University of BRISTOL

MISUNDERSTANDINGS AND MISUSES OF COMMONLY-CITED METHODS IN META-ANALYSIS

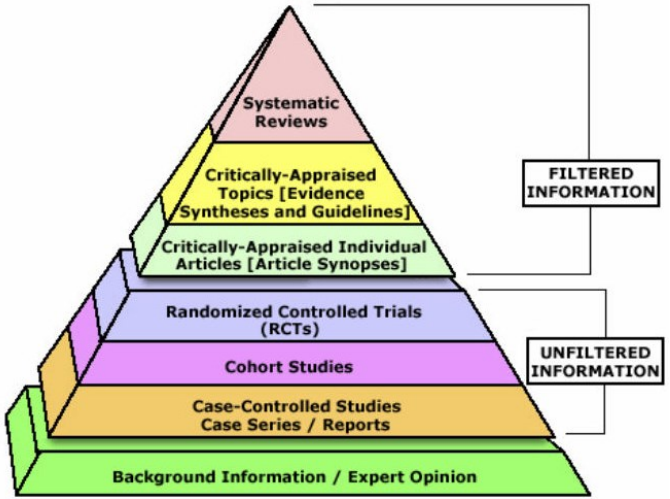
Julian Higgins

Professor of Evidence Synthesis, University of Bristol, UK

bristol.ac.uk

Introductory remarks

- Systematic reviews and meta-analyses have become critical for decision making in health as well as other areas
- They are some of the most influential types of research



Introductory remarks

- Systematic reviews and meta-analyses have become critical for decision making in health as well as other areas
- They are some of the most influential types of research
- They are among the most highly cited of research articles

- Here are the **top 5** most cited papers on “meta-analysis” as of yesterday

- …according to Scopus

	Document title	Authors	Source	Year	Citations
1	Review • *Open access* **Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement** Get it! Related documents	Moher, D., Liberati, A., Tetzlaff, J., ... Tovey, D., Tugwell, P.	PLoS Medicine, 6(7), e1000097	2009	50,896
2	Review • *Open access* **Measuring inconsistency in meta-analyses** Get it! Related documents	Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G.	British Medical Journal , 327(7414), pp. 557–560	2003	46,559
3	Article • *Open access* **Bias in meta-analysis detected by a simple, graphical test** Show abstract Get it! Related documents	Egger, M., Smith, G.D., Schneider, M., Minder, C.	British Medical Journal , 315(7109), pp. 629–634	1997	40,478
4	Article **Meta-analysis in clinical trials** Show abstract Get it! Related documents	DerSimonian, R., Laird, N.	Controlled Clinical Trials, 7(3), pp. 177–188	1986	31,845
5	Article **Quantifying heterogeneity in a meta-analysis** Show abstract Get it! Related documents	Higgins, J.P.T., Thompson, S.G.	Statistics in Medicine, 21(11), pp. 1539–1558	2002	25,501

More legibly

			Citations
1	Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (*PLoS Med*)	Moher et al 2009	50,896
2	Measuring inconsistency in meta-analyses	Higgins et al 2003	46,559
3	Bias in meta-analysis detected by a simple, graphical test	Egger et al 1997	40,478
4	Meta-analysis in clinical trials	DerSimonian & Laird 1986	31,845
5	Quantifying heterogeneity in a meta-analysis	Higgins et al 2002	25,501

Reordered: 6 topics

1	Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement
2	Measuring inconsistency in meta-analyses
5	Quantifying heterogeneity in a meta-analysis
4	Meta-analysis in clinical trials
3	Bias in meta-analysis detected by a simple, graphical test

Outline of my presentation

1	Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement	**i. Misuse of reporting guidelines**
2	Measuring inconsistency in meta-analyses	**ii. Misuse of I-squared**
5	Quantifying heterogeneity in a meta-analysis	
4	Meta-analysis in clinical trials	**iii. Misuse of random-effects meta-analysis**
3	Bias in meta-analysis detected by a simple, graphical test	**iv. Misuse of tests for funnel plot asymmetry**

OPEN ACCESS Freely available online

PLoS MEDICINE

Guidelines and Guidance

Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

David Moher[1,2]*, Alessandro Liberati[3,4], Jennifer Tetzlaff[1], Douglas G. Altman[5], The PRISMA Group[¶]

1 Ottawa Methods Centre, Ottawa Hospital
University of Ottawa, Ottawa, Ontario, Cana
Negri, Milan, Italy, **5** Centre for Statistics in

Introduction

Systematic reviews and meta-ana
important in health care. Clinician
with their field [1,2], and they are c
developing clinical practice guide
require a systematic review to e
further research [3], and some hea
this direction [4]. As with all rese
review depends on what was done,
of reporting. As with other publica
systematic reviews varies, limiting
strengths and weaknesses of those

Several early studies evaluated t
1987, Mulrow examined 50 review a
medical journals in 1985 and 1986 a
explicit scientific criteria, such as a quality assessment of included

RESEARCH METHODS AND REPORTING

The PRISMA 2020 statement: an updated guideline for reporting systematic reviews

Matthew J Page,[1] Joanne E McKenzie,[1] Patrick M Bossuyt,[2] Isabelle Boutron,[3] Tammy C Hoffmann,[4] Cynthia D Mulrow,[5] Larissa Shamseer,[6] Jennifer M Tetzlaff,[7] Elie A Akl,[8] Sue E Brennan,[1] Roger Chou,[9] Julie Glanville,[10] Jeremy M Grimshaw,[11] Asbjørn Hróbjartsson,[12] Manoj M Lalu,[13] Tianjing Li,[14] Elizabeth W Loder,[15] Evan Mayo-Wilson,[16] Steve McDonald,[1] Luke A McGuinness,[17] Lesley A Stewart,[18] James Thomas,[19] Andrea C Tricco,[20] Vivian A Welch,[21] Penny Whiting,[17] David Moher[22]

(http://www.prisma-statement.org/).

What's the paper about?

- A checklist and flow chart for the reporting of a systematic review
  - Preceded by QUOROM
  - PRISMA published in 2009 updated to PRISMA 2020

What's the problem?

PRISMA is *not* a guideline for *doing* systematic reviews

Randomly picked paper (*most recent citation of the paper in Scopus yesterday*)

Egypt. J. Vet. Sci. **Vol. 56,** No. 4, pp. 679-690 (2025)

Egyptian

[Review Article]

A meta-analysis of Prec
dried Porcine Plasma or

Momunova Aigul Abdykerimo
Karlygash [4], Sokolov Dmitri
Zhumagaliuly [7]

Methods

This meta-analysis strictly followed the guidelines set by the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [18] and the Meta-analyses Of Observational Studies in Epidemiology (MOOSE) [19].

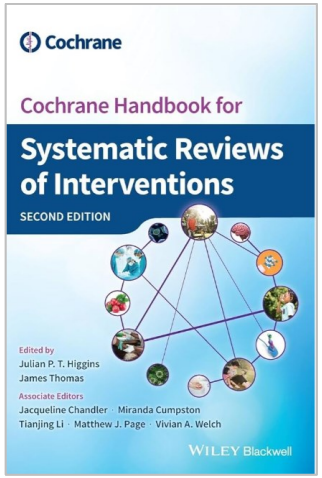
What's the problem?

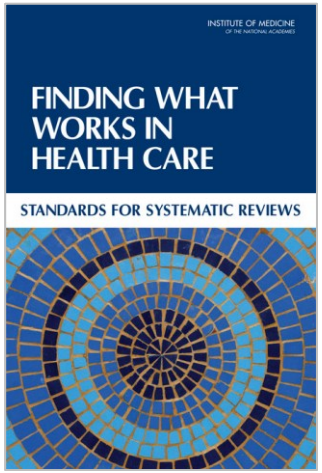
The PRISMA Statement

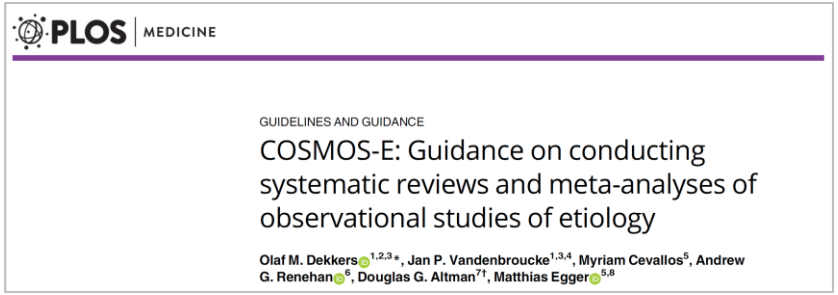
The PRISMA Statement consists of a 27-item checklist (Table 1; see also Text S1 for a downloadable Word template for researchers to re-use) and a four-phase flow diagram (Figure 1; see also Figure S1 for a downloadable Word template for researchers to re-use). The aim of the PRISMA Statement is to help authors improve the reporting of systematic reviews and meta-analyses. We have focused on randomized trials, but PRISMA can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions. PRISMA may also be useful for critical appraisal of published systematic reviews. However, the PRISMA checklist is not a quality assessment instrument to gauge the quality of a systematic review.

What's the solution?

- Follow documents that intend to provide guidance, e.g.
  - Institute of Medicine (IoM) Standards
  - COSMOS-E (for observational studies)
  - Cochrane Handbooks







Measuring inconsistency in meta-analyses

Julian P T Higgins, Simon G Thompson, Jonathan J Deeks, Douglas G Altman

Cochrane Reviews have recently started including the quantity I^2 to help readers assess the consistency of the results of studies in meta-analyses. What does this new quantity mean, and why is assessment of heterogeneity so important to clinical practice?

Systematic reviews and meta-analyses can provide convincing and reliable evidence relevant to many aspects of medicine and health care.[1] Their value is especially clear when the results of the studies they include show clinically important effects of similar magnitude. However, the conclusions are less clear when the included studies have differing results. In an attempt to establish whether studies are consistent, reports of meta-analyses commonly present a statistical test of heterogeneity. The test seeks to determine whether there are genuine differences underlying the results of the studies (heterogeneity), or whether the variation in findings is compatible with chance alone (homogeneity). However, the test is susceptible to the number of trials included in the meta-analysis. We have developed a new quantity, I^2, which we believe gives a better measure of the consistency between trials in a meta-analysis.

Need for consistency

Assessment of the consistency of effects across studies is an essential part of meta-analysis. Unless we know

intervals not overlapping. But the test of heterogeneity yields a P value of 0.09, conventionally interpreted as being non-significant. Because the test is poor at detecting true heterogeneity, a non-significant result cannot be taken as evidence of homogeneity. Using a cut-off of 10% for significance[12] ameliorates this problem but increases the risk of drawing a false positive conclusion (type I error).[10]

Conversely, the test arguably has excessive power when there are many studies, especially when those studies are large. One of the largest meta-analyses in the *Cochrane Database of Systematic Reviews* is of clinical trials of tricyclic antidepressants and selective serotonin reuptake inhibitors for treatment of depression.[13] Over 15 000 participants from 135 trials are included in the assessment of comparative drop-out rates, and the test for heterogeneity is significant (P = 0.005). However, this P value does not reasonably describe the extent of heterogeneity in the results of the trials. As we show later, a little inconsistency exists among these trials but it does not affect the conclusion of the review (that serotonin reuptake inhibitors have

MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR
Julian P T Higgins *statistician*
Simon G Thompson *director*

Cancer Research UK/NHS Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Jonathan J Deeks *senior medical statistician*
Douglas G Altman *professor of statistics in medicine*

Correspondence to:
J P T Higgins
julian.higgins@mrc-bsu.cam.ac.uk

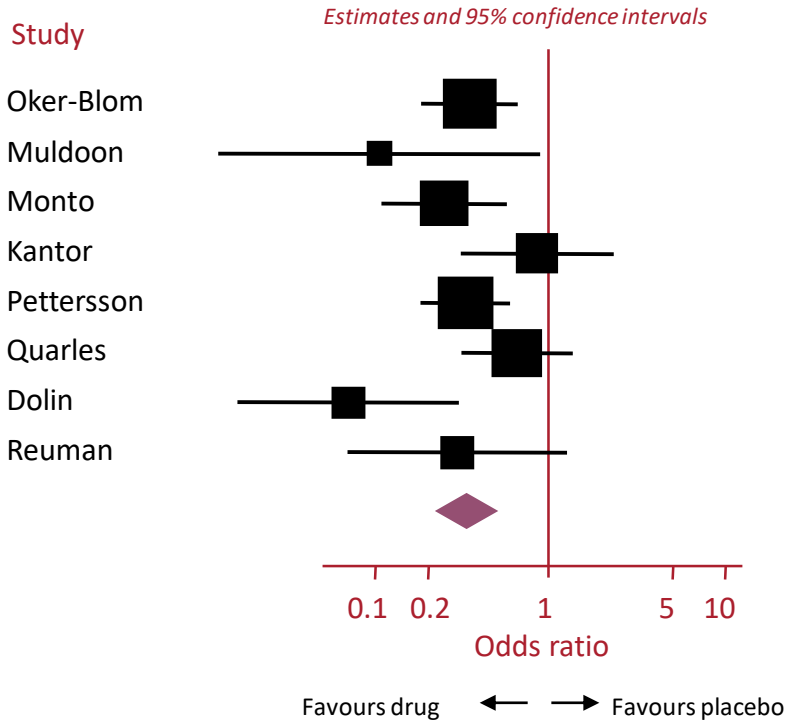
BMJ 2003;327:557–60

What's the paper about?

- Proposes a simple statistic (**I-squared**)
  - measures *inconsistency* – the extent to which results (estimates and confidence intervals) are consistent across studies
  - or interpreted as the *proportion of variability due to heterogeneity rather than chance*

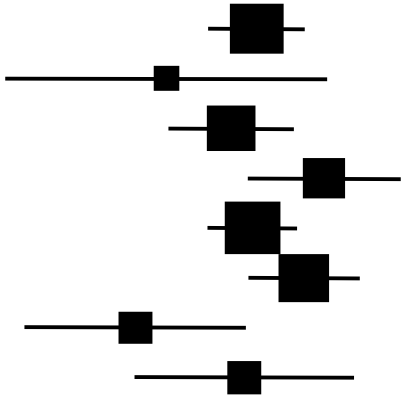
$$\iota^2 = \frac{\tau^2}{\tau^2 + \tilde{\sigma}^2} \times 100\%$$

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

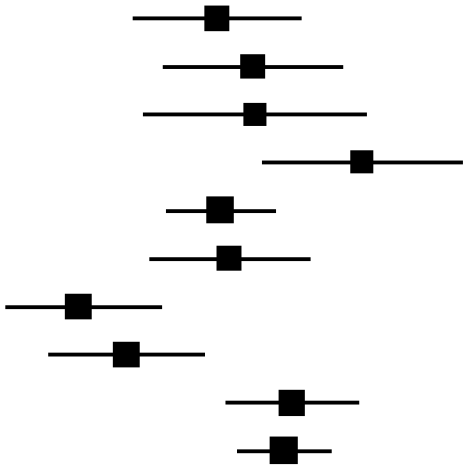


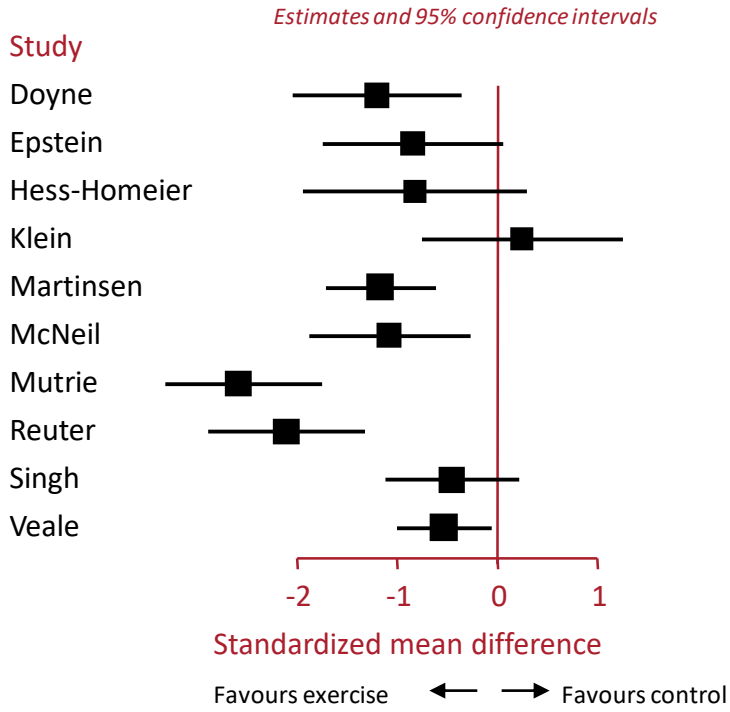
Amantadine for ‘flu

Estimates and 95% confidence intervals



Estimates and 95% confidence intervals





Exercise for depression

What's the problem?

I-squared does *not* measure the amount of heterogeneity

Randomly picked paper (*most recent citation of the paper*)

Veterinary Integrative Sciences 2025; 23(2): e2025033-1-18. DOI: 10.12982/VIS.2025.033

Faculty of Veterinary Medicine
Chiang Mai University

Veterinary Integrative Sciences
ISSN: 2629-9968 (online)

Research articl

Prevalen

Phirum Or[1], S

[1] Veterinary Public Health
[2] Department
[3] Center of Excellence in Veterinary Public Health, Faculty of Veterinary Medicine, Chiang Mai University, Chiang Mai 50100, Thailand

Pooled prevalence of ESBL *E. coli* in broiler farms

Pooled prevalence and its 95% confidence interval (95% CI) were computed using a random effect model. Cochran's *Q* test was used to estimate the heterogeneity of pooled prevalence. In addition, I squared statistic (I^2) was used to quantify the degree of heterogeneity between studies, with I^2 value of 25%, 50%, and 75% indicating low, moderate, and high degrees of heterogeneity, respectively (Higgins et al., 2003).

What's the solution?

- There's a statistic that's perfect to measure heterogeneity
  - heterogeneity standard deviation
  - often called **tau** (τ)

- Can be difficult to interpret, but may be re-expressed to describe a range of effects (see later)

Meta-Analysis in Clinical Trials*

Rebecca DerSimonian and Nan Laird

ABSTRACT: This paper examines eight published reviews each reporting results from several related trials. Each review pools the results from the relevant trials in order to evaluate the efficacy of a certain treatment for a specified medical condition. These reviews lack consistent assessment of homogeneity of treatment effect before pooling. We discuss a random effects approach to combining evidence from a series of experiments comparing two treatments. This approach incorporates the heterogeneity of effects in the analysis of the overall treatment efficacy. The model can be extended to include relevant covariates which would reduce the heterogeneity and allow for more specific therapeutic recommendations. We suggest a simple noniterative procedure for characterizing the distribution of treatment effects in a series of studies.

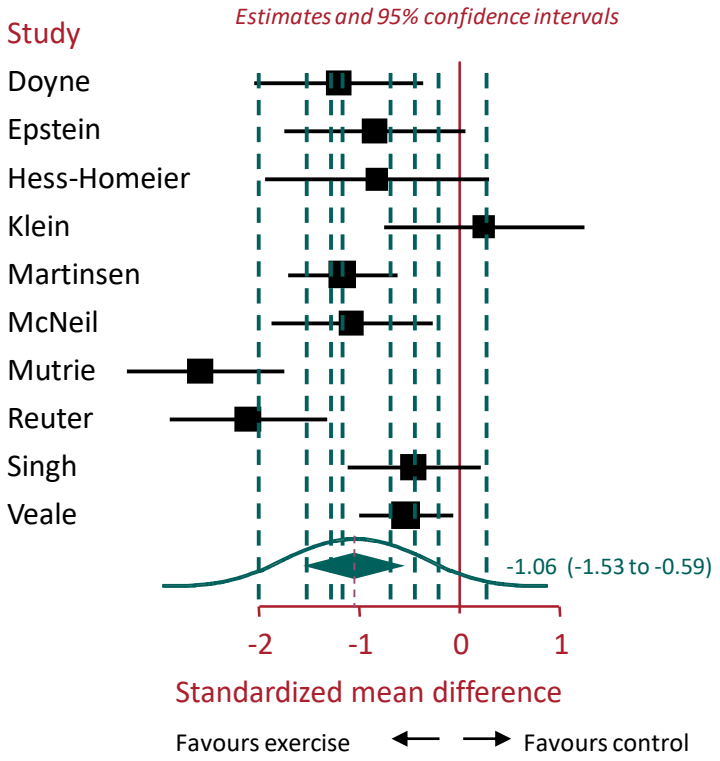
KEY WORDS: *random effects model, heterogeneity of treatment effects, distribution of treatment effects, covariate information*

INTRODUCTION

Meta-analysis is defined here as the statistical analysis of a collection of analytic results for the purpose of integrating the findings. Such analyses are becoming increasingly popular in medical research where information on

What's the paper about?

- Landmark paper describing most popular statistical method for **random-effects meta-analysis**



What's the problem?

- There are technical problems with the methods described (now mostly overcome)
- A bigger, conceptual problem is to *fail to recognize a random-effects model assumes a distribution of effects*

What's the problem?

Randomly picked paper (*n*

Aging Clinical and Experimental Rese
https://doi.org/10.1007/s40520-024-

RESEARCH

Fish consumption,
dose-response me

Justyna Godos[1,2] · Agnieszk
Alberto Dolci[10] · Cristian Ric

Received: 21 June 2024 / Accepted:

the lowest category of fish consumption, the highest consumption was related to 18%, 15% and 18% lower risk of each aforementioned outcome, respectively (RR = 0.82, 95% CI: 0.73–0.93 for dementia, RR = 0.80, 95% CI: 0.67–0.96 for Alzheimer's disease, and RR = 0.82, 95% CI: 0.75–0.90 for cognitive impairment/decline; Fig. 1). The evidence of

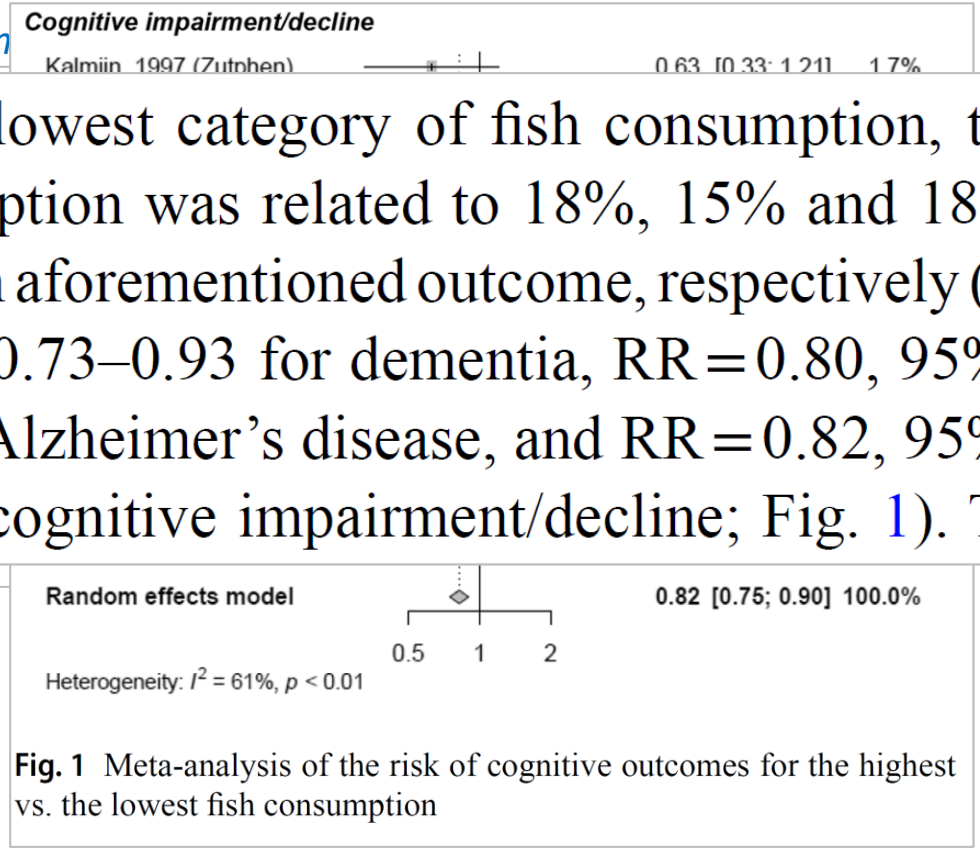


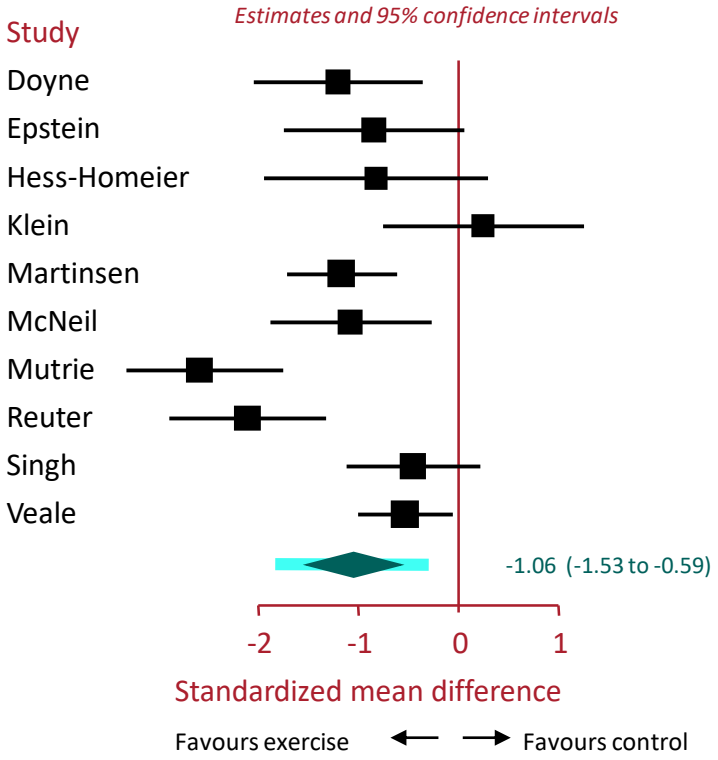
Fig. 1 Meta-analysis of the risk of cognitive outcomes for the highest vs. the lowest fish consumption

What's the problem?

Random-effects meta-analyses aim to learn about *distributions*, not single effects

What's the solution?

- A convenient way to describe the amount of heterogeneity: prediction interval
- Interval in which 95% of true effects from similar studies will lie



Higgins, Thompson & Spiegelhalter, *JRSS A* 2009

iv

Bias in meta-analysis detected by a simple, graphical test

Matthias Egger, George Davey Smith, Martin Schneider, Christoph Minder

Abstract

Objective: Funnel plots (plots of effect estimates against sample size) may be useful to detect bias in meta-analyses that were later contradicted by large trials. We examined whether a simple test of asymmetry of funnel plots predicts discordance of results when meta-analyses are compared to large trials, and we assessed the prevalence of bias in published meta-analyses.

Design: Medline search to identify pairs consisting of a meta-analysis and a single large trial (concordance of results was assumed if effects were in the same direction and the meta-analytic estimate was within 30% of the trial); analysis of funnel plots from 37 meta-analyses identified from a hand search of four leading general medicine journals 1993-6 and 38 meta-analyses from the second 1996 issue of the *Cochrane Database of Systematic Reviews*.

Main outcome measure: Degree of funnel plot asymmetry as measured by the intercept from

analyses have later been contradicted by large randomised controlled trials.[4] Such discrepancies have brought discredit on a technique that has been controversial since the outset.[5] The appearance of misleading meta-analysis is not surprising considering the existence of publication bias and the many other biases that may be introduced in the process of locating, selecting, and combining studies.[6-9]

Funnel plots, plots of the trials' effect estimates against sample size, may be useful to assess the validity of meta-analyses.[4 10] The funnel plot is based on the fact that precision in estimating the underlying treatment effect will increase as the sample size of component studies increases. Results from small studies will scatter widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot will resemble a symmetrical inverted funnel. Conversely, if there is bias, funnel plots will often be skewed and asymmetrical.

The value of the funnel plot has not been systematically examined, and symmetry (or asymme-

Department of Social Medicine, University of Bristol, Bristol BS8 2PR

Matthias Egger, *reader in social medicine and epidemiology*

George Davey Smith, *professor of clinical epidemiology*

Department of Social and Preventive Medicine, University of Berne, CH-3012 Berne, Switzerland

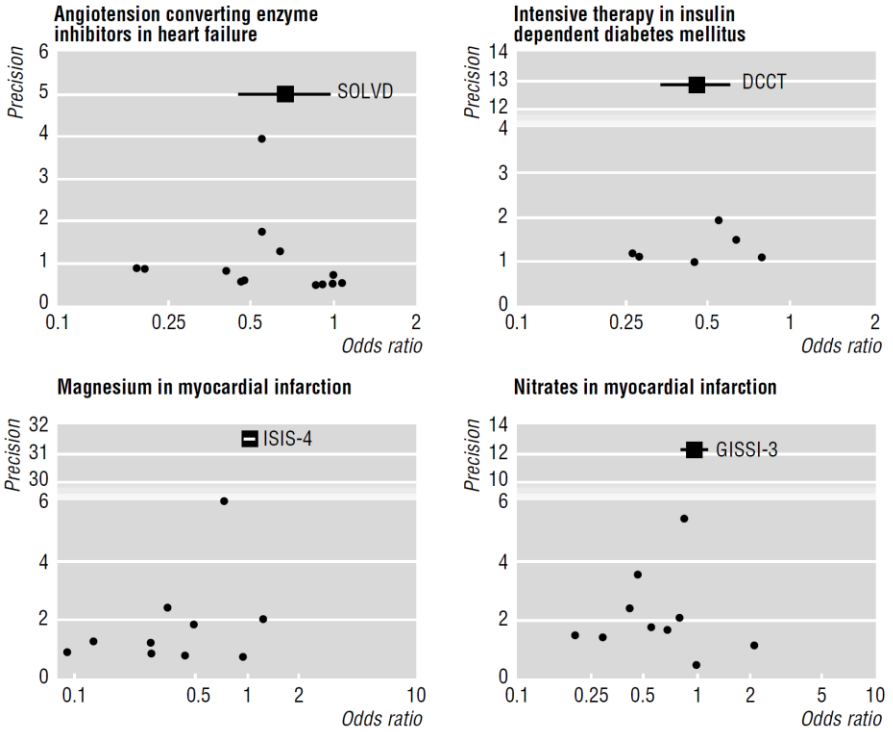
Martin Schneider, *research associate*

Christoph Minder, *head, medical statistics unit*

Correspondence to:

What's the paper about?

- **Funnel plots** plot effect size against precision (study size)
- Funnel plots would ideally be symmetrical
- The **Egger test** aims to identify asymmetry



What's the problem?

Randomly picked pa

The existence of publication bias due to the small study effect was suggested by a funnel plot as shown in Figure 4 and Egger's test which indicated a statistically significant coefficient bias (-5.29±2.54, p = 0.0449).

Veterinary Integrative Scien

Pu

Funnel plot and Eggers's test were conducted to measure publication bias, al.,

Resea

P

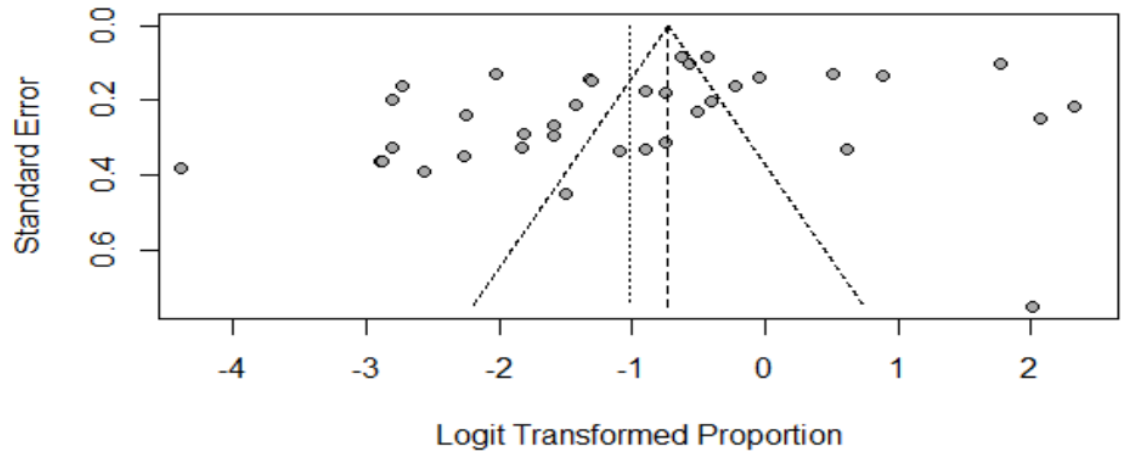


Figure 4 Funnel plot to measure publication bias of the included studies.

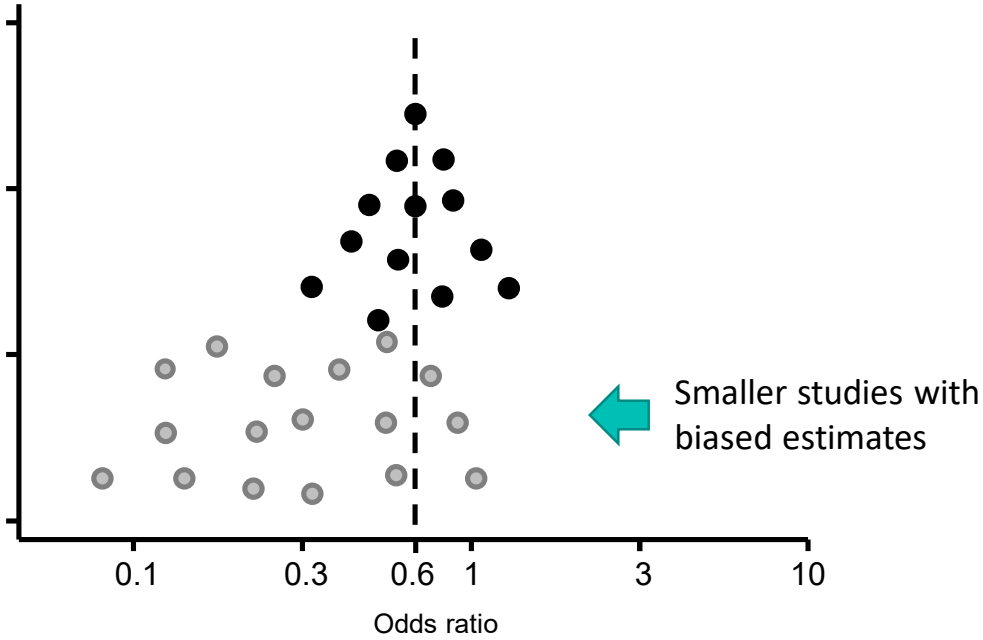
bris

What's the problem?

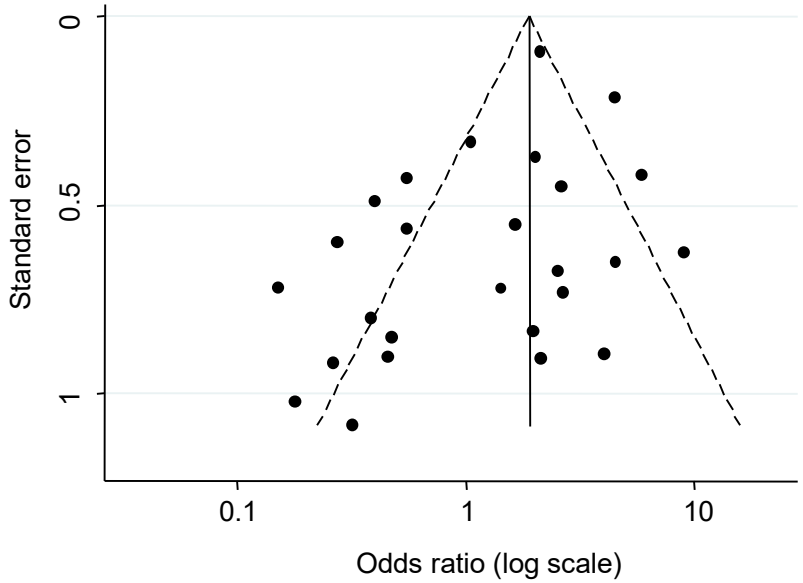
Egger's test is *not* a test for publication bias

- Plausible reasons for funnel plot asymmetry
  - Publication bias
  - Selective outcome non-reporting
  - Poor methodological quality leading to spuriously inflated effects in smaller studies
  - True heterogeneity
  - Artefactual
  - Chance

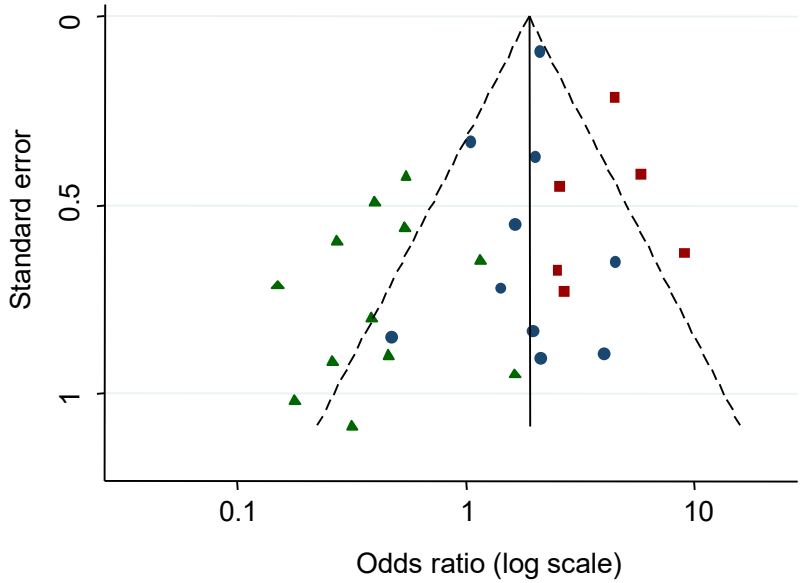
Bias due to poor quality of small trials

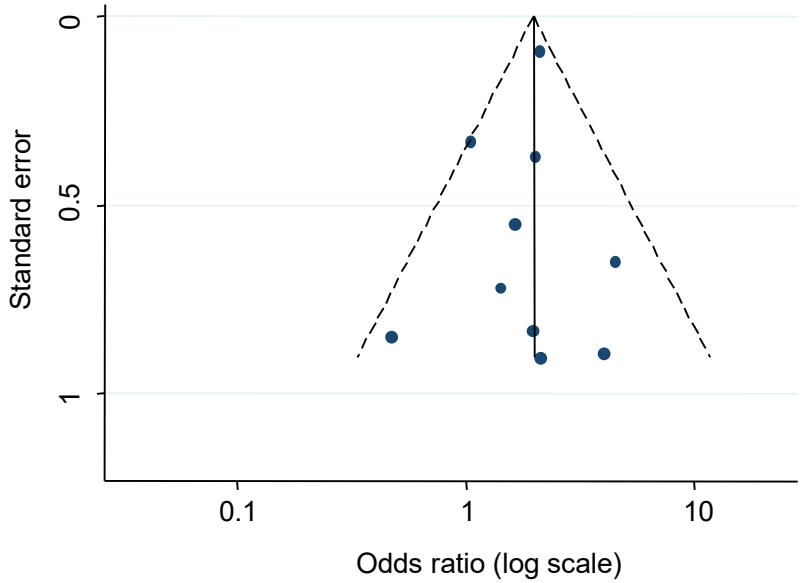


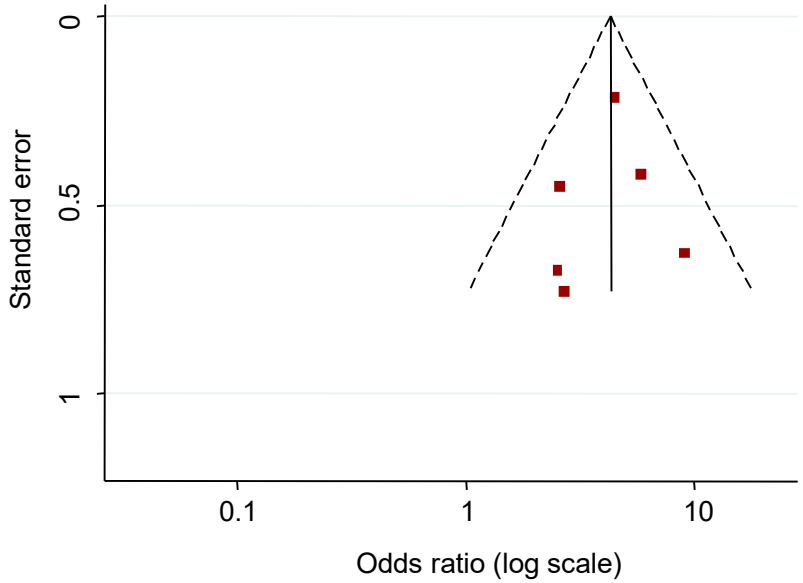
Funnel plot asymmetry?

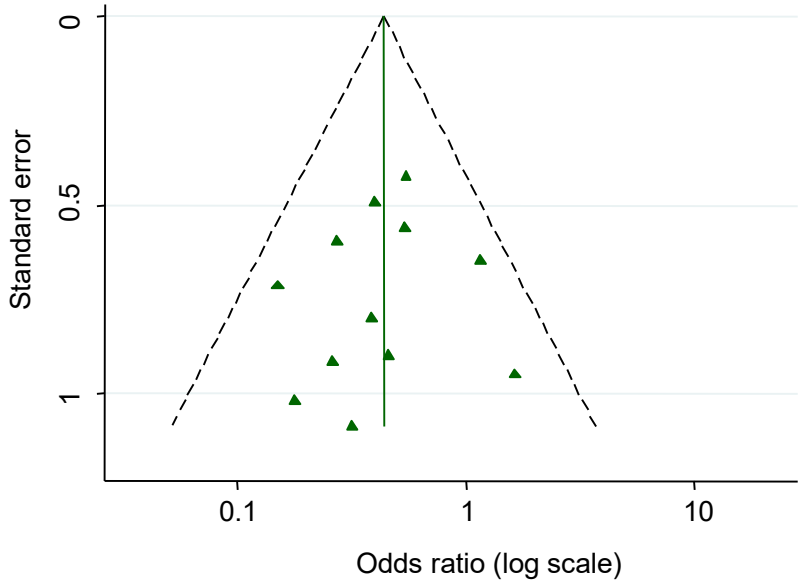


Funnel plot asymmetry?



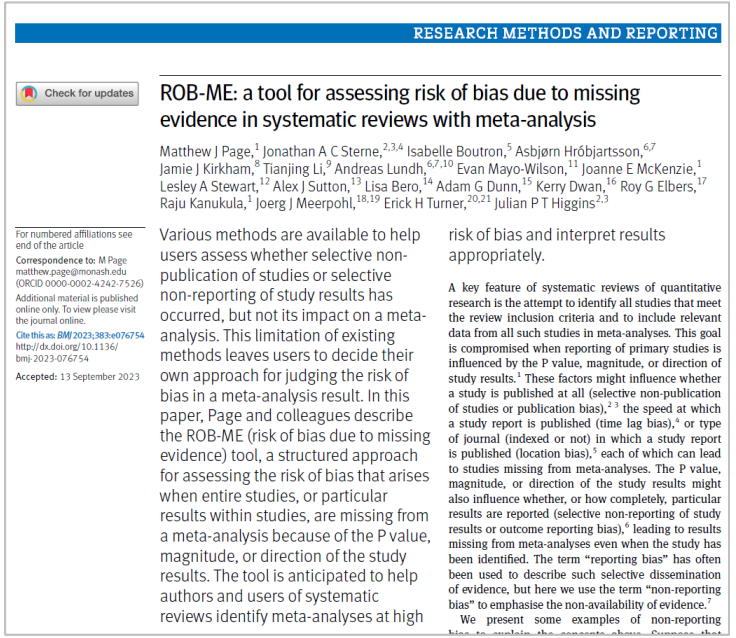

Standard error
0
0.5
1
0.1
1
10
Odds ratio (log scale)


Standard error
0
0.5
1
0.1
1
10
Odds ratio (log scale)


Standard error
0
0.5
1
0.1
1
10
Odds ratio (log scale)

What's the solution?

- Assessment of publication bias should focus on
  - likelihood of outcome non-reporting
    - some detective work can help
  - likelihood of non-reporting of studies
    - whether studies likely to be **registered**, **published** and **identified**
  - taking account of the nature of the studies



RESEARCH METHODS AND REPORTING

Check for updates

ROB-ME: a tool for assessing risk of bias due to missing evidence in systematic reviews with meta-analysis

Matthew J Page,[1] Jonathan A C Sterne,[2,3,4] Isabelle Boutron,[5] Asbjørn Hróbjartsson,[6,7] Jamie J Kirkham,[8] Tianjing Li,[9] Andreas Lundh,[6,7,10] Evan Mayo-Wilson,[11] Joanne E McKenzie,[1] Lesley A Stewart,[12] Alex J Sutton,[13] Lisa Bero,[14] Adam G Dunn,[15] Kerry Dwan,[16] Roy G Elbers,[17] Raju Kanukula,[1] Joerg J Meerpohl,[18,19] Erick H Turner,[20,21] Julian P T Higgins[2,3]

For numbered affiliations see end of the article

Correspondence to: M Page matthew.page@monash.edu (ORCID 0000-0002-4242-7526)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2023;383:e076754 http://dx.doi.org/10.1136/bmj-2023-076754

Accepted: 13 September 2023

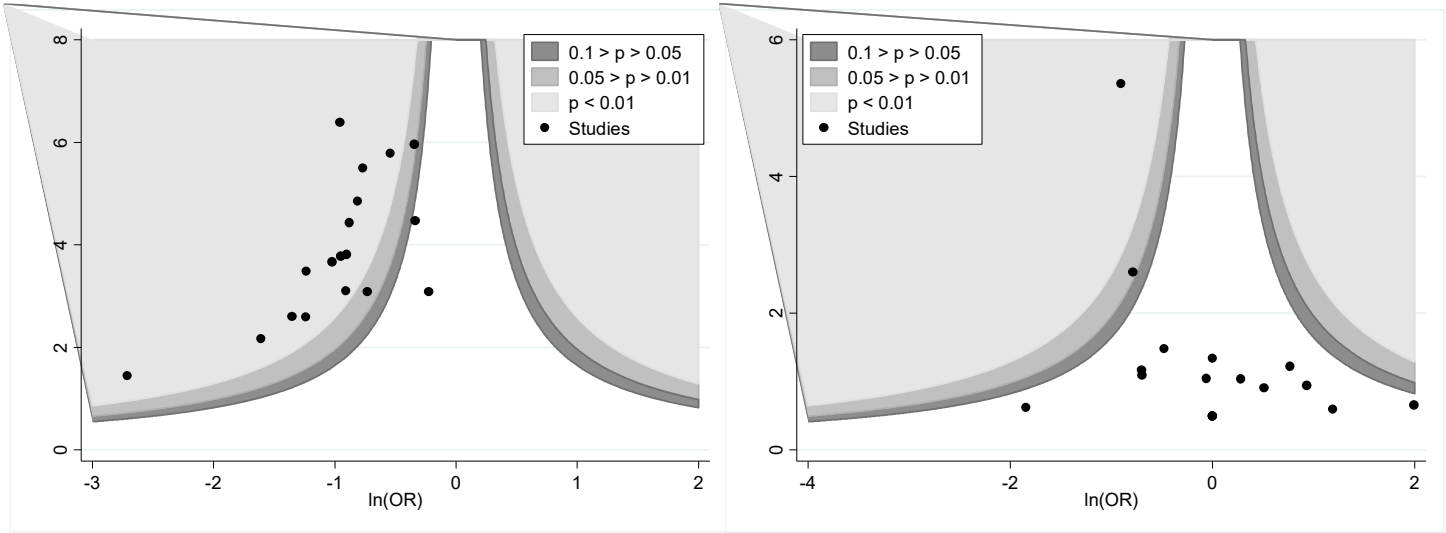
Various methods are available to help users assess whether selective non-publication of studies or selective non-reporting of study results has occurred, but not its impact on a meta-analysis. This limitation of existing methods leaves users to decide their own approach for judging the risk of bias in a meta-analysis result. In this paper, Page and colleagues describe the ROB-ME (risk of bias due to missing evidence) tool, a structured approach for assessing the risk of bias that arises when entire studies, or particular results within studies, are missing from a meta-analysis because of the P value, magnitude, or direction of the study results. The tool is anticipated to help authors and users of systematic reviews identify meta-analyses at high risk of bias and interpret results appropriately.

A key feature of systematic reviews of quantitative research is the attempt to identify all studies that meet the review inclusion criteria and to include relevant data from all such studies in meta-analyses. This goal is compromised when reporting of primary studies is influenced by the P value, magnitude, or direction of study results.[1] These factors might influence whether a study is published at all (selective non-publication of studies or publication bias),[2,3] the speed at which a study report is published (time lag bias),[4] or type of journal (indexed or not) in which a study report is published (location bias),[5] each of which can lead to studies missing from meta-analyses. The P value, magnitude, or direction of the study results might also influence whether, or how completely, particular results are reported (selective non-reporting of study results or outcome reporting bias),[6] leading to results missing from meta-analyses even when the study has been identified. The term "reporting bias" has often been used to describe such selective dissemination of evidence, but here we use the term "non-reporting bias" to emphasise the non-availability of evidence.[7]

We present some examples of non-reporting

What's the solution?

- If using funnel plots, use contour-enhanced versions
  - contours describe levels of statistical significance



Concluding remarks

- I have reviewed some of the most common errors I have observed in
  - reporting systematic reviews
  - identifying heterogeneity
  - dealing with heterogeneity
  - assessing publication bias
- … and tried to offer solutions
- I was astonished and disappointed that every time I looked at the most recent citation (Scopus), I found the method *misused*
- Being highly cited is nice, but it doesn't necessarily make you an influential researcher