

Systematic reviews, Meta-analysis in R – a practical guide

Clinical Epidemiology - Principles, Methods and Applications

Dr. Tobias Niedermaier

Division of Clinical Epidemiology and Aging Research

German Cancer Research Center (DKFZ)

Winter term 2022/23

February 14, 2023

Overview

- Background
- Systematic reviews
- Key ideas of meta-analysis
- Why naïve pooling is not reasonable
- Forest plot
- Important statistical issues in meta-analysis
- Worked examples
- Diagnostic test studies

Background – Systematic reviews

- Systematic reviews (SRs) aim at retrieving and summarizing all relevant studies on a certain topic
- Unlike narrative reviews, SRs have a clear search strategy in relevant data bases (PubMed, Web of Science, Embase, Scopus, Cochrane library, ...)
- SRs should have prespecified in- and exclusion criteria

Background – Meta-analysis

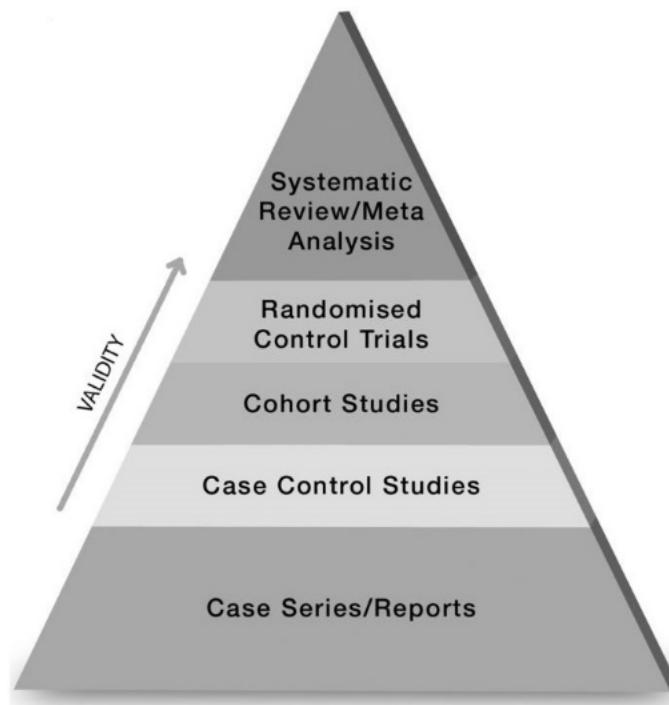
- Meta-analyses are always based on systematic reviews but apply additional steps: results of individual studies are summarized in a quantitative way to come up with a single pooled estimate
- Additional steps only done in meta-analyses: estimation of statistical heterogeneity, including meta-regression (not necessarily), assessment of publication bias, graphical presentation in the form of forest plots or similar plots. (Note: forest plots without summary estimate can also be done in systematic reviews, though it is less common)

Background

- Advantages of systematic reviews and meta-analysis (SRMA) vs. narrative reviews:
 - **Reduce bias** of the reviewer (adhering to prespecified protocols) and in the research studies (quality assessment)
 - **Higher precision** (exhaustive search, potential inclusion of all relevant articles on the research topic)
- Hierarchy of evidence: systematic reviews and meta-analyses are “on the top” compared to other study types

Background

Evidence hierarchy:



However: Meta-analyses of low-quality studies will still be considered low-quality evidence

Background

Steps in a systematic review (I):

- Define research question
- Optional, but recommended:
 - Write a study protocol
 - Register the protocol at PROSPERO (international register for systematic reviews). Why? →
 1. Avoid duplication of effort
 2. Avoid data-driven modification in eligibility criteria
 3. Avoid reporting bias and overstating of data-driven (rather than hypothesis-driven) findings
 4. Avoid publication bias in subsequent meta-analyses

Background

Steps in a systematic review (II):

- Develop a search strategy
- Define in- and exclusion criteria
- Remove duplicate articles if several data bases are searched, then screen unique articles by title and abstract
- Obtain and screen full texts of potentially relevant articles and decide whether to in- or exclude
- Extract data from relevant studies by at least two researchers independently
- Assess study quality / risk of bias
- Summarize results and write manuscript

Background

Protocols for systematic reviews

- Statement paper: Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1. doi: 10.1186/2046-4053-4-1
- Explanation and Elaboration paper: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, the PRISMA-P Group. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647. doi: 10.1136/bmj.g7647

Background

Reporting guidelines

- PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)
- PRISMA-P (PRISMA for review protocols)
- MOOSE (Meta-analyses Of Observational Studies in Epidemiology)
- STARD (Standards for Reporting Diagnostic accuracy studies)
- Others, plus explanations etc.: <https://www.equator-network.org/>
- STROBE – observational studies
- CARE – case reports
- SRQR – qualitative research
- CHEERS – economic evaluations
- AGREE – clinical practice guidelines

First 14 of the 27 PRISMA checklist items

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	

Remaining PRISMA checklist items

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

Background

Protocols for systematic reviews

- Up to the results section of a manuscript, the PRISMA-P checklist items mimic those of the PRISMA checklist items but the focus is on what you are planning to do rather than what you did
- Items regarding the Discussion are missing in PRISMA-P for obvious reasons
- Checklist, statement, and explanation and elaboration available from <http://www.prisma-statement.org/Extensions/Protocols.aspx>

Search strategy – sources

Potential article sources:

- Systematic search:

- PubMed, Web of Science, EMBASE, Scopus, Cochrane Library, PsycINFO, Europe PMC, (Google Scholar)
- Cochrane Controlled Trials Register

- “Manual” search:

- Cross-referencing (citing and cited articles)
- Find articles from the same author/research group
- Abstract books, conference proceedings etc.

Search strategy – search terms

- Having appropriate search terms is the most crucial part of a systematic review, so some time should be spent on it in order to avoid repetitive searching due to previously inappropriate search terms
- Advice: Check search terms of SRs published on similar topics in high-ranking journals
- Try to do a preliminary search, sorted by relevance, obtain a few relevant articles from it, find further relevant articles via cross-referencing, check if preliminary search would have identified those relevant articles, refine if necessary
- Have your search terms checked by your peers

Search strategy – example in PubMed and Web of Science

PubMed – example of the search for diagnostic accuracy studies using fecal immunochemical test:

(FIT[ti/ab] OR iFOBT[ti/ab] OR FOBT[ti/ab] OR stool[ti/ab] OR
fecal[ti/ab] OR faecal[ti/ab] OR occult[ti/ab]) } Stool test

AND (CRC[ti/ab] OR colorect*[ti/ab] OR rectal[ti/ab] OR
rectum[ti/ab] OR colon[ti/ab] OR colonic[ti/ab] OR bowel[ti/ab]) } Colorectal cancer

AND (carcinoma*[ti/ab] OR cancer[ti/ab] OR cancers[ti/ab] OR
cancerous[ti/ab] OR neoplas*[ti/ab] OR adenocarcinoma*[ti/ab] OR
tumor*[ti/ab] OR tumour*[ti/ab]) } Colorectal cancer

AND (sensitiv* OR accura* OR positiv* OR detect* OR miss* OR
specifi*) } Diagnostic accuracy

Search strategy – example in PubMed and Web of Science

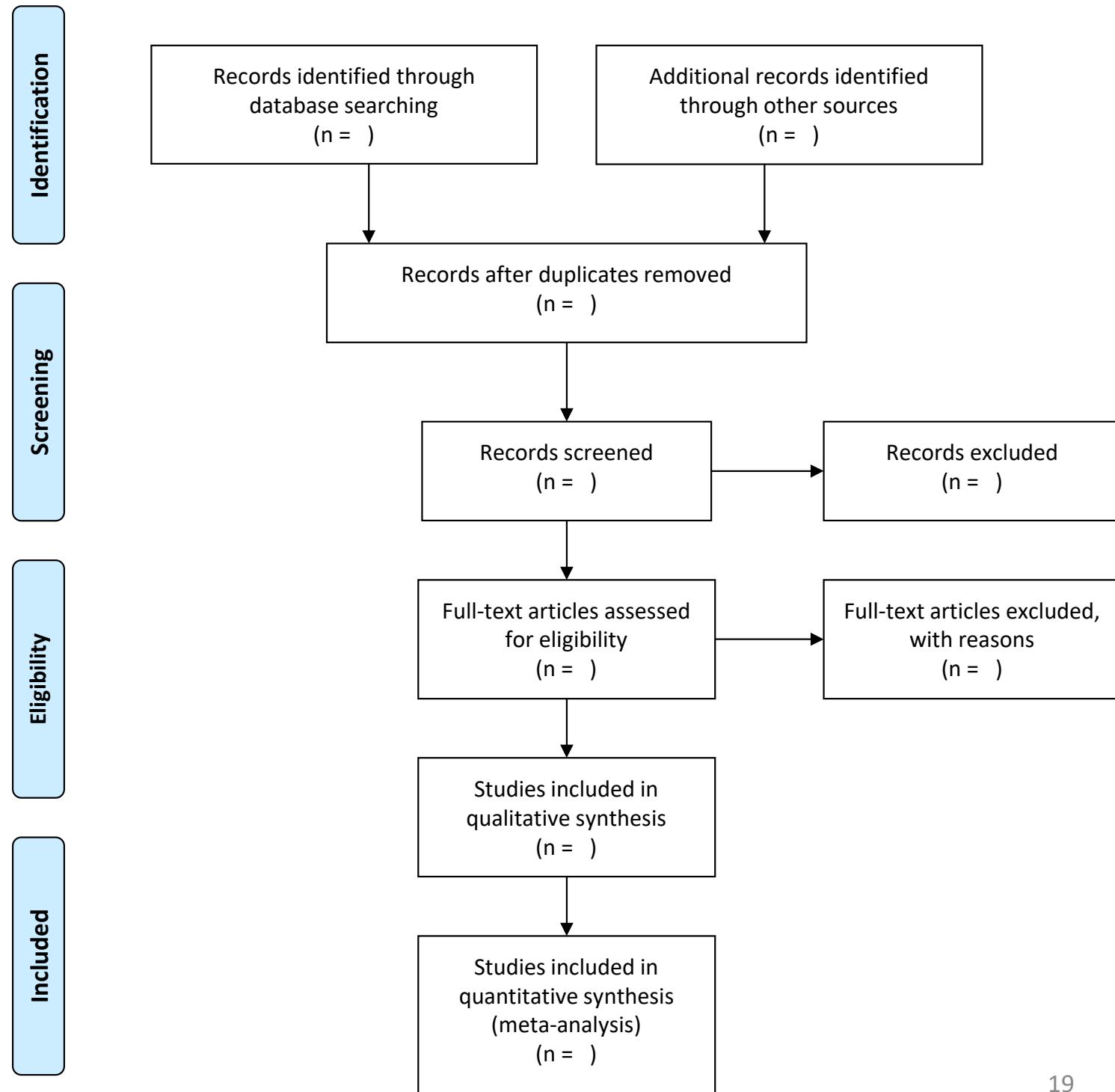
Web of Science:

```
TS=((FIT OR iFOBT OR FOBT OR stool OR fecal OR faecal OR occult)  
AND (CRC OR colorect* OR rectal OR rectum OR colon OR colonic  
OR bowel)  
AND (carcinoma* OR cancer OR cancers OR cancerous OR neoplas*  
OR adenocarcinoma* OR tumor* OR tumour*)  
AND (sensitiv* OR accura* OR positiv* OR detect* OR miss* OR  
specifi*)
```

Search

- Find terms that should appear in all relevant studies, combine them with “AND”
- Find synonyms of terms and combine them with “OR”
- Once you specified the search terms, organize your search:
 - Remove duplicate articles
 - Keep track of excluded studies and “stage” of exclusion (after screening title and/or abstract vs. after full-text screening)
 - Keep track of reasons for excluding studies read in full-text
 - Recommended: Use EndNote or Reference Manager

PRISMA Flow Diagram



Key ideas of meta-analysis

- Check if a meta-analysis actually makes sense.
Counter examples might be:
 - Too much heterogeneity (see later)
 - Similarly: inconsistency regarding exposure and/or outcome definitions / lack of comparability
- Extract effect sizes and measures of precision (variance) for each study

Why naïve pooling is not reasonable

- Pooling can generate an effect where one does not actually exist

(Bravata DM, Olkin I,
Simple pooling versus
combining in
meta-analysis,
Eval Health Prof (2001))

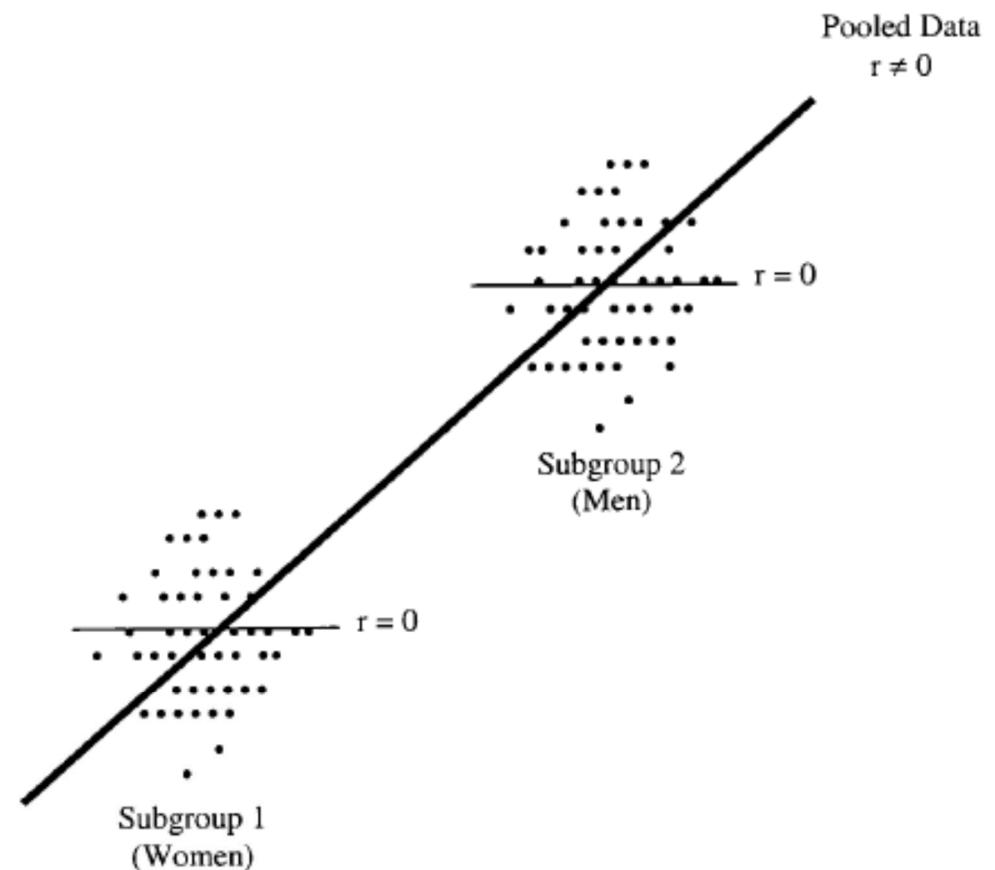


Figure 1: Pooling Finds Correlation Where There Is None

NOTE: Each of the two groups of data represents data from a subgroup (e.g., men and women). The lighter lines show that there are no correlations between x and y for each subgroup ($r = 0$). The dark line shows the correlation that is created by pooling ($r \neq 0$).

Why naïve pooling is not reasonable

- Pooling can obscure effects that do exist within the subgroups

Further reading:

Simpson's paradox

(Bravata DM, Olkin I,
Simple pooling versus
combining in
meta-analysis,
Eval Health Prof (2001))

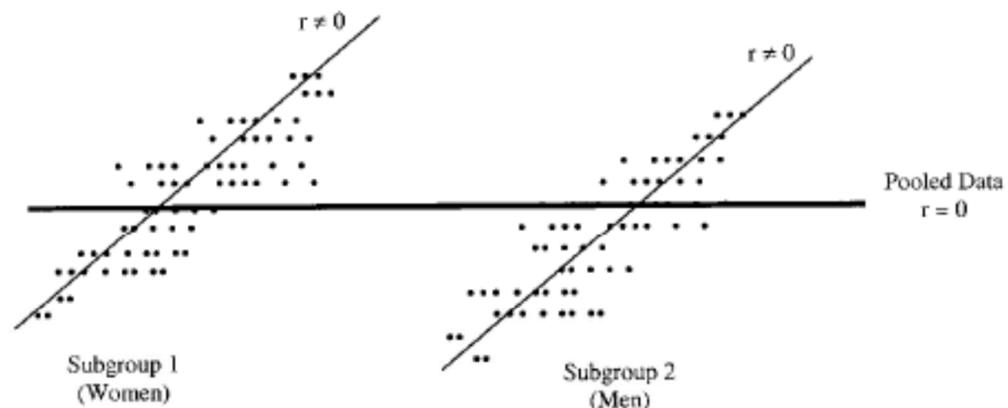
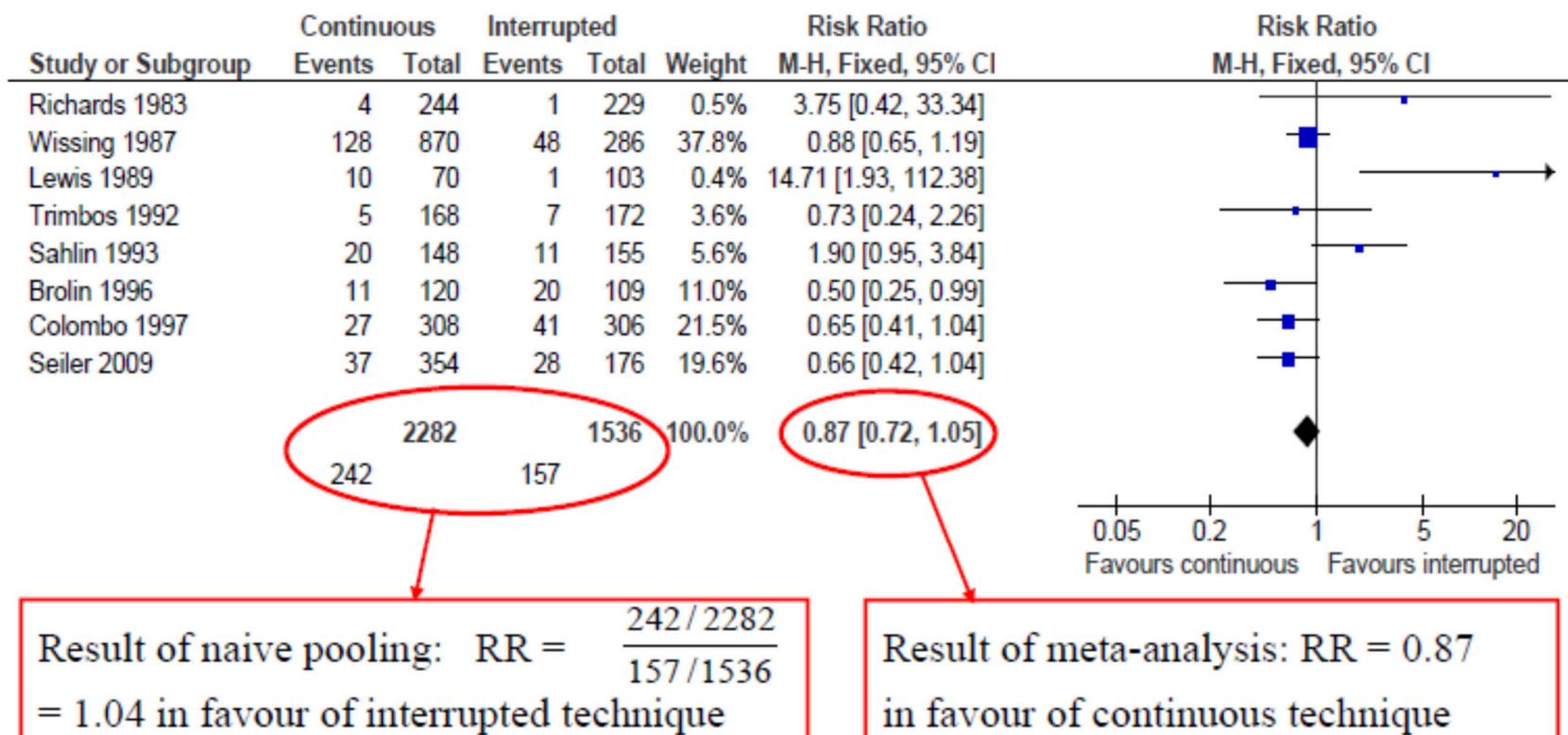


Figure 2: Pooling Obscures the Correlation That Is There

NOTE: Each of the two groups of data represents data from a subgroup (e.g., men and women). The lighter lines show the correlations between x and y for each subgroup ($r \neq 0$). The dark line shows that pooling obscures the correlations that had existed in the subgroups ($r \neq 0$).

Why naïve pooling is not reasonable

Comparison: Continuous versus interrupted suture technique
Outcome: Incisional hernia



Effect measures

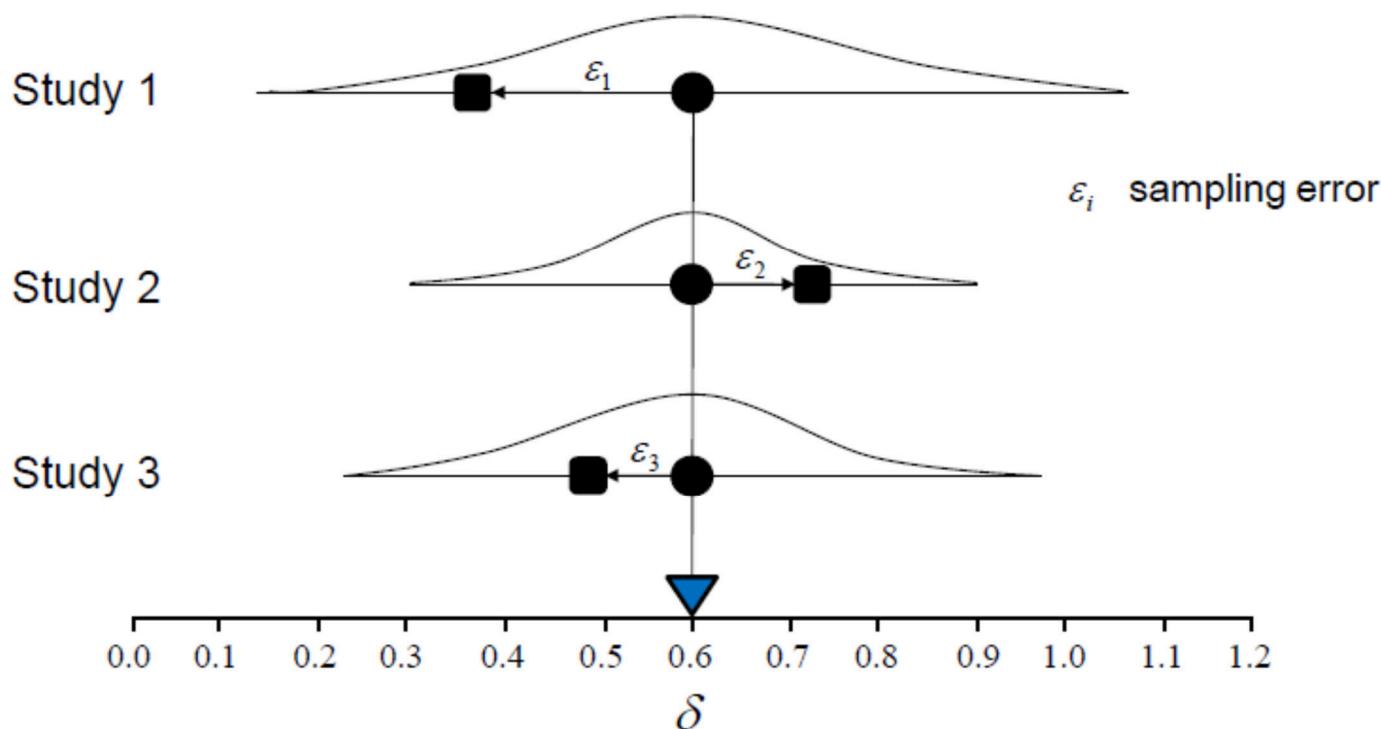
- Wide range of possible measures of effect size, e.g.:
 - For binary data:
 - RD (risk difference)
 - RR (relative risk)
 - OR (odds ratio)
 - For continuous data:
 - Mean difference
 - Standardized mean difference
 - Change in scores
 - Time-to-event data:
 - (log) HR (hazard ratio)

Important statistical issues in meta-analysis

- Fixed-effect and random-effects models
- Statistical heterogeneity
- Publication bias

Fixed-effect models

- FE models assume the presence of a common true effect among all studies and the observed estimates among each individual study



Fixed-effect models

Effect measure observed in study i ($i = 1, \dots, k$): $Y_i = \delta + \epsilon_i$

Within-study variance for study i : V_{Y_i}

Weight assigned to each study: $W_i = \frac{1}{V_{Y_i}}$ (inverse variance)

Weighted mean: $M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i}$

Standard error: $SE_M = \sqrt{V_M} = \sqrt{\frac{1}{\sum_{i=1}^k W_i}}$

Fixed-effect models

95%-confidence interval for the overall effect:

$$[M - 1.96 \cdot SE_M, M + 1.96 \cdot SE_M]$$

Test statistic to test $H_0 : \delta = 0$:

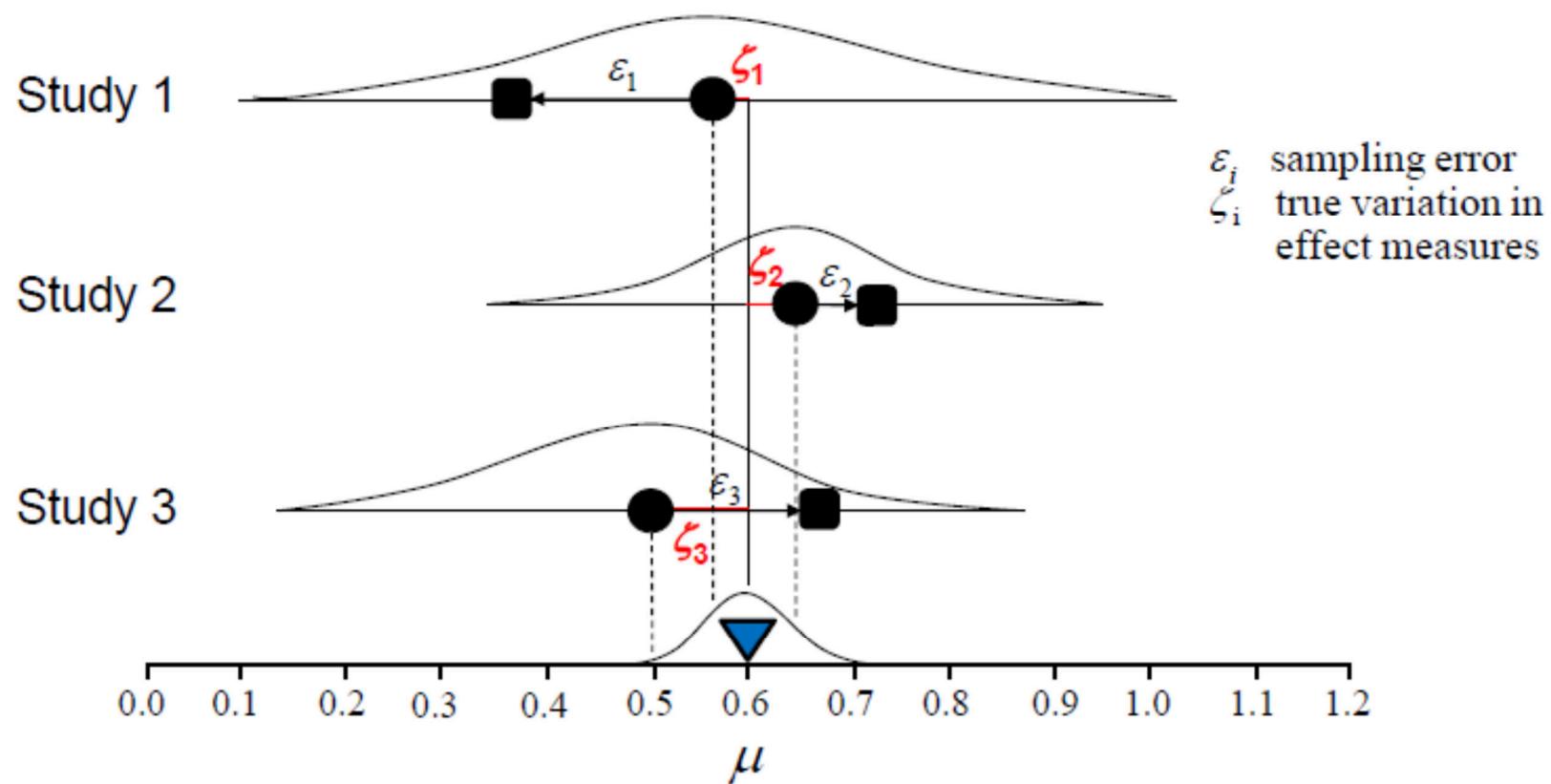
$$Z = \frac{M}{SE_M}$$

For a two-sided test the p -value is given by

$$p = 2 \cdot (1 - \Phi(|Z|))$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution

Random-effects models



Between-study variance and within-study variance

Borenstein et al. (2009)

Random-effects models

“...we not only assume that effects of individual studies deviate from the true intervention effect of all studies due to sampling error, but that there is another source of **variance introduced** by the fact that the **studies do not stem from one single population**, but are drawn from a “universe” of populations.”

- The random-effects-model pays **more attention to small studies** when pooling the overall effect in a meta-analysis (Schwarzer, Carpenter, and Rücker 2015)
- However, smaller studies are commonly thought to be more prone to bias
→ Consider excluding very small studies

*Doing Meta-Analysis in R
A Hands-on Guide*

Mathias Harrer, M.Sc.
Prof. Dr. Pim Cuijpers
Prof. Dr. Toshi A. Furukawa
Assoc. Prof. Dr. David D. Ebert

→ Differences in summary estimates to be expected if many small studies are included and results of those studies differ from large studies

Random-effects models

Effect measure observed in study i ($i = 1, \dots, k$):

$$Y_i = \mu + \zeta_i + \epsilon_i$$

Within-study variance for study i : V_{Y_i}

Between-study variance τ^2 estimated e.g. by the *DerSimonian and Laird* estimator (DerSimonian and Laird, 1986):

$$T^2 = \frac{Q - df}{C} \quad \text{with} \quad df = k - 1,$$

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{\left(\sum_{i=1}^k W_i Y_i\right)^2}{\sum_{i=1}^k W_i} \quad \text{and} \quad C = \sum_{i=1}^k W_i - \frac{\sum_{i=1}^k W_i^2}{\sum_{i=1}^k W_i}$$

Random-effects models

Variance for study i : $V_{Y_i}^* = V_{Y_i} + T^2$

Weight assigned to each study: $W_i^* = \frac{1}{V_{Y_i}^*}$ (inverse variance)

Weighted mean: $M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*}$

Standard error: $SE_{M^*} = \sqrt{V_{M^*}} = \sqrt{\frac{1}{\sum_{i=1}^k W_i^*}}$

Random-effects models

95%-confidence interval for the overall effect:

$$[M^* - 1.96 \cdot SE_{M^*}, M^* + 1.96 \cdot SE_{M^*}]$$

Test statistic to test $H_0 : \mu = 0$:

$$Z^* = \frac{M^*}{SE_{M^*}}$$

For a two-sided test the p-value is given by

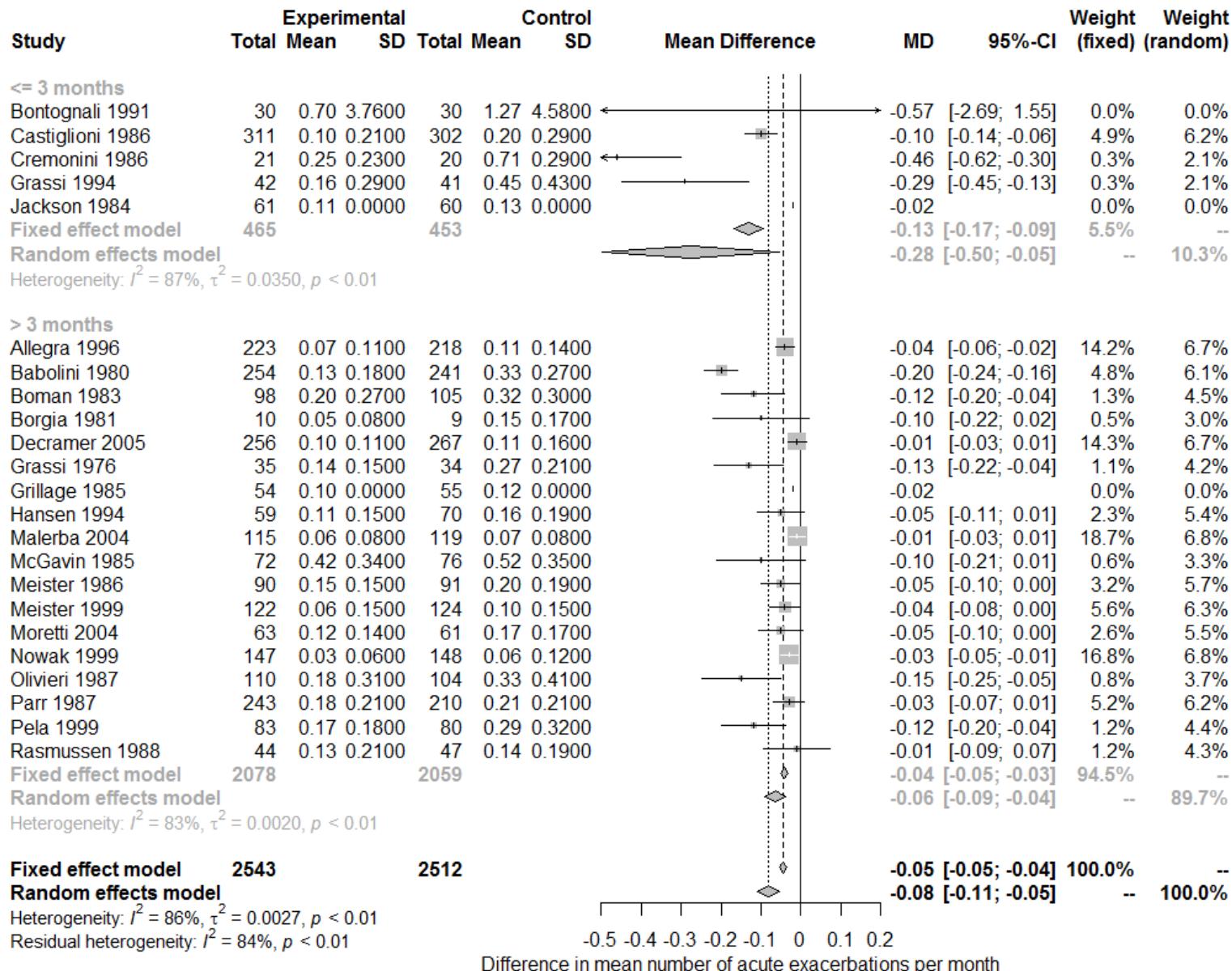
$$p^* = 2 \cdot (1 - \Phi(|Z^*|))$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution

Should I take a fixed or random-effects model?

- Fixed-effects models are appropriate if we have strong reasons to believe that included studies are “functionally identical”. Example: Drug trials within the same country
- Random-effects models are the right choice if we cannot be sure that all studies are “functionally identical/equivalent” (=in the majority of cases)
- Exception: if very few studies → better to use fixed effects if possible, otherwise Bayesian approach

Forest plot



Heterogeneity

- Present if there is more variation in the true effects of the underlying studies than expected by chance
- May be caused by differences in study characteristics (participants, exposure, outcome)
- Can be the result of bias in some studies (e.g. attrition, flawed design)
- Options to explore heterogeneity:
 - Subgroup analyses (only for discrete characteristics)
 - Meta-regression (for discrete or continuous characteristics)

Heterogeneity

- Measure of heterogeneity: Cochran's Q

$$Q = \sum_{i=1}^k W_i(Y_i - M)^2$$

- Q = observed weighted sum of squares
- P values calculated from χ^2 -distribution, (k-1) degrees of freedom (k=number of studies)
- Only the p value is meaningful, not the absolute value of Q
- Relatively low power, thus commonly higher α level (0.1)

Heterogeneity

- Measure to quantify heterogeneity: I^2

$$I^2 = \max\{0, \frac{Q - df}{Q} * 100\}$$

- Q = observed weighted sum of squares
- P values calculated from χ^2 -distribution, (k-1) degrees of freedom (k=number of studies)
- Independence on number of included studies but not on study size
- Relatively low power, thus commonly higher α level (0.1)

Heterogeneity

Guidance on the interpretation of I^2 :

$I^2 = 0\%$: no heterogeneity

$I^2 = 25\%$: low heterogeneity

$I^2 = 50\%$: moderate heterogeneity

$I^2 = 75\%$: high heterogeneity

Higgins et al. (2003)

Heterogeneity

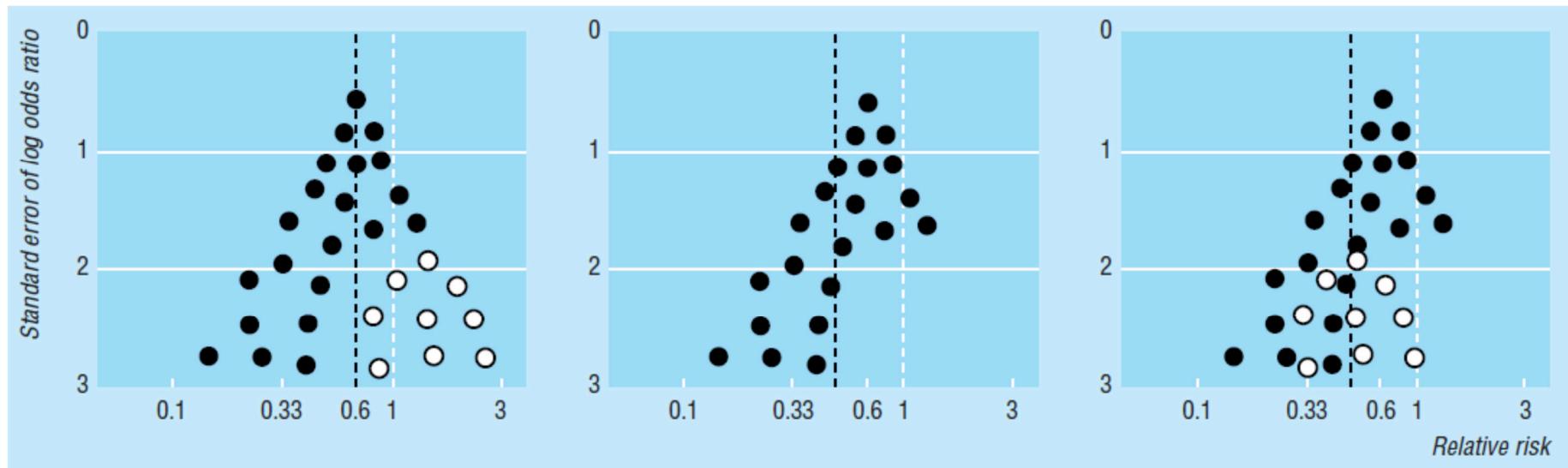
How to address statistical heterogeneity?

- Use a random-effects model which allows for heterogeneity
- Stratify into more homogeneous subgroups if source of heterogeneity could be identified
- Sometimes, a single outlier study contributes to significant heterogeneity → exclude one study at a time and compare results
- Do not perform a meta-analysis but only a systematic or narrative review

Publication bias

- If the probability of results being published depends on the “significance” of estimated effects, results of meta-analyses are biased
- No definite way of telling if publication bias is present, but *funnel plots* provide helpful information
- Variation in effect size expected to be larger in smaller studies, but over- and underestimated effect sizes should be equally frequent in the absence of publication bias (PB)
- “Missing” small studies with null results suggest PB (i.e., funnel plot asymmetry will be present)

Funnel plot



No publication bias

Publication bias present

Presence of bias due to
low methodological quality

- Testing for funnel plot asymmetry
- Begg's test and Egger's test (the latter not recommended for binary outcomes)
- Trim-and-fill method
 - Trim small studies until an unbiased estimate of effect size symmetrically about the mean effect size
 - Fill the original studies back and impute a mirror image for each (in order to correct the variance)

Publication bias – example

- Famous example: Tamiflu (Oseltamivir) from Roche
- Medication is supposed to shorten and reduce symptoms associated with influenza
- Studies were published that suggested associations between Tamiflu intake in influenza patients and shorter hospital stay
- Authors of those studies were all paid by Roche
- Their own further studies which found no or very minor beneficial effects remained unpublished
- Including all studies revealed only very minor benefit of the medication and no evidence of reductions in serious complications

Study quality assessment

- Study Quality Assessment Tools from the NIH:
<https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>
- McMaster University:
https://www.healthevidence.org/documents/our-appraisal-tools/QA_Tool&Dictionary_10Nov16.pdf
- Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses:
http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
- QUADAS-2 for diagnostic accuracy studies:
<https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/>

Example: Meta-analysis in R

- Meta-analyses can be done in various statistical programs but we will focus on R (open source, free of charge, huge number of add-on packages)
- <https://cran.r-project.org/web/views/MetaAnalysis.html>
→ overview of R packages related to meta-analysis
- Widely used: ‘meta’ and ‘metafor’¹ for meta-analysis of RRs, ORs, RDs etc., HSROC or bivariate model ('mada') for diagnostic accuracy studies

¹(>9500 and >6900 citations according to Google Scholar and Web of Science, respectively)

Worked examples using the ‘meta’ package

Example for **continuous** outcome data:

- Get dataset01.csv from <https://www.uniklinik-freiburg.de/imbi/stud-le/meta-analysis-with-r.html>
- Select the appropriate model (fixed or random effects)
- Select the appropriate function (here: **metacont**)
- Understand the meaning of the relevant input parameters / “function arguments” (see next slide)
- Select the appropriate outcome measure (e.g. mean difference)

meta package – function 'metacont'

Parameter	Function
Ne	The number of participants (N) in the intervention group
Me	The mean (M) of the intervention group
Se	The standard deviation (SD) of the intervention group
Nc	The number of participants (N) in the control group
Mc	The mean (M) of the control group
Sc	The standard deviation (SD) of the control group
data=	After '='; paste the name of your dataset here
studlab=paste()	This tells the function where the labels for each study are stored. If you named the spreadsheet columns as advised, this should be studlab=paste(Author)
fixed=	Whether to use a fixed-effects-model
random	Whether to use a random-effects-model
prediction=	Whether to print a prediction interval for the effect of future studies based on present evidence
sm=	The summary measure we want to calculate. We can either calculate the mean difference (MD) or Hedges' g (SMD)

meta package – function 'metacont'

- `m <- metacont(Ne, Me, Se, Nc, Mc, Sc,
studlab=paste(author, year), data=data)`
- `forest(m, xlab="Maximum % fall in FEV1")`

Subgroup analysis:

- `m <- metacont(Ne, Me, Se, Nc, Mc, Sc,
studlab=paste(author, year), data=data,
subgroup=stratvar, print.byvar=FALSE)`
- `forest(m, xlab="Mean difference...")`

meta package – meta-regression

- Basic idea is similar to subgroup analysis but allows use of continuous predictors (compare also: stratification vs. multiple linear/logistic/Cox regression)
- Each covariate should contain at least ten studies
- Random effects model + meta-regression → mixed effects model
- Rely on predefined scientific or theoretical questions
- Avoid “playing around” until you find sth. significant
- https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/metareg.html

meta package – meta-regression

```
require(meta) || (install.packages("meta") && require(meta))
data(Fleiss93cont)
# Add some (fictitious) grouping variables:
Fleiss93cont$age <- c(55, 65, 55, 65, 55)
Fleiss93cont$region <- c("Europe", "Europe", "Asia", "Asia",
"Europe")
m1 <- metacont(n.e, mean.e, sd.e, n.c, mean.c, sd.c,
data = Fleiss93cont, sm = "MD")
mu2 <- update(m1, subgroup = region, tau.common = TRUE, fixed =
FALSE)
metareg(mu2) ; bubble(metareg(mu2))
metareg(mu2, intercept=FALSE)
mu1 <- update(m1, byvar = region)
mu1; forest(mu1)
metareg(mu1, region + age)
```

Example outputs (plots) from the 'meta' package

```
#preparation:  
data(Olkin95)  
  
m1 <- metabin(event.e, n.e, event.c, n.c,  
data = Olkin95, subset = c(41, 47, 51,  
59), studlab = paste(author, year), sm =  
"RR", method = "I")
```

Note:

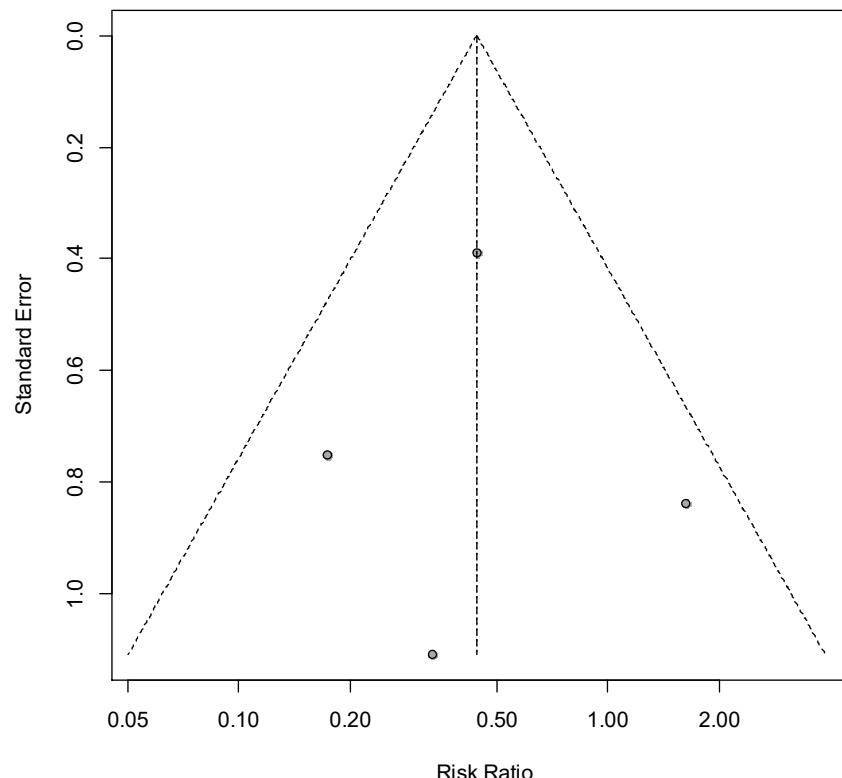
event.e=Number of events in experimental group,
n.e=Number of observations in experimental group,
event.c=Number of events in control group,
n.c=Number of observations in control group

Example output (plot) from the ‘meta’ package

- `forest()` and `funnel()` are customizable functions that are probably the most widely used ones for graphical outputs of meta-analyses

Basic funnel plot:

```
funnel(m1)
```



Worked examples using the ‘metafor’ package (I)

```
rm(list=ls(all=TRUE)) #clean up workspace  
require(metafor) || (install.packages("metafor") &&  
require(metafor))  
#load package  
data(dat.bcg) #load example data set with information on  
allocation and outcome (tpos=treatment, positive outcome,  
tneg=treatment, negative outcome, cpos=control, positive  
outcome, cneg=control, negative outcome)  
str(dat.bcg) #take a look at the variables included in the data  
set  
#ES <- escalc(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg,  
data=dat.bcg) #calculate effect measures (not needed here)  
reml <- rma(method="REML", measure="OR", ai=tpos, bi=tneg,  
ci=cpos, di=cneg, data=dat.bcg) #meta-analysis, random-effects  
model  
summary(reml)
```

Worked examples using the 'metafor' package (II)

#**Meta-regression** (mixed effects model):

```
meml <- rma(method="REML", measure="OR", ai=tpos, bi=tneg,  
ci=cpos, di=cneg, mods=dat.bcg$ablat, data=dat.bcg)
```

summary(meml) #notice the new part „, mods=dat.bcg\$ablat“ which makes it a mixed effects model

- Funnel plot, Egger's regression test & rank correlation test for funnel plot asymmetry:

```
funnel(reml) #Random-Effects Model; funnel(meml) #Mixed-Effects Model;  
regtest(reml) #Regression Test for Funnel Plot Asymmetry
```

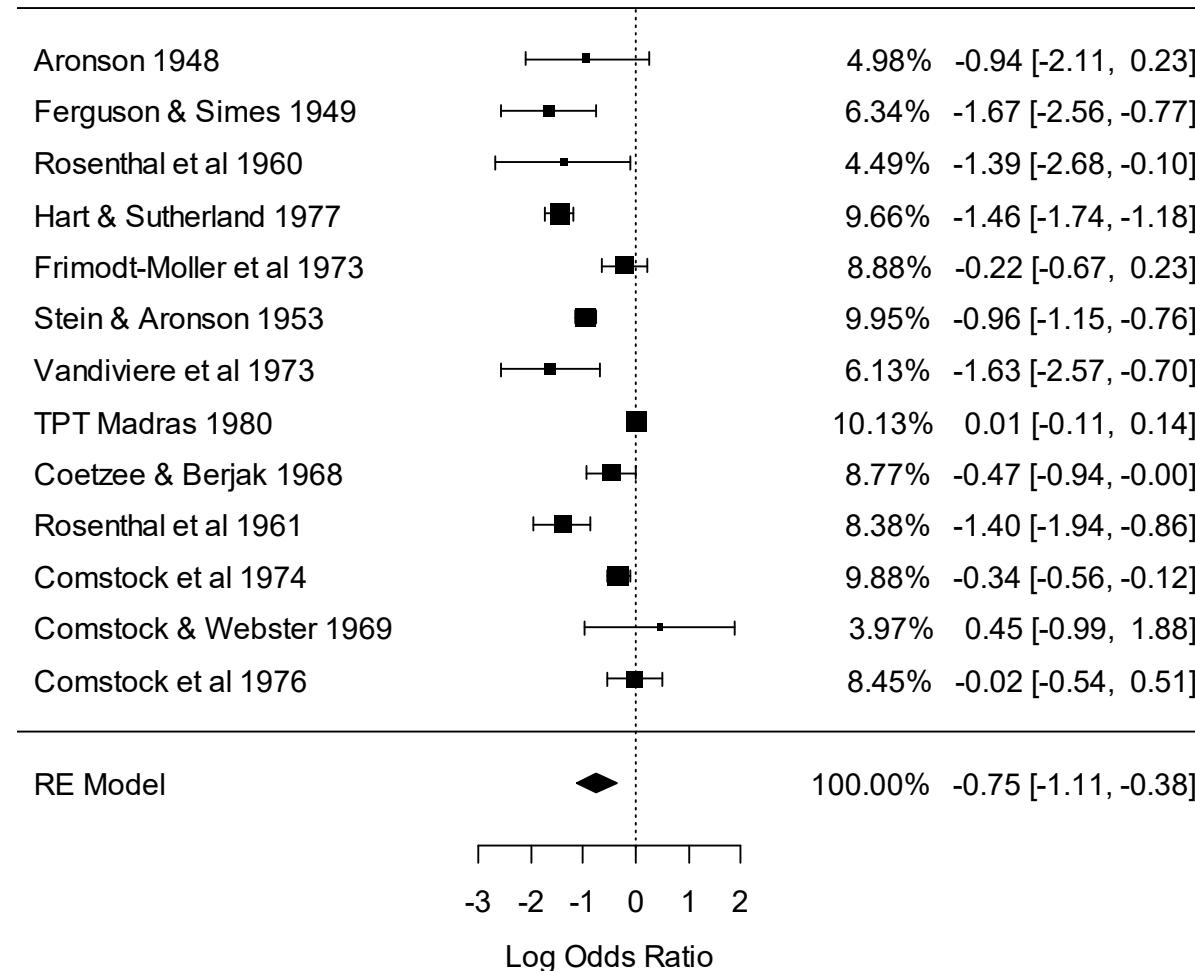
- Estimated influence of publication bias: trimfill(reml)
- leave-one-out meta-analysis: leave1out(reml, trans=exp, digits=3)
- Each study's influence on heterogeneity: baujat(reml)
- normal QQ-Plots: qqnorm(reml, main="Q-Q normal plot")
- Cumulative meta-analysis: forest(cumul(reml,
order=order(dat.bcg\$year)))

Worked examples using the ‘metafor’ package (III)

- Forest plots:

```
forest(reml) #basic forest plot, log scale on x-axis  
  
forest(reml, transf=exp, refline=1) #linear x-axis, i.e. non-symmetrical  
confidence intervals  
  
forest(reml, addcred=T) #add 95% credibility region to summary estimate  
  
forest(reml, order="prec") #order from most precise to least precise  
study  
  
forest(cumul(reml, order=order(dat.bcg$year))) #cumulative meta-  
analysis, ordered by year (pooled estimates after adding new studies one  
by one)  
  
forest(reml, slab=paste(dat.bcg$author, dat.bcg$year)) #add author names  
and publication year  
  
forest(reml, slab=paste(dat.bcg$author, dat.bcg$year), showweights=TRUE)  
#show the weight of each study  
  
#Subgroup analysis: see helppage of function forest.rma, section #### forest plot with  
subgrouping of studies
```

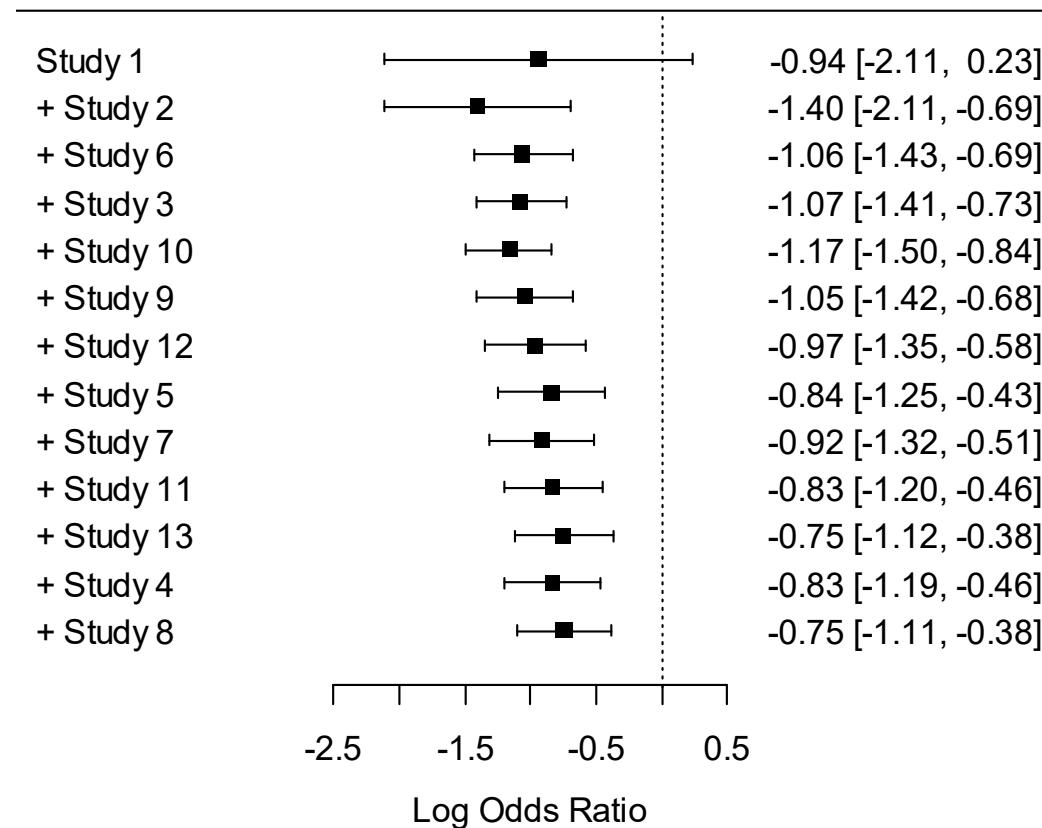
Example outputs from metafor



```
forest(reml, slab=paste(dat.bcg$author, dat.bcg$year), showweights=TRUE)
```

Example outputs from metafor

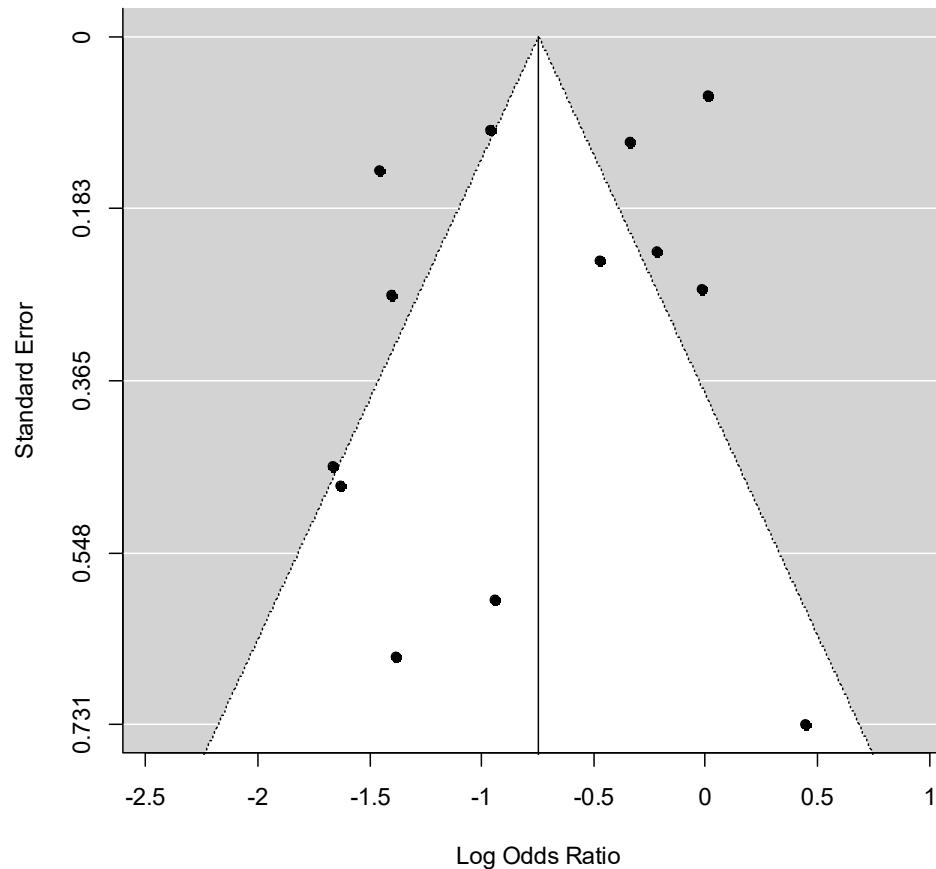
Cumulative meta-analysis:



```
forest(cumul(reml, order=order(dat.bcg$year)))
```

Example outputs from metafor

Normal, “classical”, funnel plot:



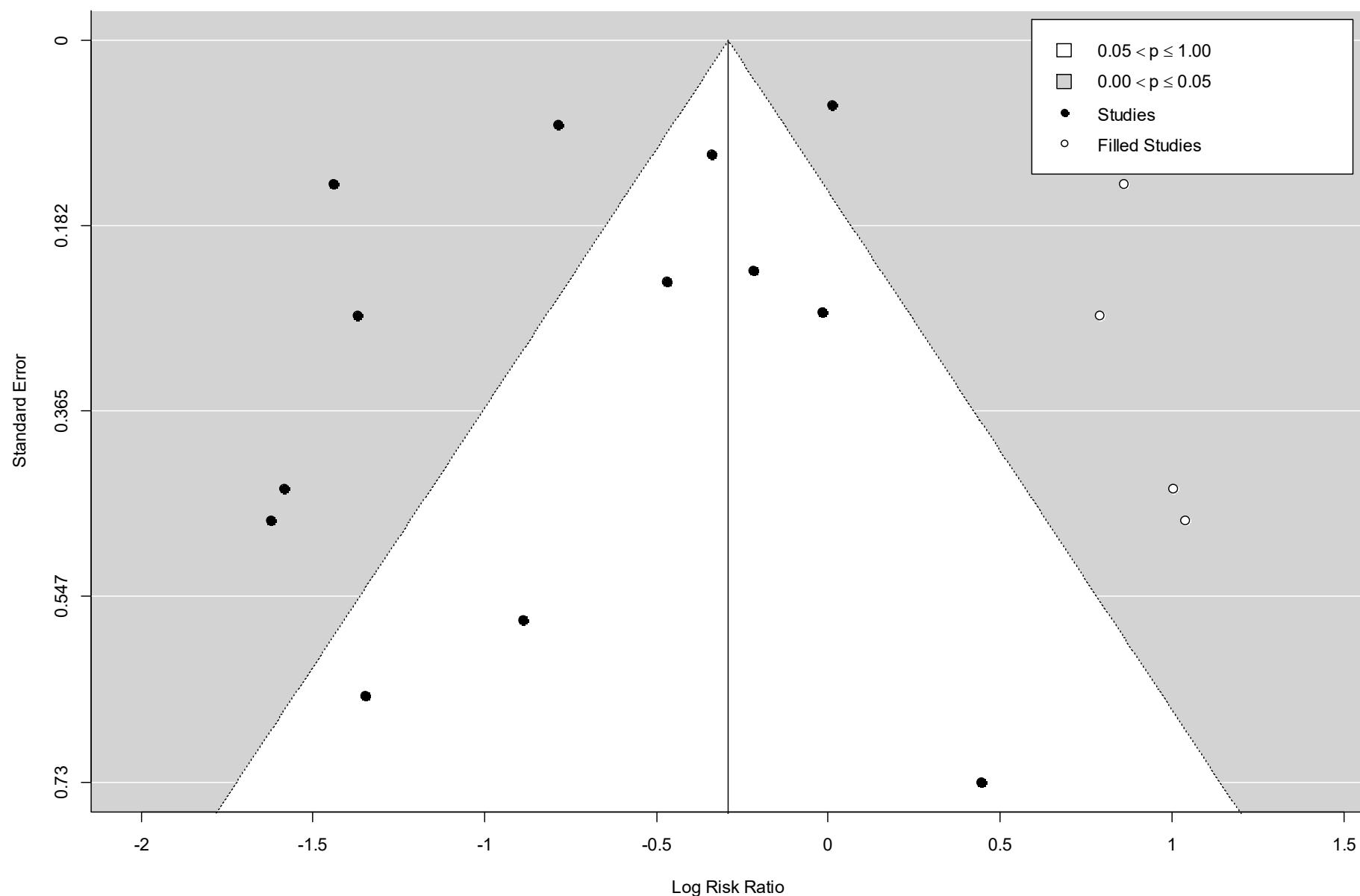
funnel (reml)

Trim-and-fill method

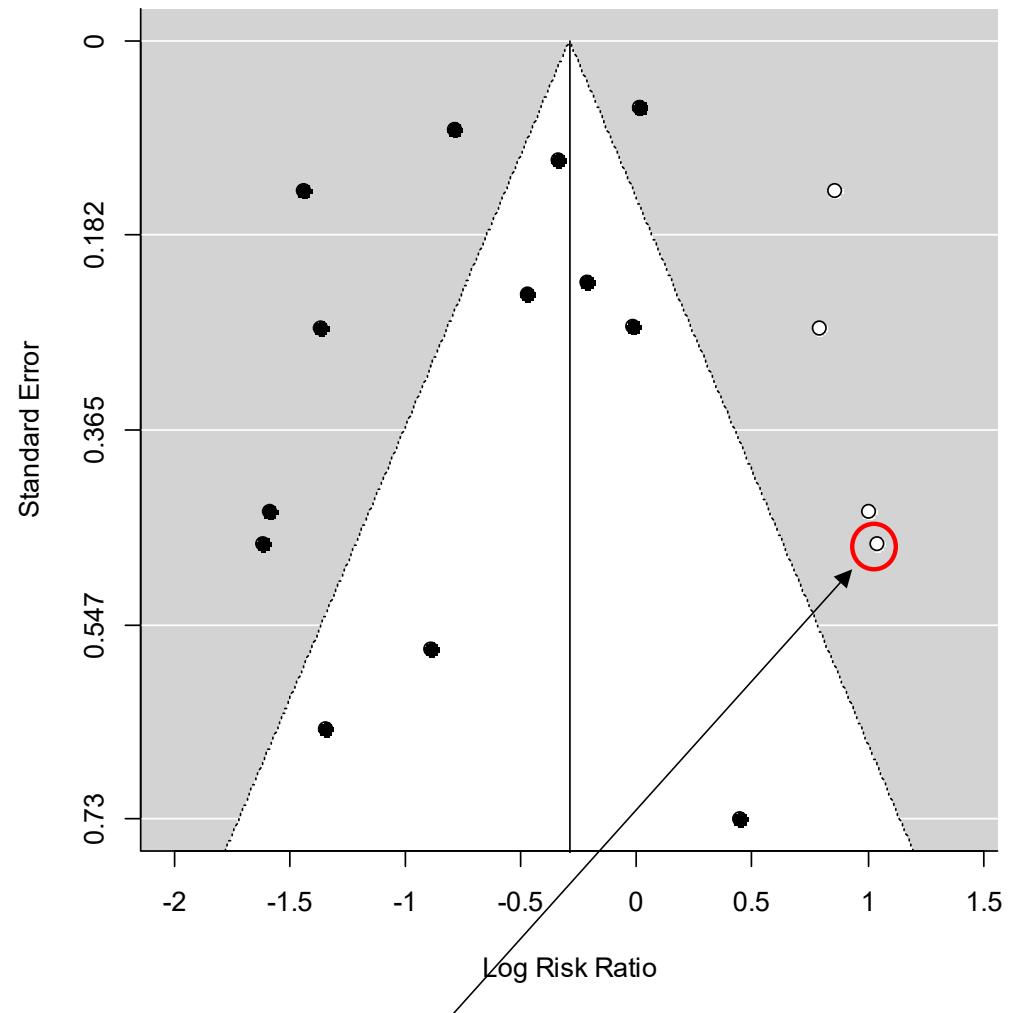
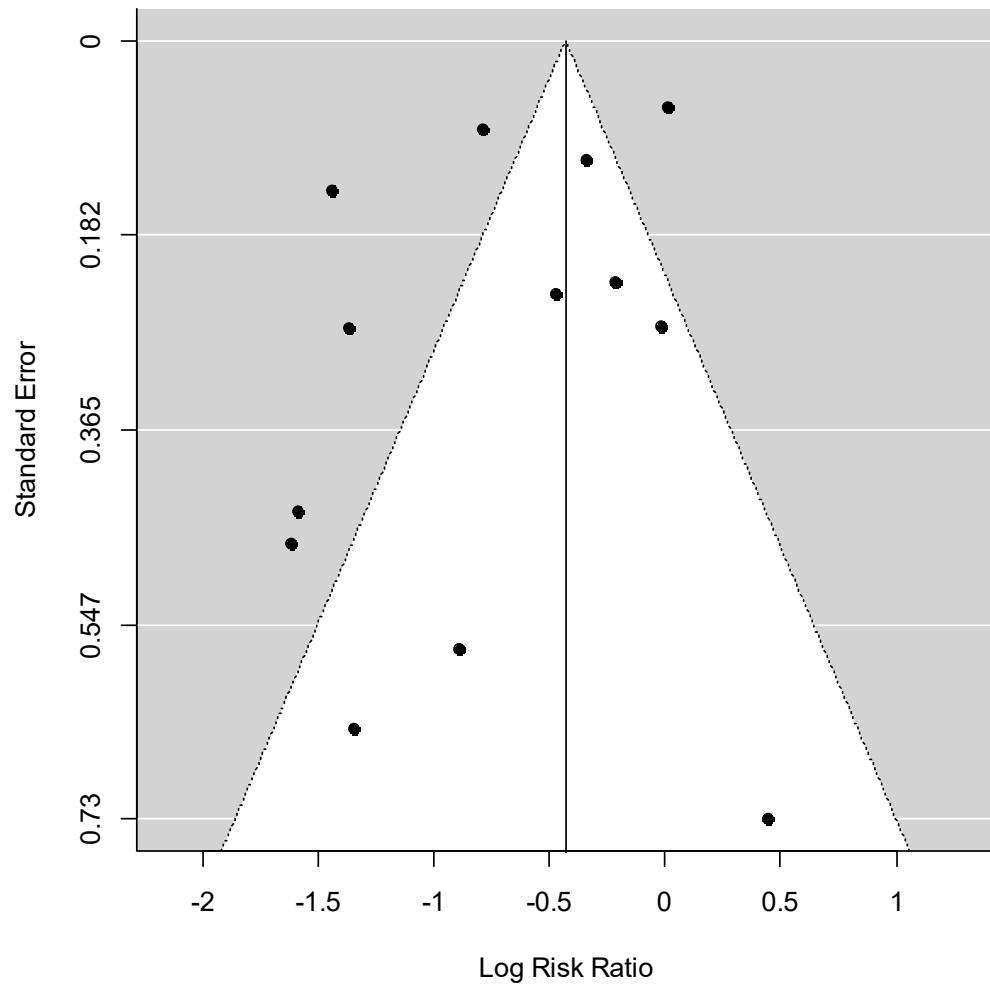
```
res <- rma(measure="RR", ai=tpos, bi=tneg, ci=cpos,  
di=cneg, data=dat.bcg, method="FE")  
  
res.tf <- trimfill(res)  
  
layout(matrix(c(1,2), ncol=2))  
  
funnel(res)  
  
funnel(res.tf, legend=TRUE, cex=1.2)
```

“The method can be used to **estimate the number of studies missing** from a meta-analysis **due to the suppression of the most extreme results on one side** of the funnel plot. The method then augments the observed data so that the funnel plot is more symmetric and recomputes the summary estimate based on the complete data. The trim and fill method can only be used in the context of a fixed- or random-effects model (i.e., in models without moderators). The method **should not be regarded as a way of yielding a more "valid" estimate** of the overall effect or outcome, but as a **way of examining the sensitivity of the results** to one particular selection mechanism (i.e., one particular form of publication bias).” (metafor help page on the function “trimfill”)

Trim-and-fill funnel plot:



Comparison of original and trim-and-fill funnel plot:



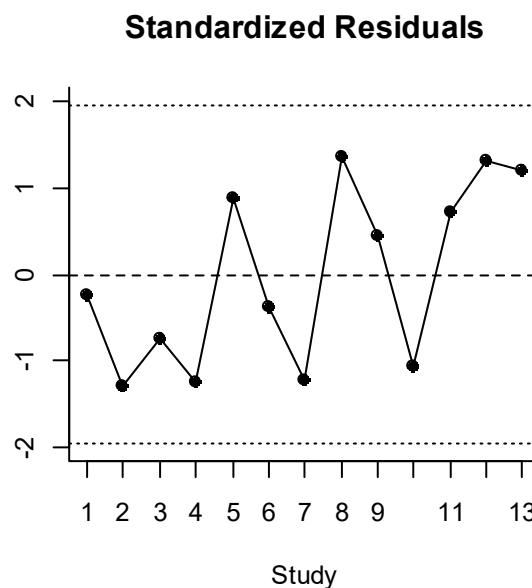
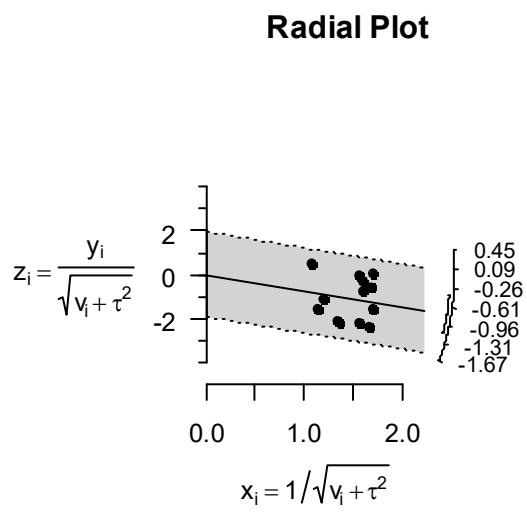
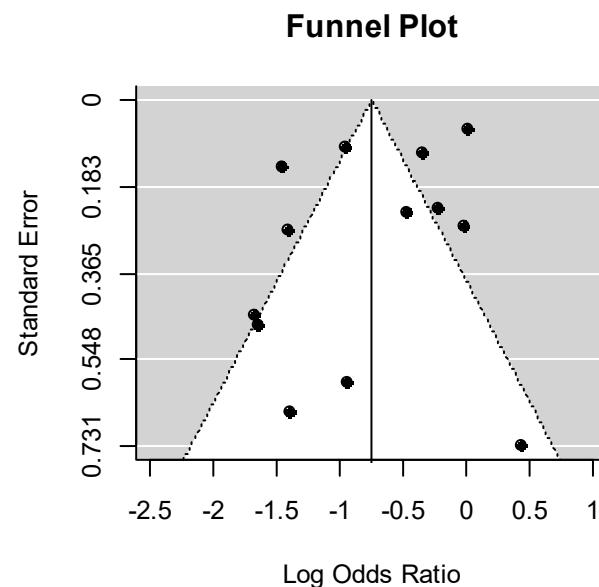
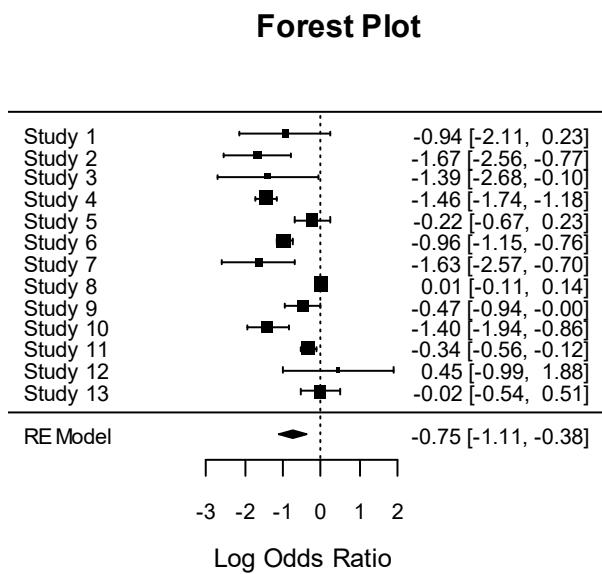
unpublished “mirror image” of a study

Further notes

`plot(reml)`

- Creates a 4-in-1 plot comprising a forest plot, funnel plot, radial plot, and standardized residuals
- Less versatile than individual functions regarding customizations, but gives a good first impression

Result of plot (reml):



Choice of the “method” used with rma.uni() in metafor

- `rma(measure="RR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg, method="FE") #fixed effects`
- `..., method="DL") #DerSimonian-Laird estimator`
- "HE" = Hedges estimator
- "HS" = Hunter-Schmidt estimator
- "SJ" = Sidik-Jonkman estimator
- "ML" = maximum-likelihood estimator
- "REML" = restricted maximum-likelihood estimator
- "EB" = empirical Bayes estimator
- "PM" = Paule-Mandel estimator
- "GENQ" = generalized Q-statistic estimator

Choice of the “method” used with rma.uni() in metafor

- Details can be found in the help page
(<http://127.0.0.1:31466/library/metafor/html/rma.uni.html>)
- Results of the fixed-effects model can be quite different compared to the random-effects models
- Results of the various random-effects models are typically very similar
- Restricted maximum likelihood estimator (REML) and DerSimonian-Laird (DL) are frequently used

Meta-analysis if only RR (or OR) and confidence intervals are known

```
library(meta)

data2 <- data.frame(matrix(nrow=13, ncol=0)) #columns will be added subsequently
data2$Study <- LETTERS[1:13] #just an example, better use real author names here
data2$rr <- c(0.1952, 0.1890, 0.2331, 0.2463, 0.2500, 0.3911, 0.3836, 0.6239,
0.7112, 0.8032, 0.9828, 1.0121, 1.5630)

data2$lcl <- c(0.0774, 0.0797, 0.1766, 0.1450, 0.0707, 0.1280, 0.3272, 0.3917,
0.5717, 0.5155, 0.5821, 0.8946, 0.3740)

data2$ucl <- c(0.4925, 0.4482, 0.3075, 0.4182, 0.8838, 1.1950, 0.4497, 0.9939,
0.8847, 1.2515, 1.6593, 1.1450, 6.5331)

#Compute the log(RR) and its standard error SElog(RR):
data2$log.rr <- log(data2$rr)

data2$se.log.rr <- (log(data2$ucl) - log(data2$lcl)) / (2*1.96)

#Meta-analysis using the metagen-function (generic meta-analysis):
ma2 <- metagen(TE = log.rr, seTE = se.log.rr, studlab = Study, data = data2, sm =
"RR")

forest(ma2) #Forest plot

#See also: https://training.cochrane.org/handbook/current/chapter-06#section-6-3-1
```

Take-home messages

- Follow best-practice when conducting systematic reviews (specify review question, in-/exclusion criteria, search strategy) and meta-analyses (select adequate model, consider heterogeneity, possible bias when pooling observational studies, publication bias)
- Even a perfect analysis plan and adequate methods cannot compensate poor data/biased individual studies (“garbage in – garbage out”)
- My advice: spend some time in finding an interesting and novel research question instead of replicating or updating existing meta-analyses

Thank you for your attention!

Acknowledgements:

PD Dr. Christian Stock and Dr. Katrin Jensen (IMBI Heidelberg) for many of the slides used here.

Contact:

Dr. Tobias Niedermaier

Division of Clinical Epidemiology and Aging Research

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 581

69120 Heidelberg

Tel: +49 (0)6221 42 1312

Email: t.niedermaier@dkfz-heidelberg.de

Useful R packages for meta-analysis

- meta
- metafor
- rmeta
- forestplot
- mada

Check CRAN Task View: Meta-Analysis for packages more specific to a certain type of meta-analysis (uni- or multivariate, network, individual participant data) or specific tasks (graphics, small study bias, meta-regression, simulation etc.)

Add-on: Diagnostic test studies

- Diagnostic tests aim at classifying individuals correctly into those with the target condition (who should be test-positive) and those without the target condition (who should be test-negative)

	D	\bar{D}
P	TP	FP
N	FN	TN

D=disease present

“D dash”=disease absent

P=test positive

N=test negative

TP=true positive, FP=false p.

TN=true neg., FN=false neg.

Diagnostic test studies

- **Sensitivity:** $TP/(TP+FN)$
- **Specificity:** $TN/(TN+FP)$
- Positive predictive value: $TP/(TP+FP)$
- Negative predictive value: $TN/(TN+FN)$
- Diagnostic odds ratio: $(TP/FN)/(FP/TN)$
- Sensitivity and specificity are typically negatively correlated and dependent on the marker cutoff
- Separate pooling of both measures is not appropriate, requiring more sophisticated models

Diagnostic test studies

- Historically: diagnostic odds ratios were pooled like normal odds ratios (fixed or random effects model)
- Popular up-to-date models: HSROC and bivariate model
- Those two models are mathematically equivalent in the absence of covariates (Harbord et al. 2006)
- Requirements:
 - Reasonable number of studies (5 parameters estimated)
 - Numbers of TP, FN, FP, TN available for each study
 - No cells of any study containing 0 (else: continuity corr.)

Dataset preparation (I)

- If available, extract numbers of true positives, true negatives, false positives, false negatives directly
- If not available but sensitivities, specificities and total numbers of cases and non-cases are reported: calculate as follows:

$$TP = \text{sens} * N(\text{cases})$$

$$FN = (1-\text{sens}) * N(\text{cases})$$

$$TN = \text{spec} * N(\text{non-cases})$$

$$FP = (1-\text{spec}) * N(\text{non-cases})$$

First 13 of the 30 STARD checklist items

Section & Topic	No	Item
TITLE OR ABSTRACT	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
<i>Study design</i>	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
<i>Participants</i>	6	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
<i>Test methods</i>	10a	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard

Remaining STARD checklist items

<i>Analysis</i>	14	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
<i>Participants</i>	19	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
<i>Test results</i>	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

Dataset preparation (II)

- Create an excel sheet which looks approximately like this:

First author	Year	TP	FN	FP	TN	strat_var
Smith	2010	204	15	163	2016	1
Johnson	2012	134	6	194	1432	1
Williams	2013	86	20	12	157	1
Brown	2015	353	40	126	1643	0
Jones	2015	117	27	56	736	0
Miller	2017	166	14	120	961	1
Davis	2018	50	7	34	390	0

Analysis in R (I)

- Read in the data set in R:
 - `setwd("X:\\Path\\to\\file")`
 - `dat <- read.csv2("data_extraction_sheet.xlsx", header=TRUE, stringsAsFactors=FALSE)`
- Make sure the relevant variables are recognized as numeric:
 - `dat[, ("TP", "FN", "FP", "TN")] <- as.numeric(as.character(dat[, ("TP", "FN", "FP", "TN")]))`
- If necessary, select relevant subset:
 - `dat_sub1 <- dat[dat$strat_var==1,]`
- Load R package "mada": `library(mada)`

Analysis in R (II)

- Show descriptive data and double-check with original publication:

- `(x <- madad(dat, correction=0, method="clopper-pearson"))`
- `crosshair(dat, correction=0, method="clopper-pearson")`
- `ROCellipse(x, correction=0, method="clopper-pearson")`
- `forest(x, type="sens")`
- `forest(x, type="spec")`

More information on different CIs of proportions:

<https://towardsdatascience.com/five-confidence-intervals-for-proportions-that-you-should-know-about-7ff5484c024f>

Analysis in R (III)

- Perform meta-analysis:

```
reitsma_model <- reitsma(dat)

plot(reitsma_model, main="SROC curve, summary estimate and
confidence\nregion of the bivariate Reitsma model")

points(fpr(dat), sens(dat), pch=2, cex=0.5)

legend("bottomright", c("data", "summary estimate"), pch =
c(2,1))

legend("bottomleft", c("SROC", "conf. region"), lwd = c(2,1))

summary(reitsma_model)
```

Analysis in R (IV) - Result of summary(reitsma_model):

Call: reitsma.default(data = AuditC)

Bivariate diagnostic random-effects meta-analysis

Estimation method: REML

Fixed-effects coefficients

	Estimate	Std. Error	z	Pr(> z)	95%ci.lb	95%ci.ub	
tsens.(Intercept)	2.100	0.338	6.215	0.000	1.438	2.762	***
tfpr.(Intercept)	-1.264	0.174	-7.249	0.000	-1.605	-0.922	***
sensitivity	0.891	-	-	-	0.808	0.941	
false pos. rate	0.220	-	-	-	0.167	0.285	

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

Variance components: between-studies Std. Dev and correlation matrix

	Std.	Dev	tsens	tfpr
tsens	1.175	1.000	.	
tfpr	0.638	0.854	1.000	

logLik	AIC	BIC
31.564	-53.128	-46.467

AUC: 0.887

Partial AUC (restricted to observed FPRs and normalized): 0.861

HSROC parameters

Theta	Lambda	beta	sigma2theta	sigma2alpha
-0.083	3.262	-0.610	0.695	0.218

Analysis in R (V) – Explanation of the output

- Row “sensitivity”, column “Estimate”: Point estimate for pooled sensitivity
- Row “sensitivity”, column “95%ci.lb”: lower bound of 95% confidence interval for pooled sensitivity
- Row “sensitivity”, column “95%ci.ub”: upper bound of 95% confidence interval for pooled sensitivity
- Row “false pos. rate”: equals 1-specificity
- Variance components: typically not necessary
- AUC: overall accuracy when pooling all studies
- pAUC: same but restricted to 100-80% specificity
- HSROC parameters: typically not interesting

R packages suitable for meta-analysis of diagnostic tests

- **mada** (bivariate Reitsma model incl. meta-regression, *Journal of Clinical Epidemiology*, 58, 982–990)
- **Metatron** (bivariate Reitsma model)
- **Metamisc** (Riley model, *Biostatistics* 2008; 9: 172–186)
- **bamdit** (Bayesian bivariate model, *Statistics in Medicine*. 29(30):3088-102)
- **meta4diag** (Bayesian bivariate model, *Statistics in Medicine* 36(19): 3039–3058)
- **CopulaREMADA** (Copula Mixed Models, doi:10.1002/sim.6595)
- **diagmeta** (doi:10.1186/s12874-016-0196-1)
- **HSROC** (HSROC, *Statistics in Medicine*, 20, 2865–2884)
- **CopulaDTA** (Copula Based Bivariate Beta-Binomial Model, doi:10.18637/jss.v082.c01)

Worked examples for selected packages (I)

```
#1.) mada

library(mada)

data(Dementia)
Dementia

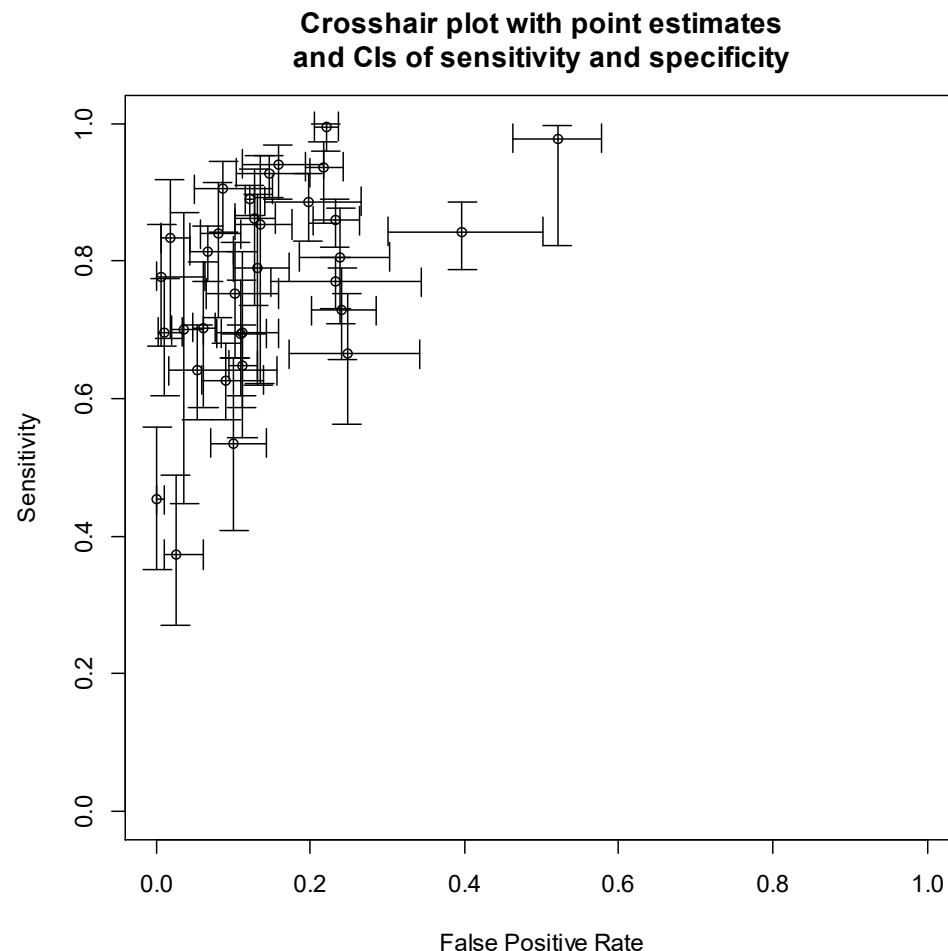
summary(reitsma(Dementia)) [1]
summary(reitsma(Dementia, correction.control="single", correct=0.0001)) [1]

#Plotting:

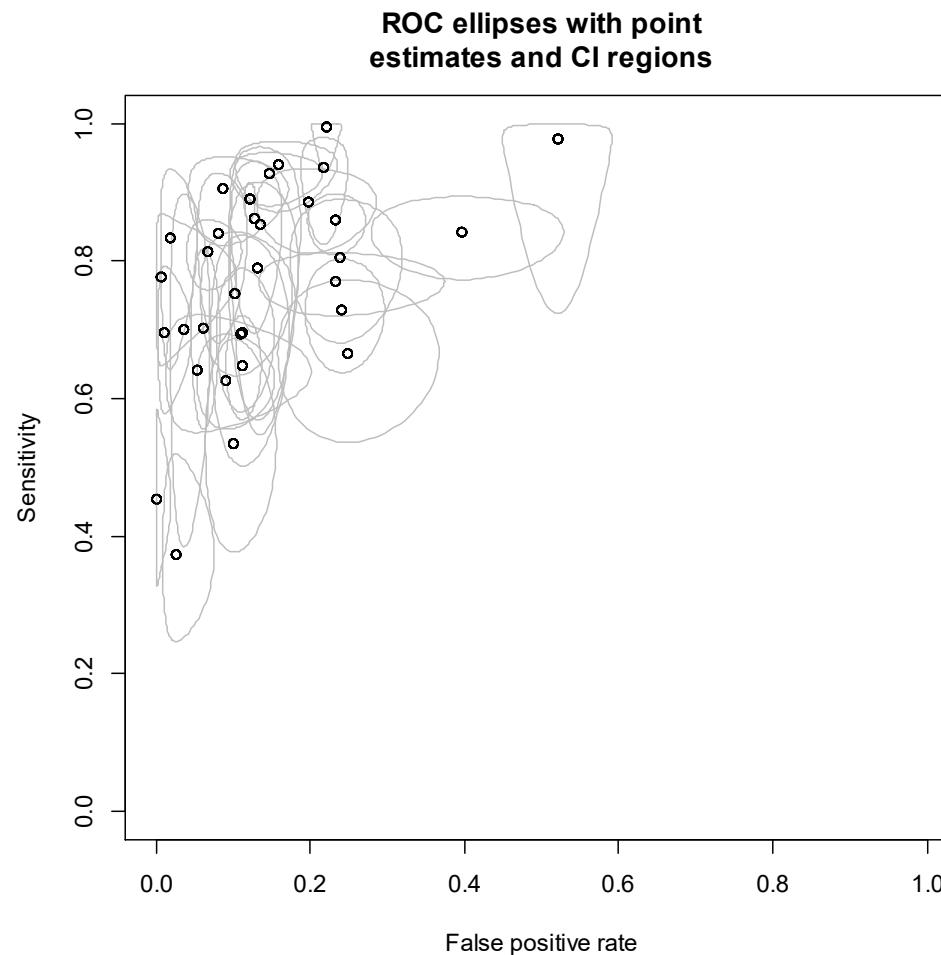
crosshair(Dementia)
ROCellipse(Dementia, pch=". ", cex=3)
reitsma_model <- reitsma(Dementia, correction.control="single", correct=0.0001)
plot(reitsma_model, main="SROC curve, summary estimate and confidence\nregion of the bivariate Reitsma model")
points(fpr(Dementia), sens(Dementia), pch=2, cex=0.5)
legend("bottomright", c("data", "summary estimate"), pch = c(2,1))
legend("bottomleft", c("SROC", "conf. region"), lwd = c(2,1))
forest(madad(Dementia, correction.control="single", correction=0.0001), type="sens", snames=c(LETTERS[1:26],
letters[1:7]), main="Sensitivity")

forest(madad(Dementia, correction.control="single", correction=0.0001), type="spec", snames=c(LETTERS[1:26],
letters[1:7]), main="Specificity")
```

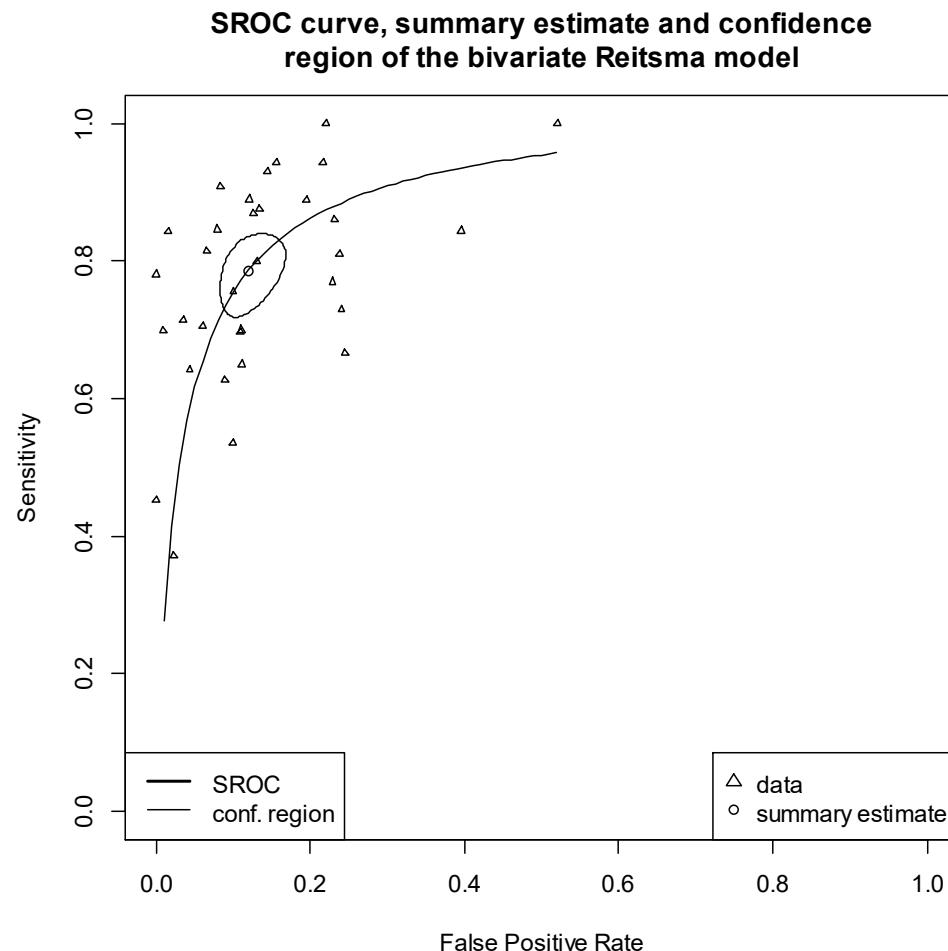
Example output ,mada' package (I)



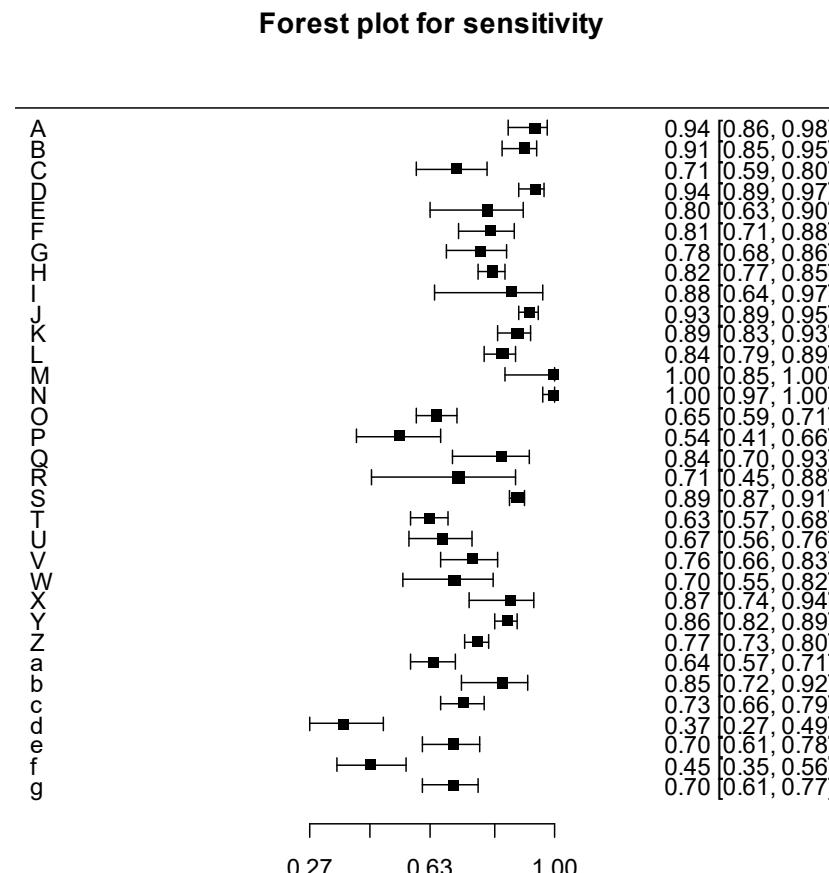
Example output ,mada' package (II)



Example output ,mada' package (III)



Example output ,mada' package (IV)



Worked examples for selected packages (II)

#2.) Metatron

```
library(Metatron)

Dementia$study_id <- c(LETTERS, letters[1:7])

x <- fit.bivar(TP=TP, FN=FN, TN=TN, FP=FP, study=study_id,
data=Dementia)

summary(x)

#coefficient logit sensitivity: e.g. 1.427642
#coef. logit specificity: e.g. 2.198466

#https://stat.ethz.ch/R-manual/R-devel/library/boot/html/inv.logit.html

#The inverse logit is defined by exp(x) / (1+exp(x))

#https://www.ncbi.nlm.nih.gov/books/NBK115744/

→sens: 80.7%, spec: 90.0%. 95%-CIs: -/+ 1.96* std. error
```

Worked examples for selected packages (III)

#3.) metamisc

```
library(metamisc)

summary(riley(Dementia))

plot(riley(Dementia, slab=c(LETTERS[1:26], letters[1:7])))
#forest plots

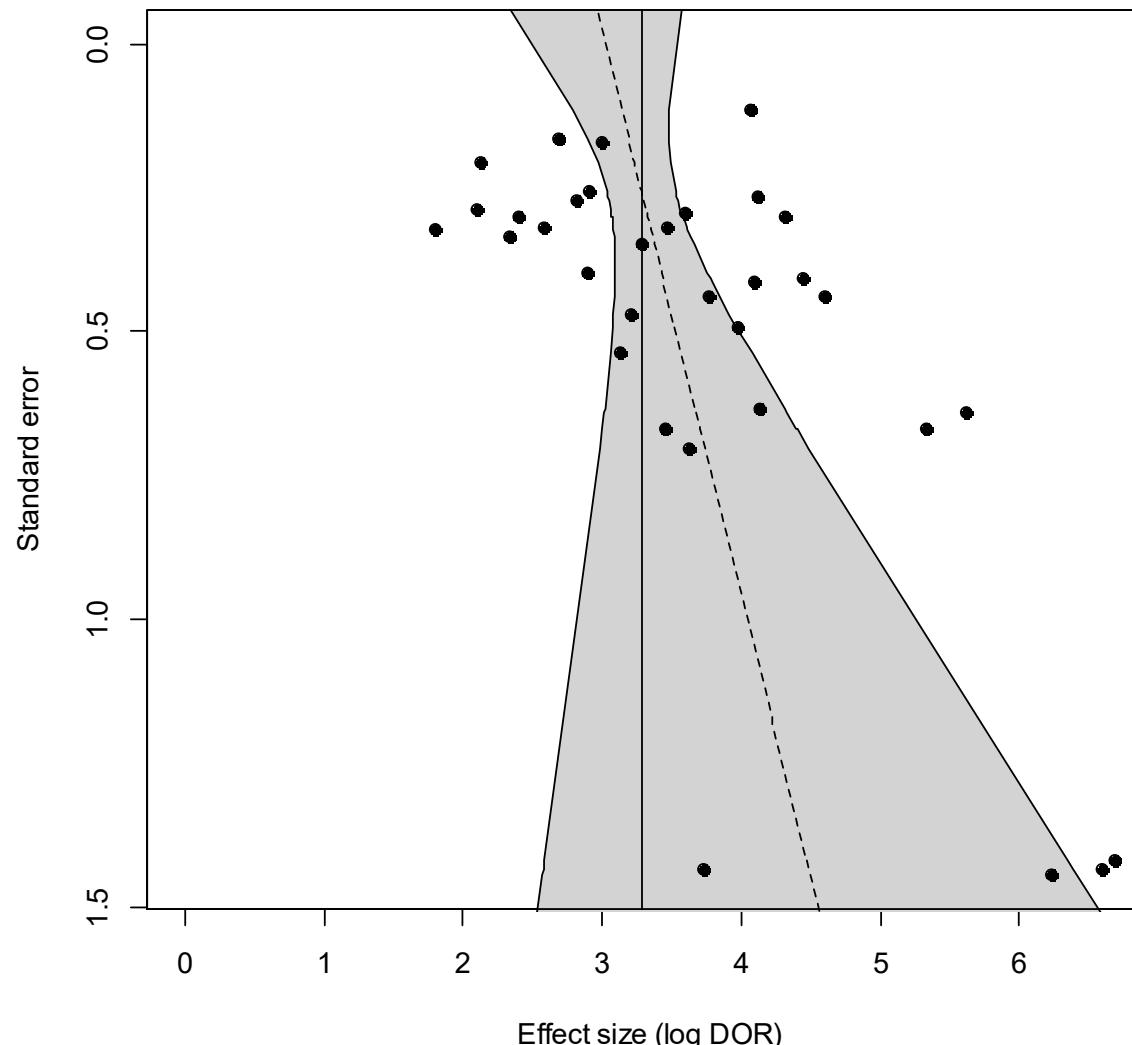
#funnel plot asymmetry test:

(fat_Dementia <- fat(log(mada::madad(Dementia)$DOR$DOR),
madad(Dementia)$DOR$se.lnDOR)) Default: "E-FIV" (Egger's test
with multiplicative dispersion)

plot(fat_Dementia)

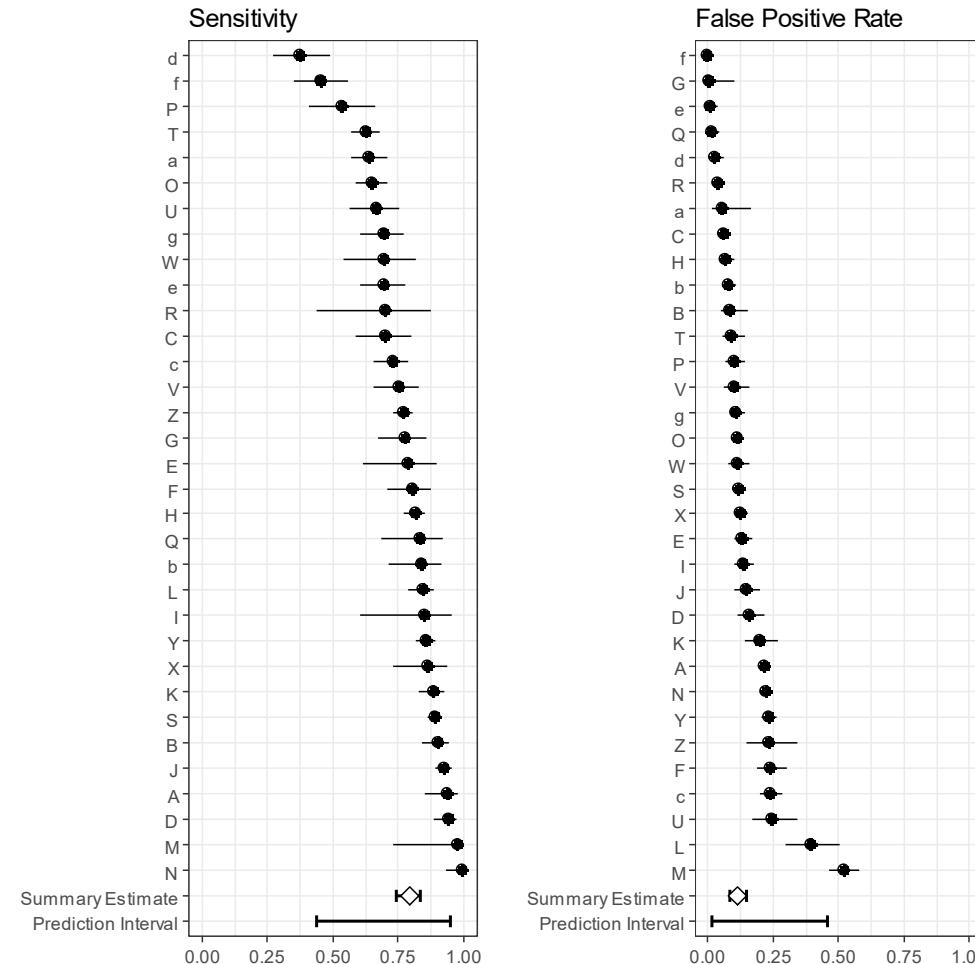
plot(fat_Dementia, confint = TRUE, confint.level = 0.1,
confint.col = "lightgrey", confint.density = NULL, xlab =
"Effect size (log DOR)")
```

Plot of $\log(\text{diagnostic odds ratio})$ vs. standard error (publication bias assessment)



Example output ,metamisc' package

```
plot(riley())
```



Worked examples for selected packages (IV)

#4.) bamdit

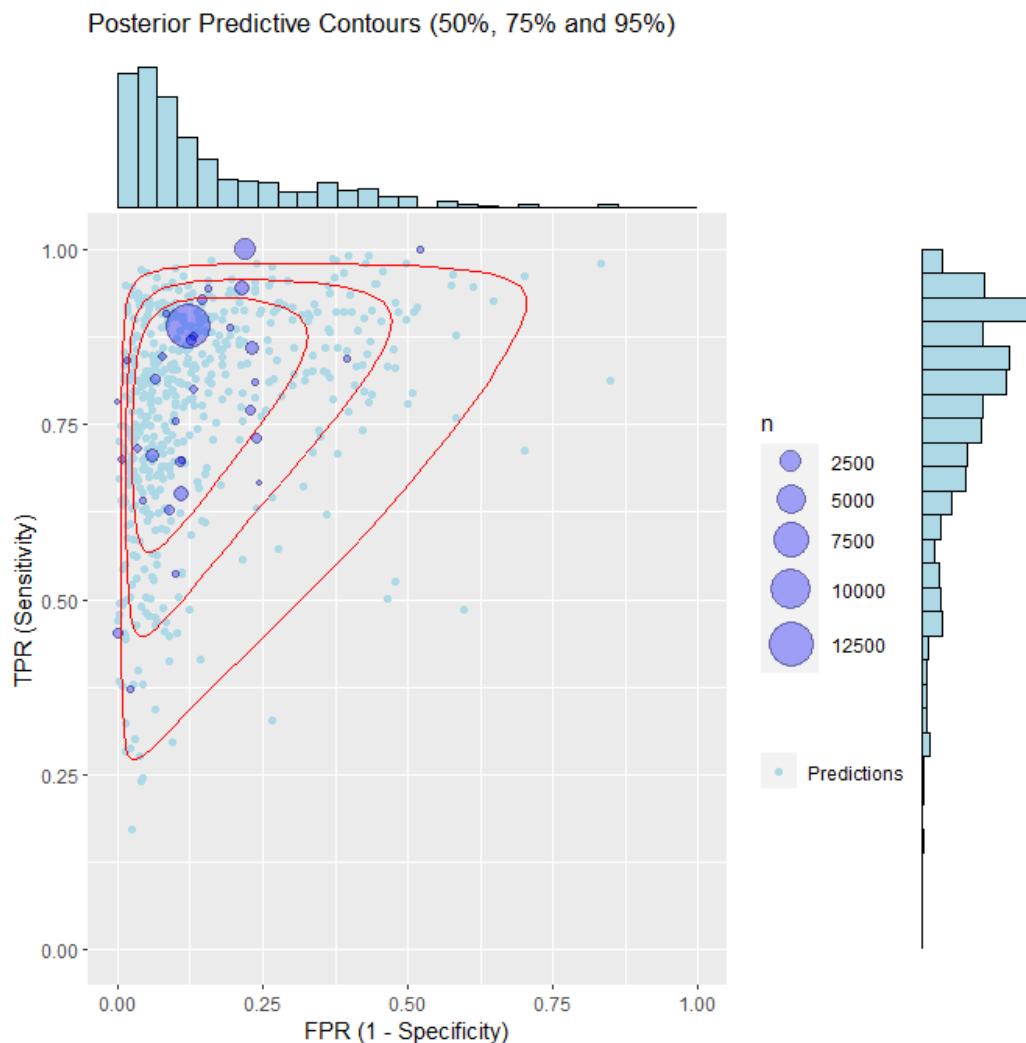
```
library(bamdit)

res <- metadiag(Dementia, two.by.two=TRUE, re.model="SeSp")

summary(res)
plot(res)
res[]
```

Example output ,bamdit' package

```
plot(metadiag(Dementia  
, two.by.two=TRUE,  
re.model="SeSp"))
```



Worked examples for selected packages (V)

#5.) meta4diag

```
library(meta4diag)

#install.packages("INLA", repos=cgetOption("repos"),
INLA="https://inla.r-inla-download.org/R/stable"),
dep=TRUE)

library(INLA)

res <- meta4diag(data=Dementia)

summary(res)

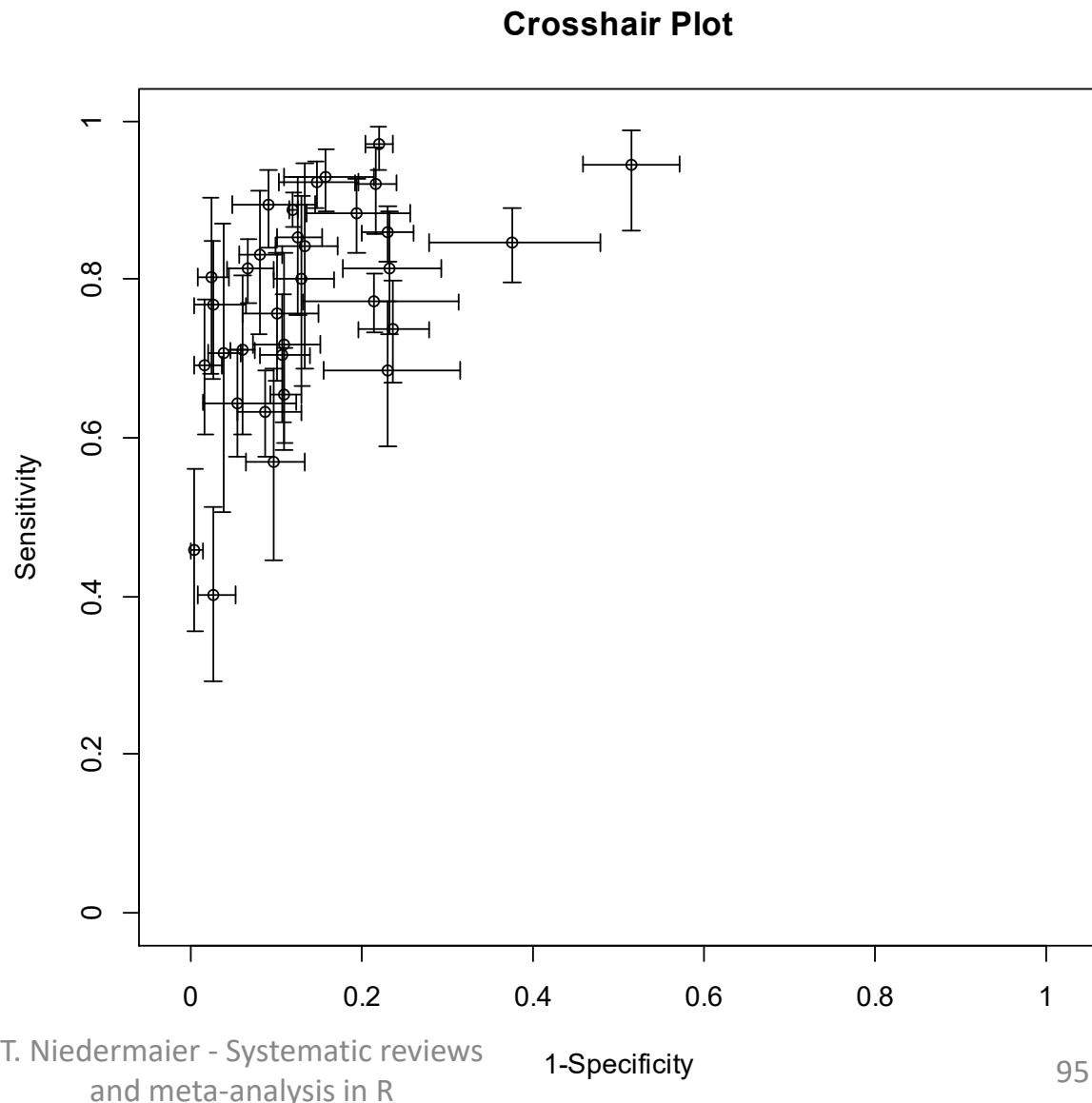
outdata = makeData(data=Dementia, model.type=1)

outpriors = makePriors(var.prior = "invgamma", cor.prior
= "normal",
var.par = c(0.25, 0.025), cor.par = c(0, 5))

model <- runModel(outdata, outpriors, link = "logit",
quantiles = c(0.025, 0.5, 0.975), verbose = FALSE)
```

Example output ,meat4diag' package

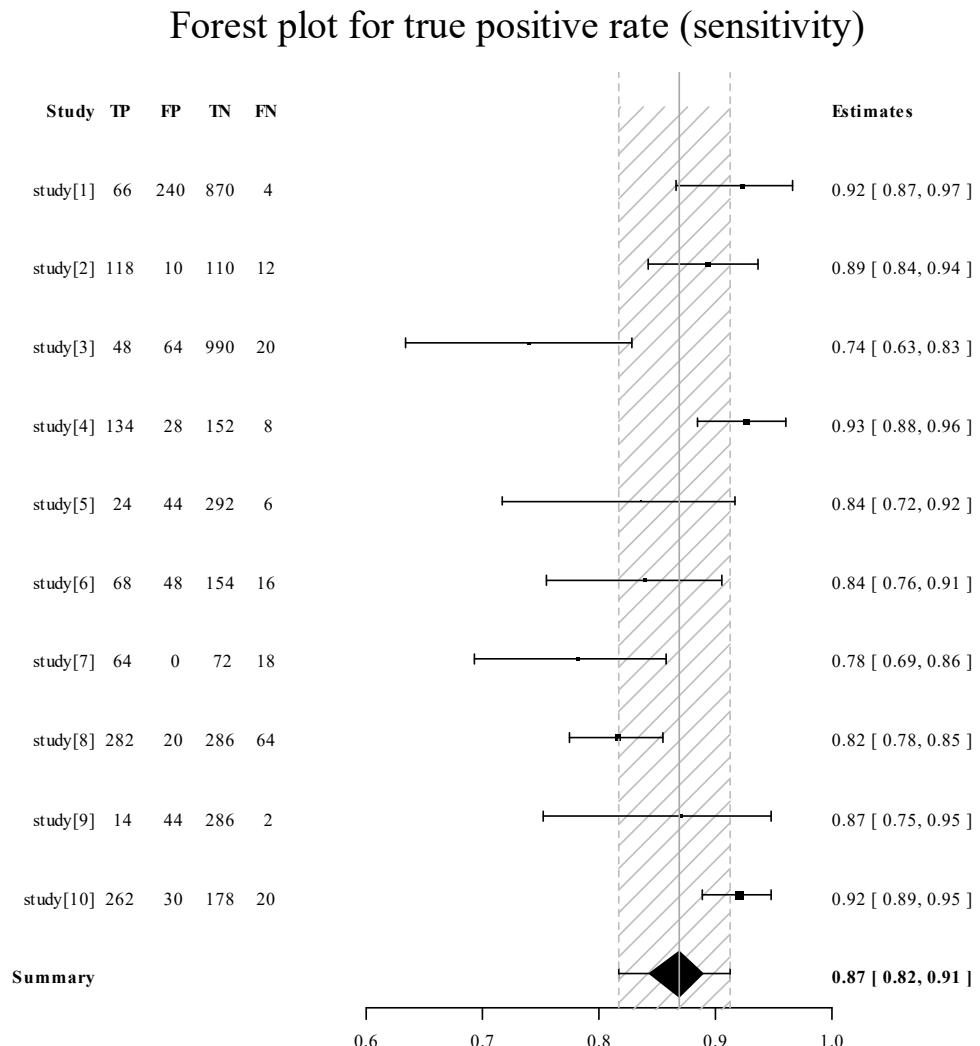
```
crosshair(res)
```



Example output ,meat4diag' package

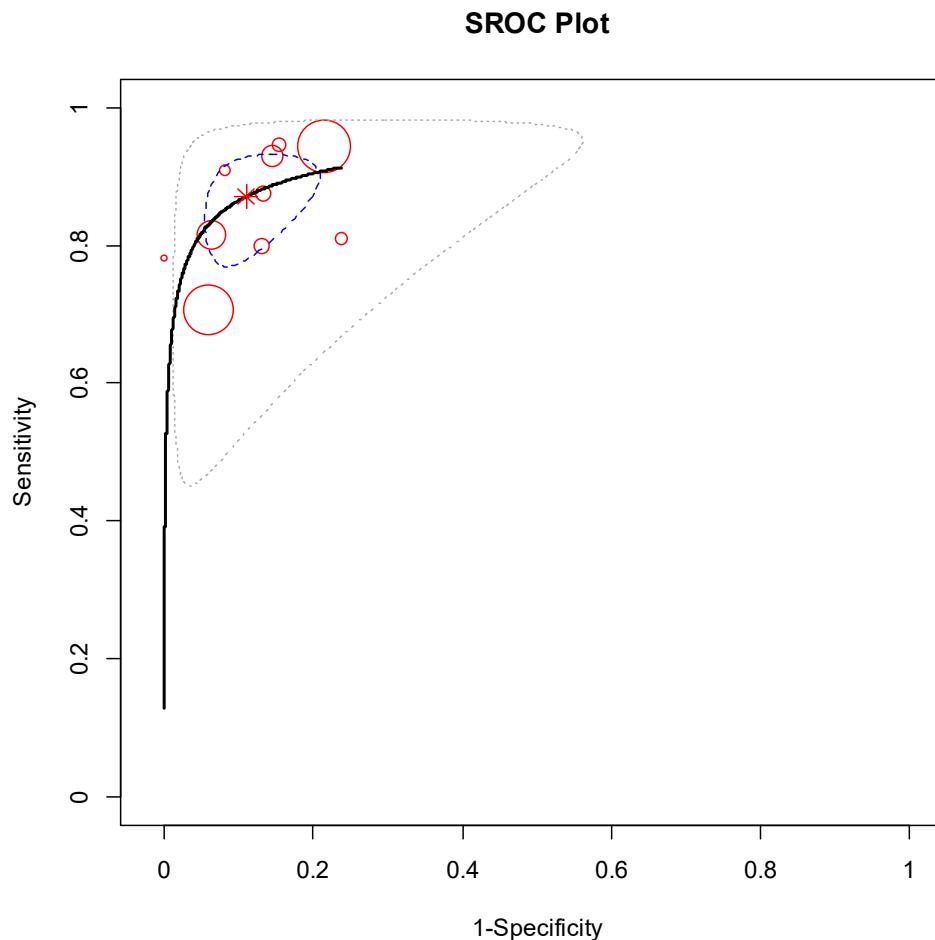
```
forest(res,  
accuracy.type="sens")
```

```
forest(res,  
accuracy.type="spec")
```



Example output ,meat4diag' package

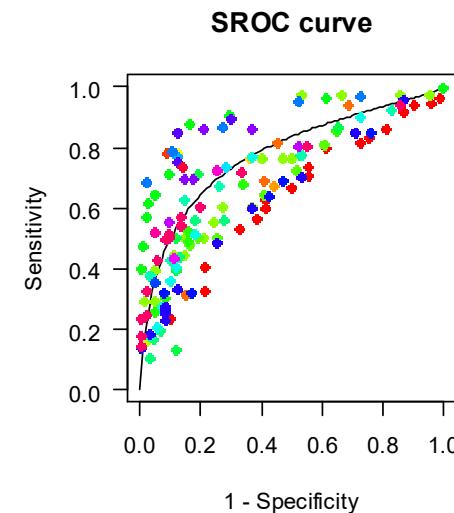
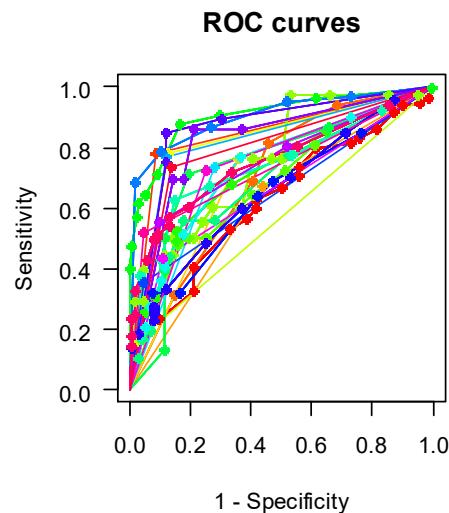
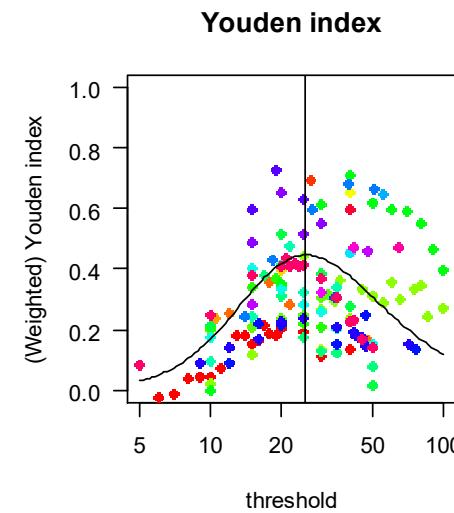
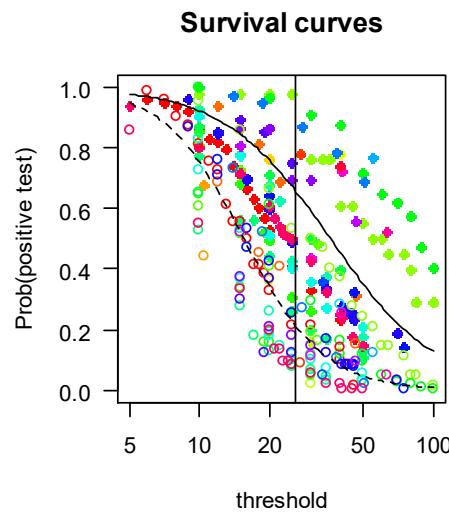
```
#ROC-Plotting  
function:  
SROC(res)
```



Worked examples for selected packages (VI)

```
#6.) diagmeta (considers multiple cutoffs)
library(diagmeta)
data(Schneider)
data(Schneider2017)
diag1 <- diagmeta(tpos, fpos, tneg, fneg, cutpoint,
studlab = paste(author, year, group), data =
Schneider2017, model = "DIDS", log.cutoff = TRUE)
summary(diag1)
plot(diag1)
summary(diag1)
diagstats(diag1)
```

Example output ,diagmeta' package



References and further reading

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024725/>
- https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/
- <http://www.metafor-project.org/doku.php>
- <https://cran.r-project.org/web/packages/mada/vignettes/mada.pdf>
- CRAN Task View: Meta-Analysis (<http://cran.r-project.org/web/views/MetaAnalysis.html>)
- Meta-analysis of diagnostic accuracy studies in Stata:
doi 10.1007/978-1-0716-1566-9_11
- And in SAS: google „MetaDAS: A SAS macro for meta-analysis of diagnostic accuracy studies”

References and further reading

- Schwarzer, Guido, Carpenter, James R., Rücker, Gerta. **Meta-Analysis with R**, Springer (Use R! series)
- Bravata DM, Olkin I, **Simple pooling versus combining in meta-analysis**, **Eval Health Prof** (2001)
- Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2009). **Introduction to Meta-Analysis**. Wiley.
- Chen, D.-G. and Peace, K. E. (2013). **Applied Meta-Analysis with R**. Chapman & Hall/CRC Biostatistics Series
- Cochrane Collaboration (2020). **Cochrane Handbook for Systematic Reviews of Interventions**. Version 6.1. Wiley. URL: <https://training.cochrane.org/handbook/current>
- Egger, M., Davey-Smith, G., and Altman, D. G. (2000). **Practical Meta-Analysis**. Wiley.
- Welton, N. J., Sutton, A. J., Cooper, N. J., Abrams, K. R., and Ades, A. E. (2012). **Evidence Synthesis for Decision Making in Healthcare**. Statistics in Practice. Wiley

References and further reading

- Duval, S. and Tweedie, R. (2000). **Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis.** Biometrics, 56(2):455–463.
- DerSimonian, R. and Laird, N. (1986). **Meta-analysis in clinical trials.** Control Clin Trials, 7(3):177–188.
- Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). **Bias in meta-analysis detected by a simple, graphical test.** BMJ, 315(7109):629–634.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). **Measuring inconsistency in meta-analyses.** BMJ, 327(7414):557–560.
- <https://www.meta-analysis-workshops.com/download/bookChapterSample.pdf>
- Leeflang, M.M.G. (2014). **Systematic reviews and meta-analyses of diagnostic test accuracy.** Clin Microbiol Infect 2019 Nov;25(11):1315-1327
- Doebl P. (2020). **Meta-Analysis of Diagnostic Accuracy with mada** (R package vignette)