beta})}\; t_{k-2;\alpha/2}\;.$$

Example

Results (log RR) for slope with covariate latitude

Method $\hat{\tau}^2$	Estimate	95% CI (classical)	95% CI (KH)
Hunter-Schmidt	-0.0296	[-0.0398, -0.0193]	[-0.0447, -0.0144]
Hedges	-0.0282	[-0.0489, -0.0075]	[-0.0493, -0.0071]
DerSimonian-Laird	-0.0292	[-0.0424, -0.0160]	[-0.0467, -0.0118]
Sidik-Jonkman	-0.0281	[-0.0497, -0.0065]	[-0.0495, -0.0067]
ML	-0.0295	[-0.0403, -0.0188]	[-0.0452, -0.0139]
REML	-0.0291	[-0.0432, -0.0150]	[-0.0472, -0.0111]
Paule-Mandel	-0.0286	[-0.0463, -0.0108]	[-0.0485, -0.0086]

Explaining Heterogeneity

Log relative risk scale

Method $\hat{\tau}^2$	MA	MR	Reduction (in %)
Hunter-Schmidt	0.2284	0.0291	87.26
Hedges	0.3285	0.2090	36.38
DerSimonian-Laird	0.3087	0.0633	79.50
Sidik-Jonkman	0.3455	0.2318	32.90
ML	0.2800	0.0344	87.73
REML	0.3132	0.0764	75.62
Paule-Mandel	0.3180	0.1421	55.31

Categorical Covariate

```
### load package
load(metafor)
### load BCG vaccine data
data(dat.colditz1994)
### calculate log relative risks and corresponding sampling variances
dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos,
                            di=cneg, data=dat.colditz1994)
### using a model formula to specify the same model
rma(yi, vi, mods = ~ factor(alloc), data=dat, method="REML", btt=c(2,3))
```

Categorical Covariate

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

tau^2 (estimated amount of residual heterogeneity):     0.3615
tau (square root of estimated tau^2 value):             0.6013
I^2 (residual heterogeneity / unaccounted variability): 88.77%
H^2 (unaccounted variability / sampling variability):   8.91
R^2 (amount of heterogeneity accounted for):            0.00%

Test for Residual Heterogeneity:
QE(df = 10) = 132.3676, p-val < .0001

Test of Moderators (coefficient(s) 2,3):
QM(df = 2) = 1.7675, p-val = 0.4132
```

Categorical Covariate

Model Results:

	estimate	se	zval	pval	ci.lb	ci.ub
intrcpt	-0.5180	0.4412	-1.1740	0.2404	-1.3827	0.3468
random	-0.4478	0.5158	-0.8682	0.3853	-1.4588	0.5632
systematic	0.0890	0.5600	0.1590	0.8737	-1.0086	1.1867

Results for type of allocation

Further Applications

- Methods can be easily extended to more than one covariate.
- Meta-regression is primarily used for explaining heterogeneity between study results.
- Meta-regression technique can be also used for other applications of combining results; e.g. combining results from controlled and uncontrolled studies in meta-regression model where the covariate indicates whether the result comes from a controlled or from an uncontrolled study
- Subgroup analyses can be performed, too. In random-effects model, we can assume either a common heterogeneity variance in all subgroups or different heterogeneity variances.