University of Zurich[UZH]

Biostatistics

Publication Bias in Meta-Analysis

Giuachin Kreiliger

May 21, 2019

Publication Bias

Studies with large effects are more likely to be published

Incentives for bad research practices (p-hacking...)

Meta-Analysis

Aim to summarize evidence & reach conclusions

classical meta-analyses rely on random sample assumption

Fixed Effects Meta-Analysis

Study i of n studies, effects θ_i and variances v_i, s.e. s_i

Fixed effects estimate: weighted mean with $w_i = 1/\hat{v}_i$:

$$\theta_f = \frac{\sum_{i=1}^n w_i \theta_i}{\sum_{i=1}^n w_i}$$

$$\mathrm{se}(\theta_f) = \frac{1}{\sqrt{\sum_{i=1}^n w_i}}$$

Random Effects Meta-Analysis

Let $\hat{\theta}_i|\theta_i \sim N(\theta_i, v_i)\theta_i \sim N(\theta, \tau^2)$

Random effects estimate: weigthed mean of $\hat{\theta}_i$ with weights $w_i = 1/(\hat{v}_i + \tau^2)$

Evidence Based Medicine

Likely to base decisions on over-optimistic findings

Direct consequences: patient harm and unnecessary research

Cochrane Organisation

Aim: summarise findings in primary clinical research and health care

Provide peer-reviewed, systematic reviews

Public access (for some countries)

Research Question

Quantify the abundance and impact of publication bias in the Cochrane Library

Cochrane Library Dataset

5,016 systematic reviews with studies published until 2018.

52,995 studies.

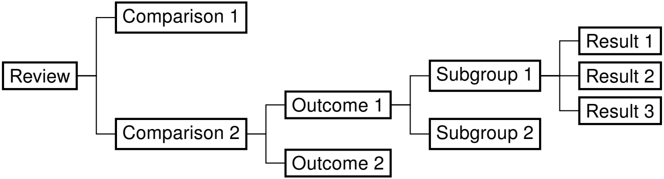
463,820 study results.

Review Example: Binary Outcome

Barbiturate efficacy for head injury treatment

Study	Comparison	Outcome	Events	Total	Events_c	Total_c
Bohn 1989	Barbiturate vs no b	Death at the end of	11	41	11	41
Bohn 1989	Barbiturate vs no b	Death or severe dis	18	41	13	41
Eisenberg 1988	Barbiturate vs no b	Uncontrolled ICP du	25	37	30	36
Eisenberg 1988	Barbiturate vs no b	Hypotension during	23	37	18	36
Perez-Barcena 2008	Pentobarbital vs Th	Death at the end of	16	21	9	21
Perez-Barcena 2008	Pentobarbital vs Th	Death or severe dis	17	21	13	21
Perez-Barcena 2008	Pentobarbital vs Th	Uncontrolled ICP du	18	22	11	22
Perez-Barcena 2008	Pentobarbital vs Th	Hypotension during	20	22	21	22
Schwartz 1984	Barbiturate vs Mann	Death at the end of	6	15	7	14
Schwartz 1984	Barbiturate vs Mann	Uncontrolled ICP du	19	28	12	31
Ward 1985	Barbiturate vs no b	Mean ICP during tre	0	27	0	26
Ward 1985	Barbiturate vs no b	Mean arterial press	0	27	0	26
Ward 1985	Barbiturate vs no b	Mean body temperatu	0	27	0	26

Dataset Structure



Dataset Properties

Review or study level:

	5% quantile	median	mean	95% quantile
Number of studies	1	7	12	40
Number of comparisons	1	2	4	12
Number of meta-analyses	2	19	37	132
Study years	1981	2002	2000	2013
Study sample size	13	78	750	890

Small Study Effects

"The tendency for the smaller studies to show larger treatment effects" (Sterne et al., 2001)

Small Study Effects

Causes:

- Selective publication of studies with significant results - publication bias
- Systematic differences in study settings

Small Study Effect Tests

Different approaches:

- Simple linear regression
- Rank correlation

Special methods for binary outcomes

Regression based Tests

study i of n studies, effects θ_i and variances v_i, s.e. s_i

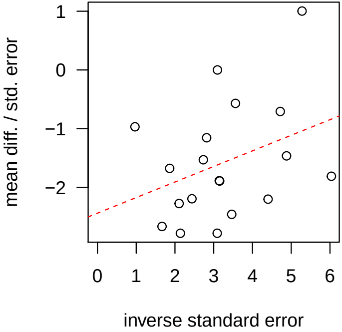
θ_M is the pooled effect and τ^2 the between-study variance.

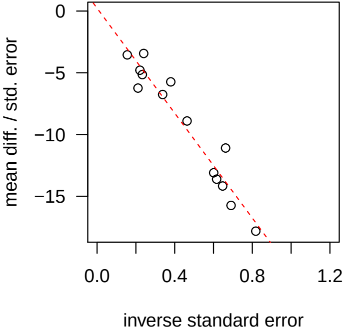
Let $y_i = \theta_i/_i$ and $x_i = 1/s_i$

- Egger et al. (1997) : Simple linear regression
  $y_i = \beta_0 + \beta_1 x_i, \epsilon_i \sim N(0, \sigma)$
- Thompson and Sharp (1999) : extension of Egger with study weights $v_i + \tau^2$

Egger's Test examples

Test for non-zero intercept β_0





Regression Tests for Binary Outcomes

- Peters et al. (2006) :$x_i = 1/n_i$ instead $1/s_i$, inverse variances as weight.
- Harbord et al. (2006) :x_i = score of the log-likelihood of a proportion and inverse variances as weights.
- Rücker et al. (2008) :Use arcsine variance stabilizing transformation for variances and effects, do e.g. Egger's test.

Rank based tests

Begg (1988):

Let y_i be $\frac{\theta_i - \theta_M}{v_i}$ and x_i its variance ($\neq v_i$)

u the number of pairs (y_i, x_i) ranked in the same order, l the number of pairs in the opposite order

$Z = \frac{(u-l)}{\sqrt{n(n-1)(2n+5)/18}}$ is a test statistic

Rank based tests

Schwarzer et al. (2007):

e_t number of events in the treatment group

E_t follows hypergeometric distribution: calculate $\mathbb{E}(E_t)$ and variances

proceed as in Begg (1988)

Test Results

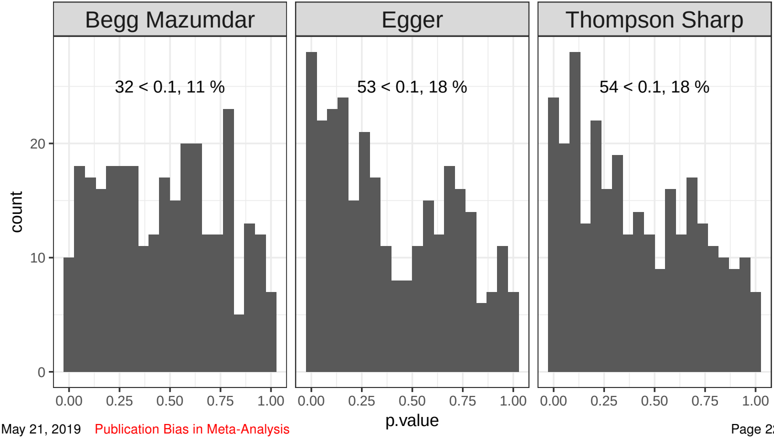
Inclusion criteria (from Ioannidis and Trikalinos (2007)):

- $n \geq 10$
- at least one statistically significant effect in a study
- $\frac{\hat{v_{\max}}^2}{\hat{v_{\min}}^2} > 4$
- $I^2 < 0.5$

From 5338 with $n \geq 10$, 1484 remain.

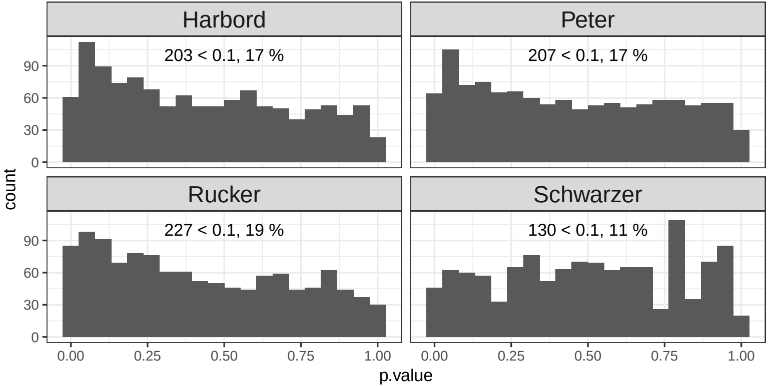
Continuous Outcome Test Results

p-values distribution, $n = 294$:



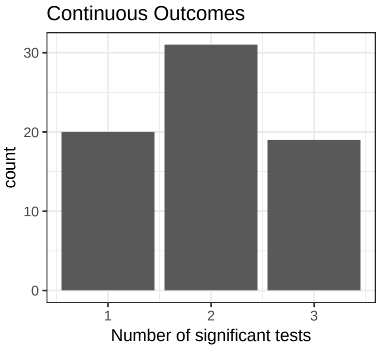
Binary Outcome Test Results

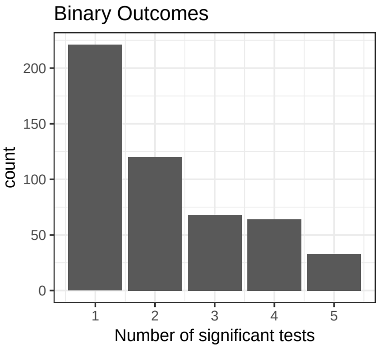
p-values distribution, n = 1190:



Agreement in significance

Number of significant test results per meta-analysis:





Small Study Effect Adjustment

Three methods:

- Regression
- Copas selection model
- Trim-and-fill

Adjustment by regression

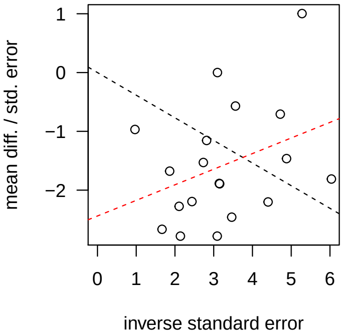
$y_i = \theta_i / s_i, x_i = 1/s_i$

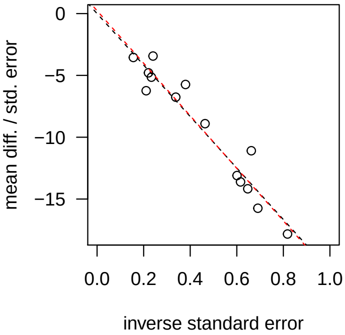
$y_i = \beta_0 + \beta_1 x_i, \epsilon_i \sim N(0, \sigma)$

β_1 is the weighted mean treatment effect if $\beta_0 = 0$

Adjustment by regression

Radial plots (continuous outcome examples):





Limit Meta-Analysis

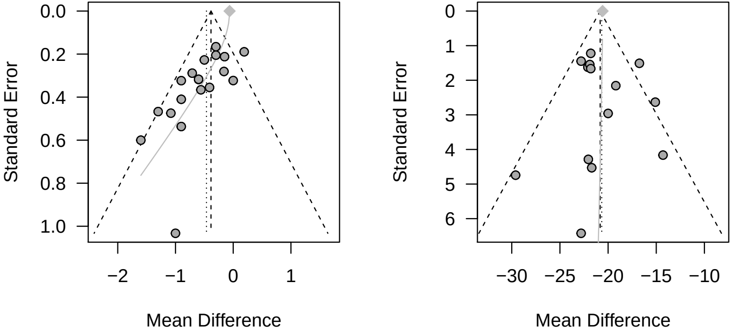
Extended random effects model:

$$\theta_i = \beta_0 + \beta_1(\sqrt{v_i + \tau^2}) + \epsilon_i(\sqrt{v_i + \tau^2}),$$
$$\epsilon_i \overset{\text{iid}}{\sim} N(0, 1)$$

Use $\mathbb{E}(\theta_i) \rightarrow \beta_0 + \beta_1 \tau$ for $\sqrt{v_i} \rightarrow 0$ as corrected treatment efffect.

Limit Meta-Analysis

Funnel plot with effect with infinite precision:



Selection model

Copas and Shi (2001): model based on a bivariate normal distribution:

$$\hat{\theta}_i = \theta_i + \sigma_i \epsilon_i \quad (1)$$

$$\theta_i \sim N(\theta, \tau^2) \quad (2)$$

$$z_i = a + b/\hat{s}_i + \delta_i \quad (3)$$

1,2 is called population model, 3 the selection model

(ϵ_i, δ_i) are standard normal residuals with correlation $\rho = cor(y_i, z_i)$.

Sensitivity Analysis

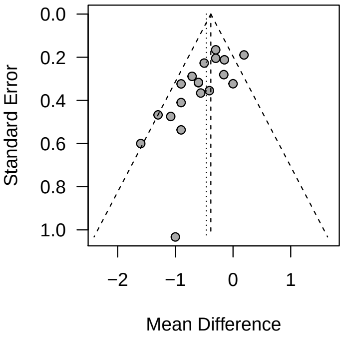
Model the selection process with different a, b

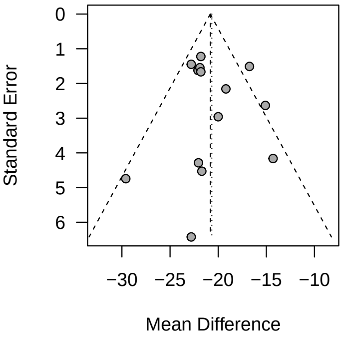
Test if small study effect is significant, by including
$\theta_i = \theta_i + \beta s_i + v_i \epsilon_i$

Estimation: Select a, b such that $H0$ can not be rejected and estimated number of unpublished studies is minimal.

Trim-and-Fill

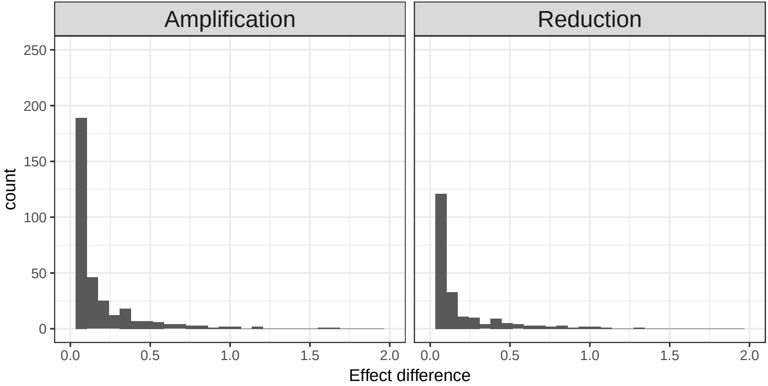
Mirror studies that cause asymmetry:





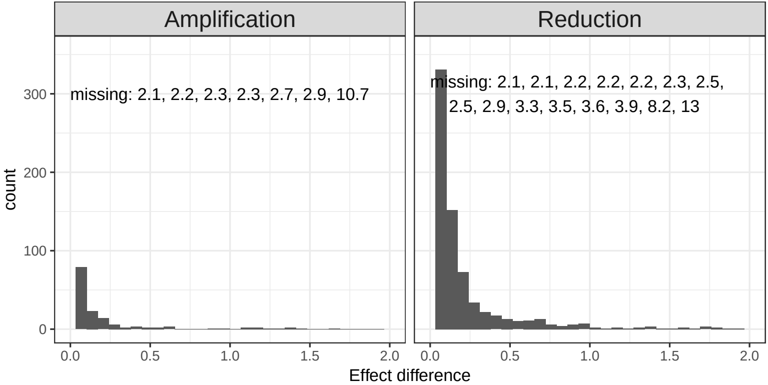
Results:

Difference between random and fixed effects meta-analysis estimate: $|\theta_f - \theta_r|$



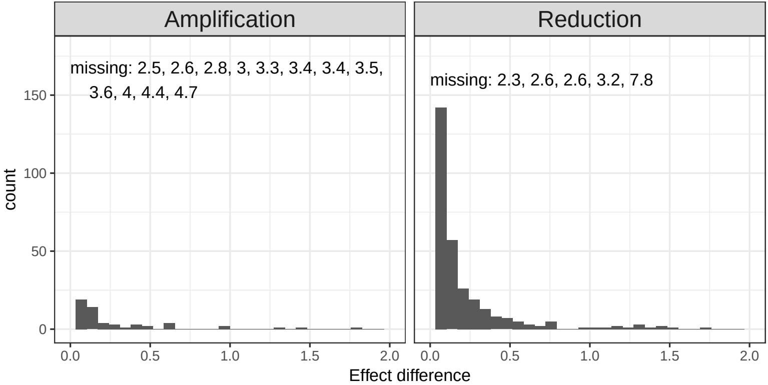
Adjustment Results: Trim-and-fill

Absolute difference between adjusted and fixed effects meta-analysis estimate: $|\theta_f - \theta_{\text{adjusted}}|$



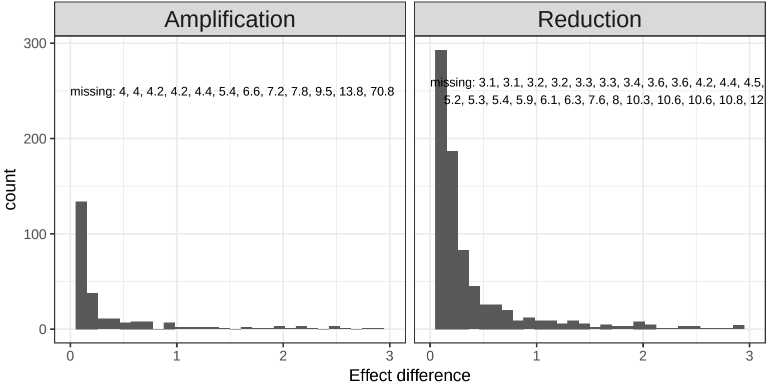
Adjustment Results: Copas

Absolute difference between adjusted and fixed effects meta-analysis estimate: $|\theta_f - \theta_{\text{adjusted}}|$



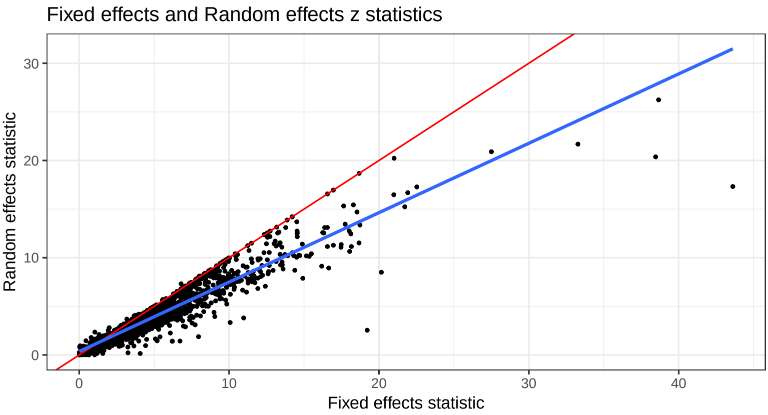
Adjustment Results: Regression

Absolute difference between adjusted and fixed effects meta-analysis estimate: $|\theta_f - \theta_{\text{adjusted}}|$



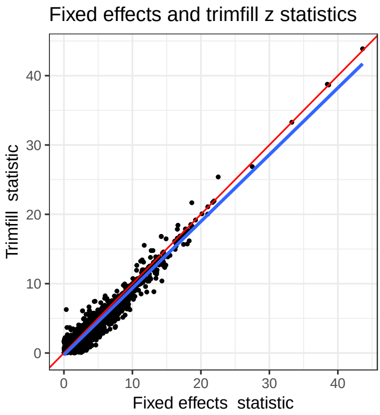
Results:

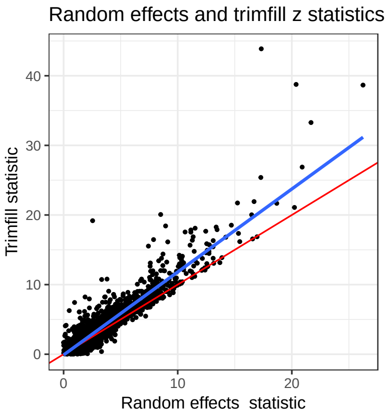
Random and fixed effects meta-analyses test statistics:



Adjustment Results: Trim-and-fill

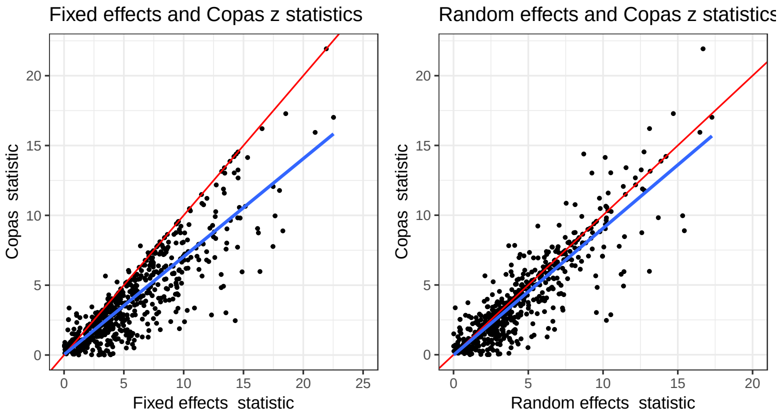
Adjusted and meta-analysis test statistics:





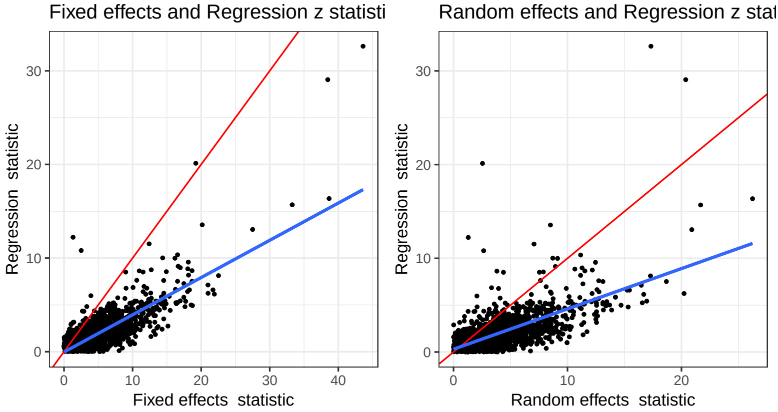
Adjustment Results: Copas

Adjusted and meta-analysis test statistics:



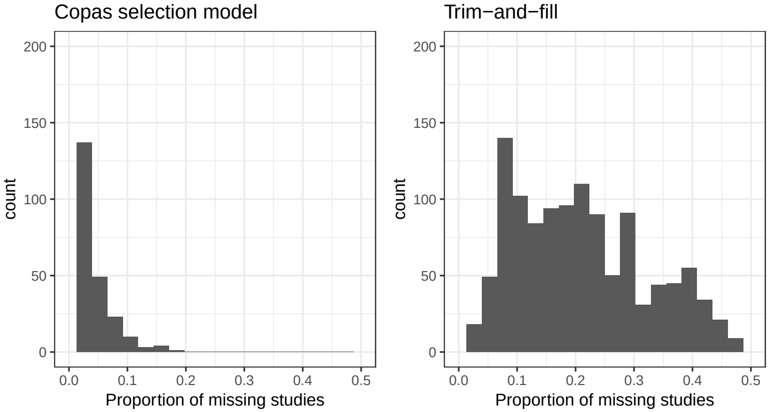
Adjustment Results: Regression

Adjusted and meta-analysis test statistics:



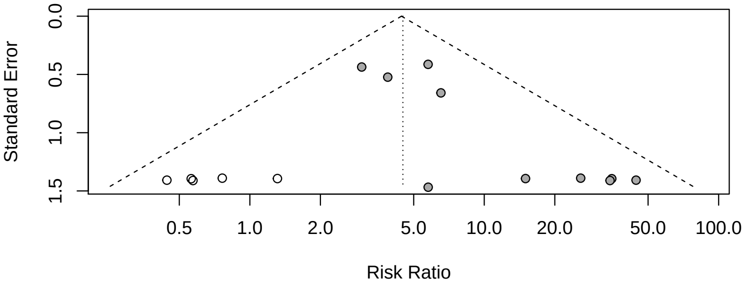
Adjustment Results

Missing study proportions:



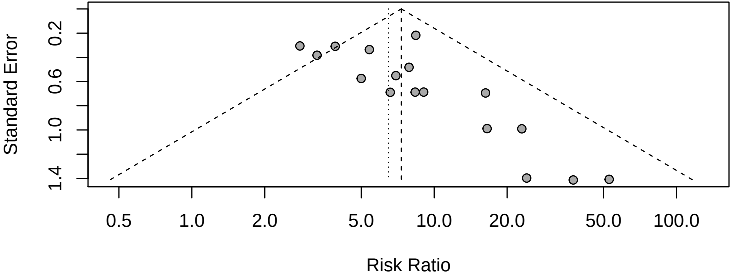
Extreme Results

RR reduction by trimfill (-3.9), side effects



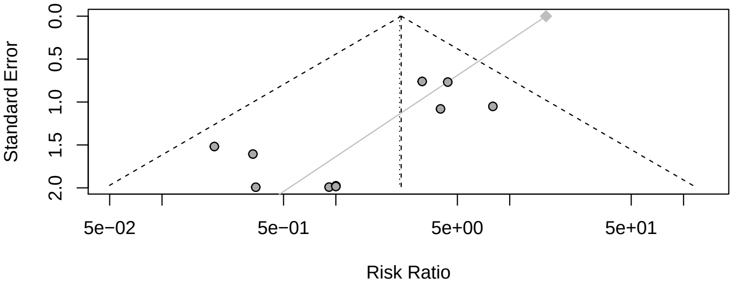
Extreme Results

RR Reduction by copas selection model (-4), pain relief



Extreme Results

RR Amplification by regression (+14), side effects



Discussion

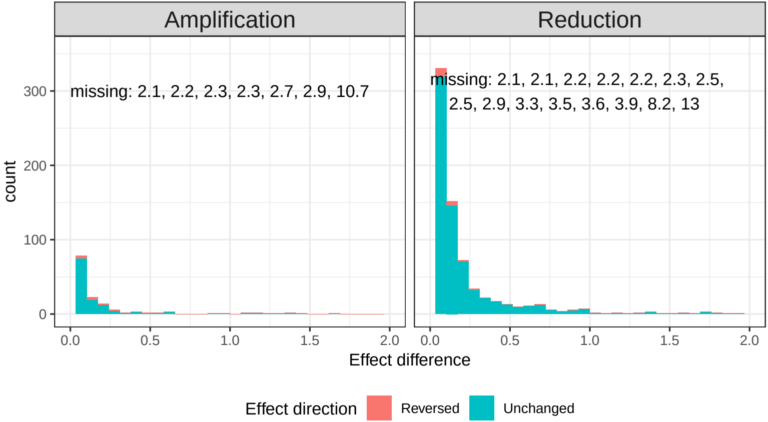
- Proportion of positive tests is well above 10%
- Effect sizes and evidence for treatment effect is diminishued
- Limitations: not only primary outcomes, adjustment methods known to perform poorly under the 0

Outlook

– Connect results with different medical fields, look for differences
– Connect results with single studies and journals (?)

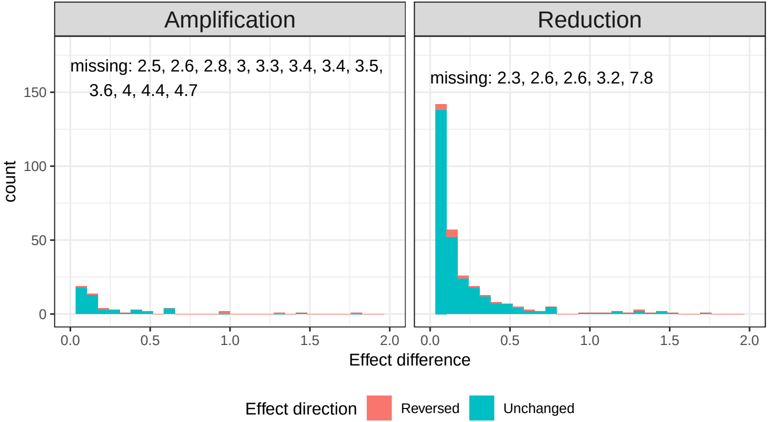
Adjustment Results: Trim-and-fill

Treatment effect difference:



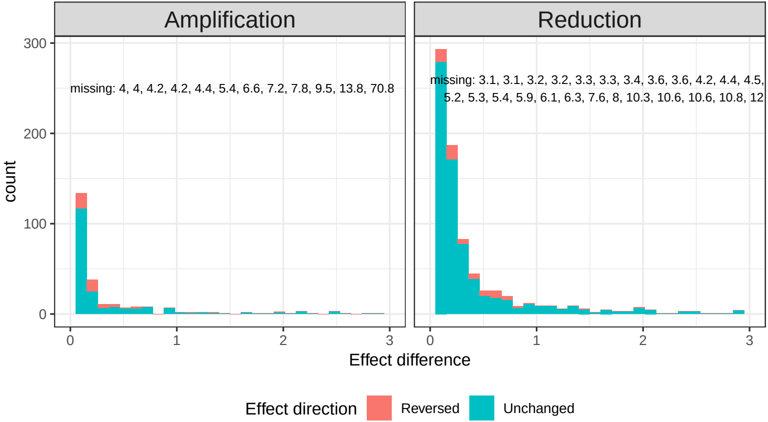
Adjustment Results: Copas

Treatment effect difference:



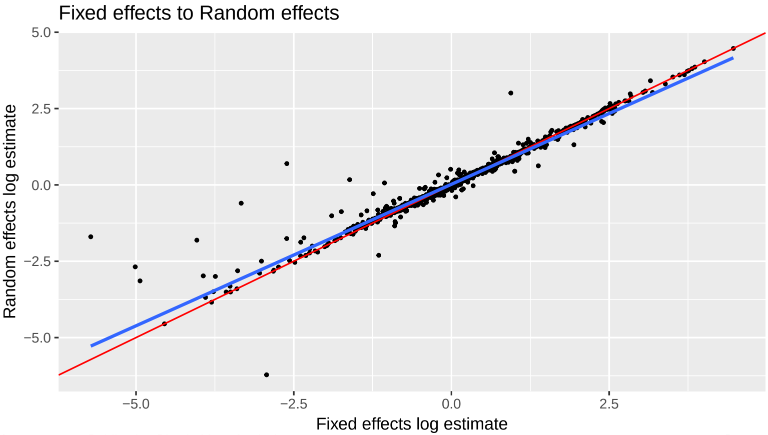
Adjustment Results: Regression

Treatment effect difference:



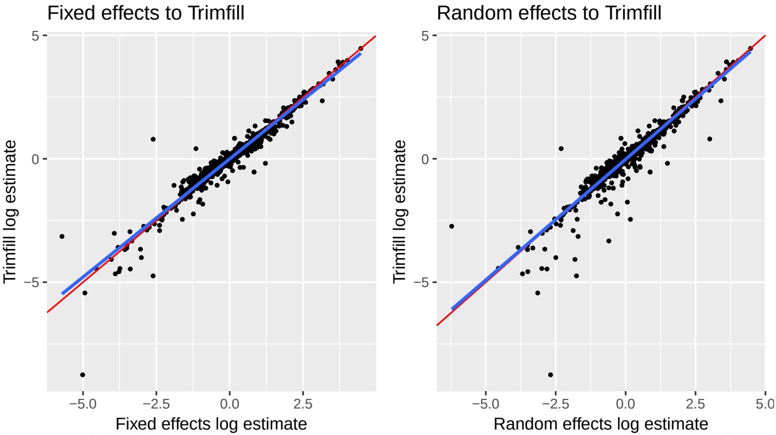
Results

log treatment effect estimates:



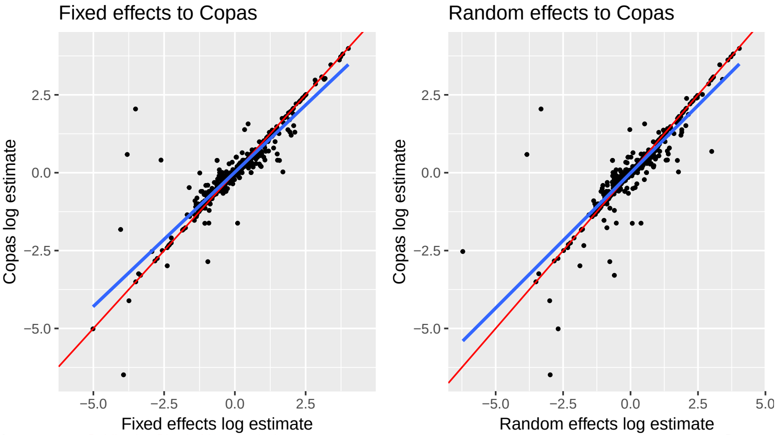
Adjustment Results: Trim-and-fill

log treatment effect estimates:



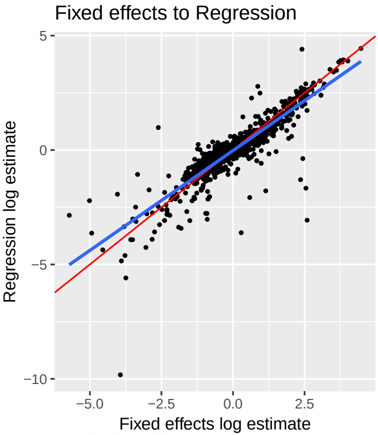
Adjustment Results: Copas

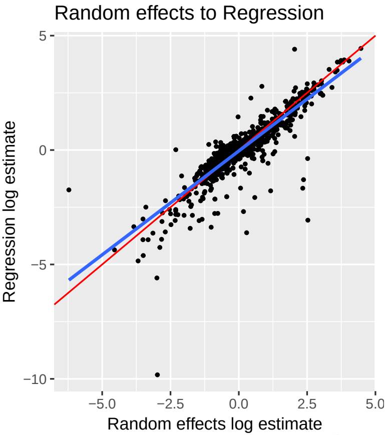
log treatment effect estimates:



Adjustment Results: Regression

log treatment effect estimates:





References

Begg, C. B. (1988). Statistical methods in medical research p. armitage and g. berry, blackwell scientific publications, oxford, u.k., 1987. no. of pages: 559. price Â£22.50. *Statistics in Medicine*, 7(7):817–818.

Copas, J. B. and Shi, J. Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research*, 10(4):251–265. PMID: 11491412.

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.

Harbord, R. M., Egger, M., and Sterne, J. A. C. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, 25(20):3443–3457.