McGill

An Introduction to Systematic Reviews and Meta-Analyses

(Sometimes)

James Brophy, MD,PhD

Professor
Departments Medicine (Cardiology) and Epidemiology, Biostatistics, and Occupational Health
McGill University

Conflicts of Interest

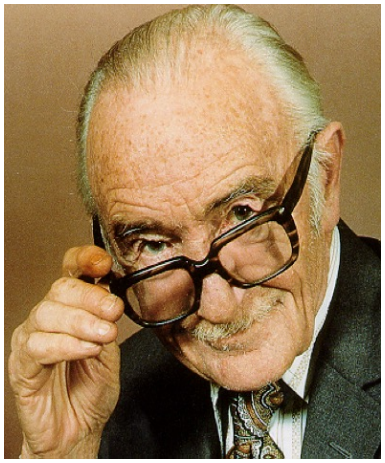
- None

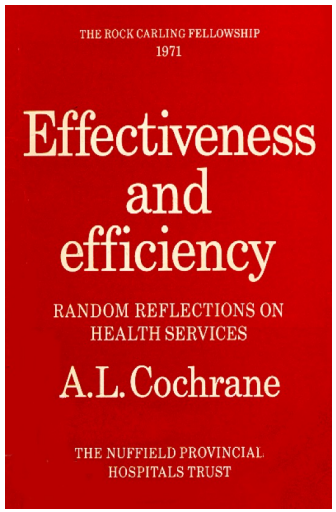
Learning Objectives

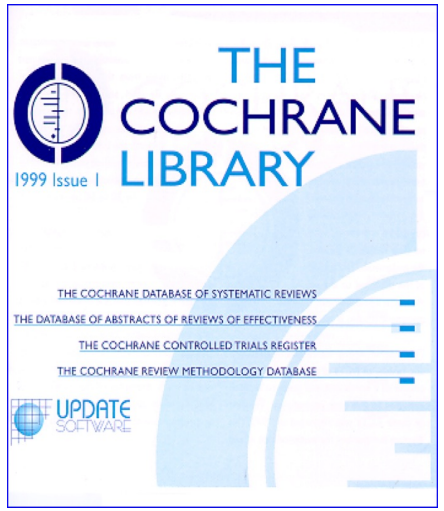
1. Explain the rationale for conducting a systematic review and/or meta-analysis. (MDCM 7.4)
2. Differentiate between narrative reviews, systematic reviews, and meta-analyses. (MDCM 7.4)
3. Describe the key components of a systematic review and meta-analysis. (MDCM 7.4)

History

- MA term first used by Glass (1976). "Primary, secondary, and meta-analysis of research" Educational Researcher 5 (10): 3–8
- Basic idea > 100 years prior (cf Pearson K (1904). "Report on certain enteric fever inoculation statistics". BMJ 3 (2288): 1243–1246)
- Also much influenced by 1971 publication



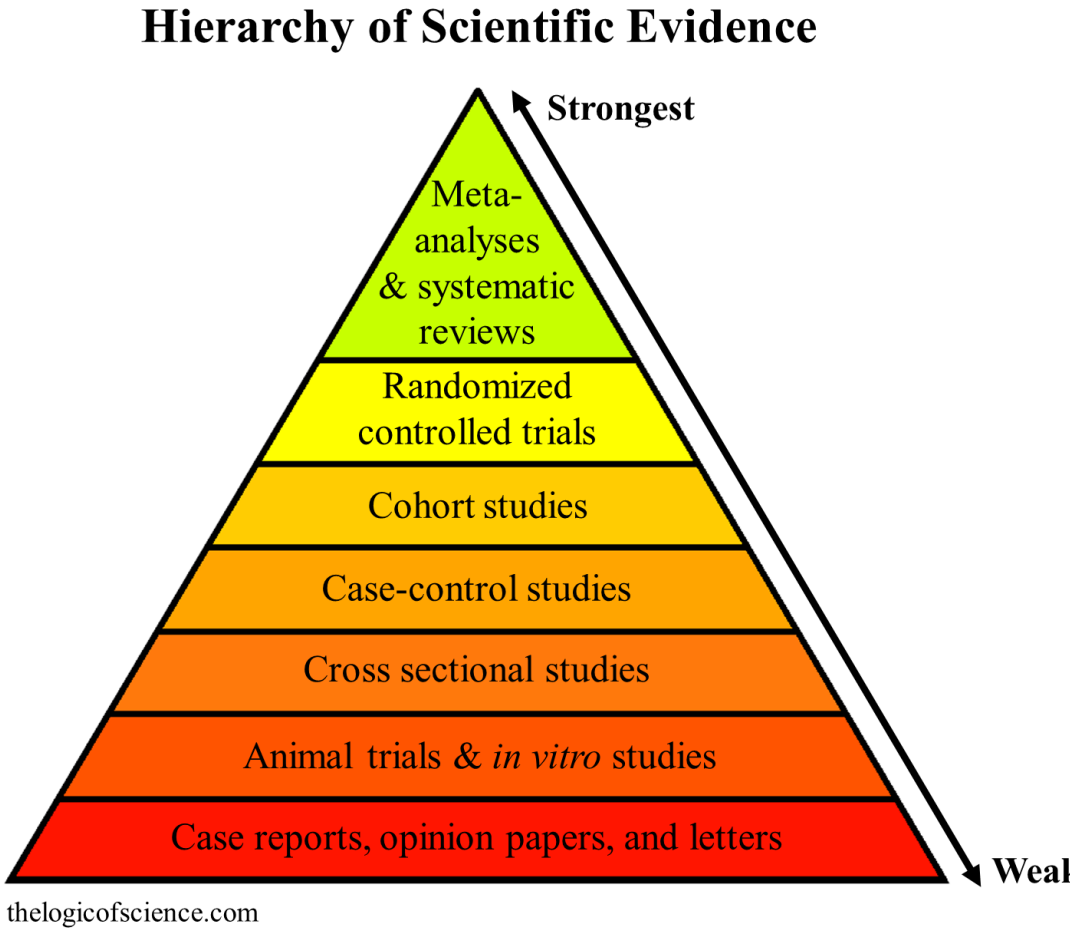




- As science is cumulative, scientists should cumulate scientifically
- Harder than it seems - "Il est plus aisé de dire des choses nouvelles que de concilier celles qui ont étés dites". Luc de Clapiers Vauvenarques, 1715-47 (Réflexions et Maximes)

Expert opinion

» “I think it is preferable to accustom a baby to sleeping on his stomach from the start if he is willing. He may change later when he learns to turn over.”-Spock (1958)



The evidence

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology
doi:10.1093/ije/dyi088

Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002

Ruth Gilbert,[1]* Georgia Salanti,[2] Melissa Harden[1] and Sarah See[1,3]

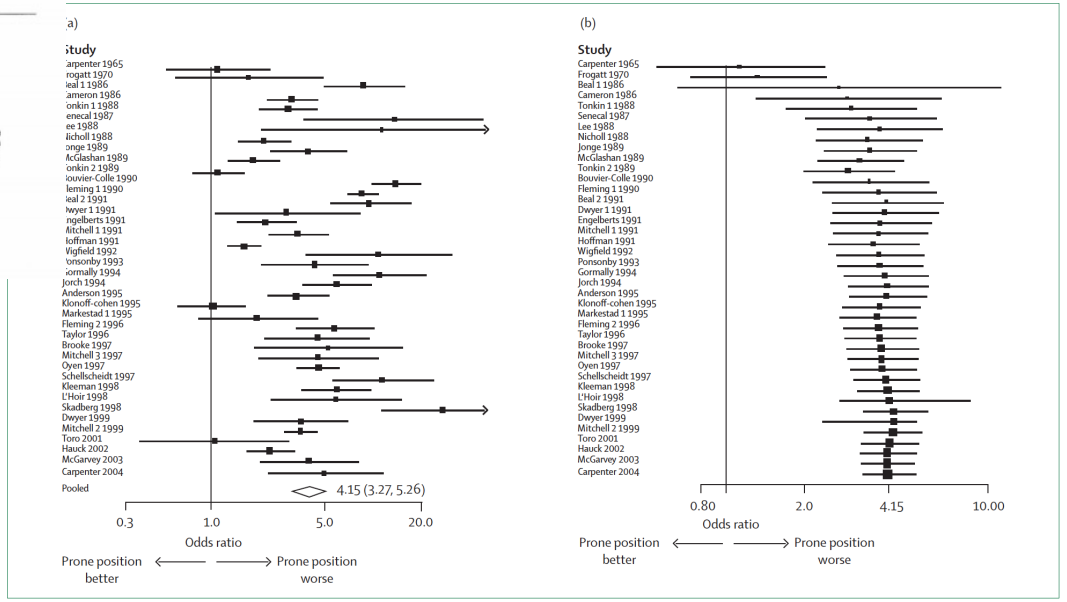
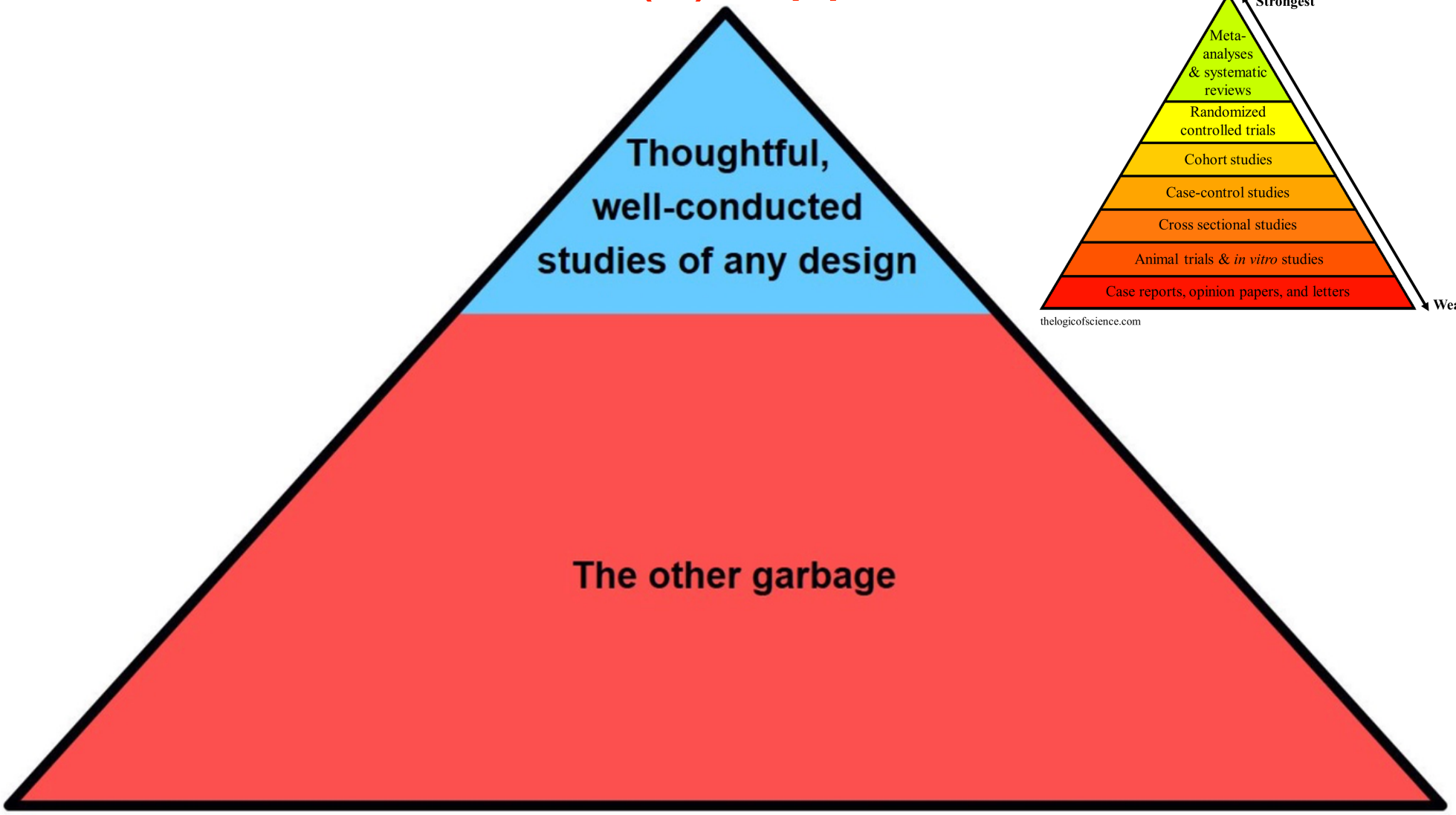


Figure 2.4: (a) Odds ratios and pooled odds ratio and (b) cumulative odds ratios for front versus non-front sleeping position in comparisons of cases of sudden infant death syndrome (SIDS) and controls (Reproduced from Gilbert et al.,[73] with permission of Oxford University Press and the International Epidemiological Association)

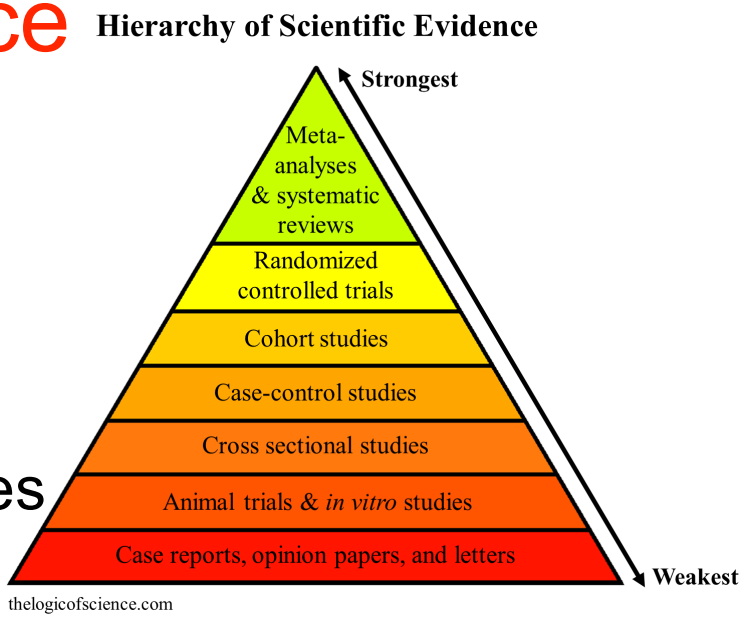
"Advice to put infants to sleep on the front for nearly half a century was **contrary to evidence** available from 1980s that this was likely to be **harmful**. Systematic review of preventable risk factors for SIDS from 1980 would have led to **earlier** recognition of the risks of sleeping on the front and might have **prevented over 10 000 infant deaths in the UK** and at least 50 000 in Europe the USA and Australasia."

A better (?) approach



Adapted from: https://twitter.com/statsepi/status/895012576714731520

Levels of evidence

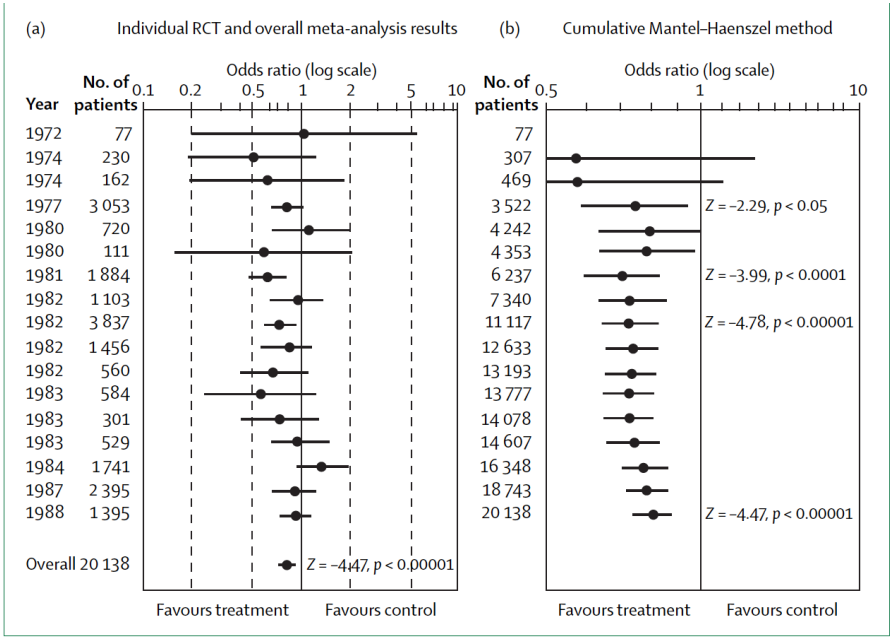


1. Systematic review (double-blind) RCTs
2. Individual (double-blind) RCTs
3. ≧ 1 well-conducted (large) cohort studies
4. ≧ 1well-conducted case-control studies
5. A dramatic uncontrolled experiment
6. Expert committee sitting in review; peer opinion leader
7. Personal experience (anecdotes

Basic premise: Results of a particular research study cannot be interpreted with any confidence unless they have been considered, systematically, together with the results of other studies addressing the same or similar questions. How well is this premise reflected in papers published in major general medical journals?

Need for evidence synthesis

» Antman et al. (JAMA 1992) cumulative meta-analysis of RCTs assessing beta blockers on secondary heart attacks (**n=20138**)



» Chance likely could have been ruled out as early as 1981, after only 6 trials and 6237 patients randomized.

» Scientifically and ethically questionable whether nearly 14,000 additional patients needed to participate in further such studies.

Background

- **Challenge** find the common truth behind all conceptually similar scientific studies that have been measured with a certain error
- **Systematic review (SR)**: A critical, impartial, systematic assessment and evaluation of research attempts to answer a focused clinical question without bias
- **Meta-analysis** - **statistical analysis** that combines independent clinical trials considered by the analyst to be "**combinable**" by weighted averages and specified error estimate distributions
- **Distinction between SR and MA** - SR always appropriate and desirable but MA may often be inappropriate, or even misleading
- **Meta-analysis** a**ims** to find the unknown common truth to reduce bias, improve precision, ultimately enhance knowledge -> better individual and population health decisions & outcomes

Essay

Doing New Research? Don't Forget the Old

Nobody should do a trial without reviewing what is known

Mike Clarke

On May 2, 1898, George Gould used his address to the founding meeting of the Association of Medical Librarians in Philadelphia to present a vision of the future of health information. 'I look forward,' he said, 'to such an organisation of the literary records of medicine that a puzzled worker in any part of the civilised world shall in an hour be able to gain a knowledge pertaining to a subject of the experience of every other man in the world' [1]. Has his vision been realised?

good quality, but some of it is not. Thus, anyone wishing to use the health literature to make well-informed decisions must both identify the relevant research from amidst this vast amount of information and then appraise it. This is an impossible task for many. Even though making access to the literature easier and cheaper will increase the ability of people to find research, it will also reveal just how much information there is out there and how daunting is the task of making sense of it.

with one or more search engines? Almost certainly, as the speed of the search increased through these four

Citation: Clarke M (2004) Doing new research? Don't forget the old. PLoS Med 1(2): e35.

Mike Clarke is dir Cochrane Centre, mclarke@cochra

Box 1. Practical Suggestions for Researchers

- Conduct a systematic review of your research question before embarking on a new study, or identify a relevant review done by someone else.
- Design your study to take account of the relevant successes and failures of the prior studies, and of the evidence within them.
- Discuss the findings of your study in the context of an updated systematic review of relevant research.
- Publish the systematic review within, alongside, or shortly after the report of your study.
- Provide information from your study to others doing systematic reviews of similar topics.

Clarke M. PLoS Med 2004

Researchers do not behave systematically

Classification	May 1997 (*n* = 26)	May 2001 (*n* = 33)	May 2005 (*n* = 18)
First trial addressing the question	1	3	3
Contained an updated systematic review integrating the new results	2	0	0
Discussed a previous review but did not attempt to integrate the new results	4	3	5
No apparent systematic attempt to set the results in the context of other trials	19	27	10

**The Lancet, New England Journal of Medicine, BMJ, JAMA* and *Annals of Internal Medicine*. Data from Clarke and co-workers.[49–51]

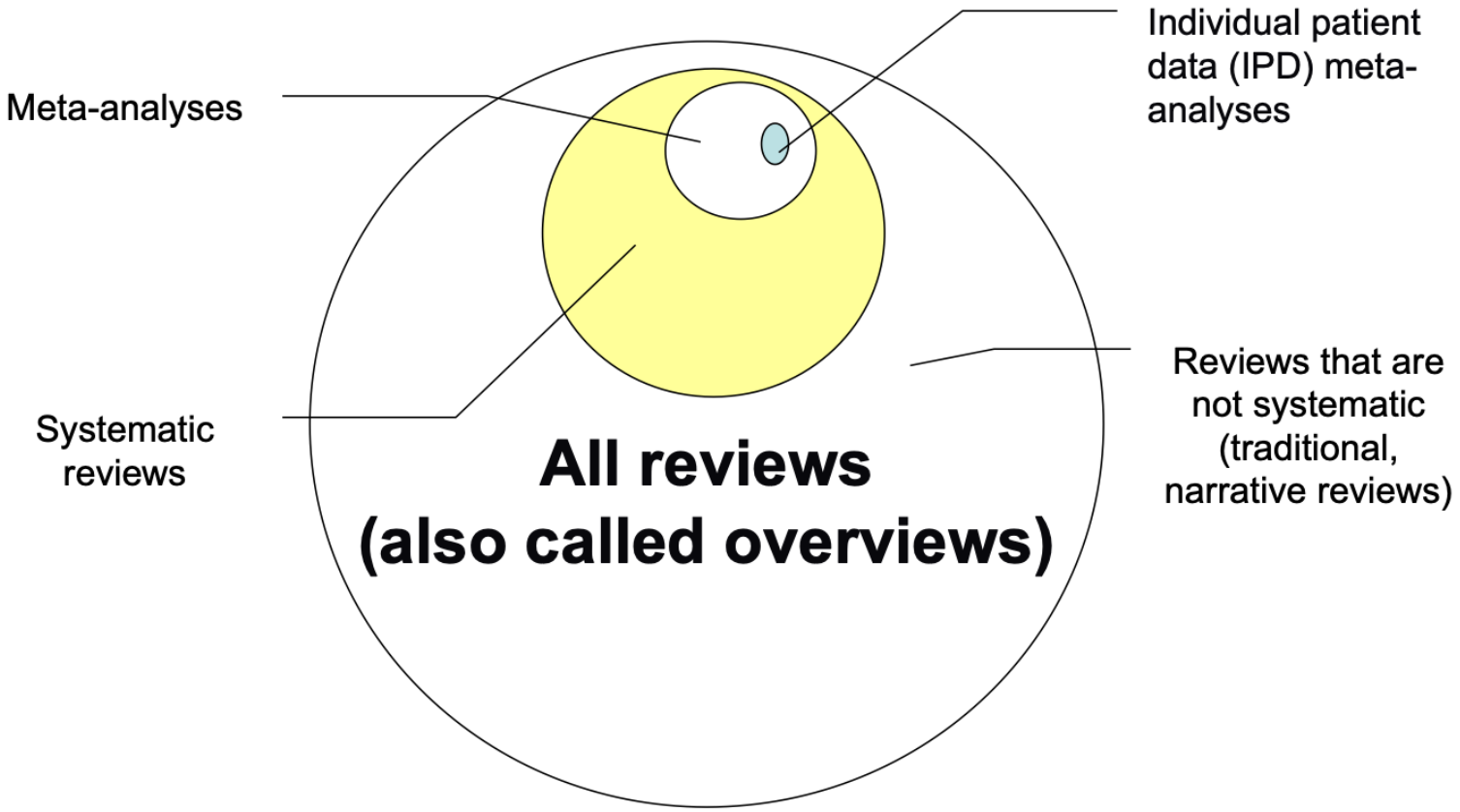
Table 2.1: **Classification of discussion sections in reports of randomised controlled trials published in May 1997, May 2001 and May 2005 in five general medical journals***

<u>Conclusions</u> No evidence of progress between 1997 and 2005 in the proportion of published trials which discussed new results within the context of up-to-date systematic reviews of relevant evidence from other controlled trials.

Most researchers appear not to have considered a systematic review when designing or discussing their trial.

Chalmers (2007)

Types of Review Articles



•A meta-analysis is the statistical combination of at least 2 studies to produce a single estimate of the effect of the healthcare intervention under consideration."

•Individual patient data meta-analyses (pooled analyses) involves obtaining raw data on all patients from each of the trials directly and then re-analyzing them.

Pai M, et al. Natl Med J India 2004.

Getting started - importance of a protocol

1. Develop a documented peer reviewed protocol addressing
   A. Specify Problem
   B. Search for and Identify Studies (electronic & hand search, ask a librarian)
   C. Enter studies into database
   D. Select Studies for Review
   E. Review Studies
   F. Develop Coding Scheme
   G. Abstract / Code Studies
   H. Select Effect Size Statistic
   I. Transform and Weight Effect Sizes
   J. Assess heterogeneity
   K. Assess Bias
   L. Synthesize and Present Results

PRISMA checklist - to improve quality

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.	

PRISMA checklist - to improve quality

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.	

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

Quality Assessment - RCTs

Table 1 | Version 2 of the Cochrane risk-of-bias assessment tool for randomised trials: bias domains, signalling questions, response options, and risk-of-bias judgments

Bias domain and signalling question*	Response options: Lower risk of bias	Higher risk of bias	Other
Bias arising from the randomisation process			
1.1 Was the allocation sequence random?	Y/PY	N/PN	NI
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Y/PY	N/PN	NI
1.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?	N/PN	Y/PY	NI
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the predicted direction of bias arising from the randomisation process?			
Bias due to deviations from intended interventions			
2.1 Were participants aware of their assigned intervention during the trial?	N/PN	Y/PY	NI
2.2 Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	N/PN	Y/PY	NI
2.3 If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the trial context?			
2.4 If Y/PY/NI to 2.3: Were these deviations likely to have affected the outcome?			
2.5 If Y/PY to 2.4: Were these deviations from intended intervention balanced between groups?			
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?			
2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?			
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the predicted direction of bias due to deviations from intended interventions?			
Bias due to missing outcome data			
3.1 Were data for this outcome available for all, or nearly all, participants randomised?			
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?			
3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?			
3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?			
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the predicted direction of bias due to missing outcome data?			
Bias in measurement of the outcome			
4.1 Was the method of measuring the outcome inappropriate?	N/PN	Y/PY	NI
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	N/PN	Y/PY	NI
4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	N/PN	Y/PY	NI
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	N/PN	Y/PY	NA/NI
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	N/PN	Y/PY	NA/NI
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the predicted direction of bias in measurement of the outcome?			
Bias in selection of the reported result			
5.1 Were the data that produced this result analysed in accordance with a prespecified analysis plan that was finalised before unblinded outcome data were available for analysis?	Y/PY	N/PN	NI
Is the numerical result being assessed likely to have been selected, on the basis of the results, from:			
5.2 ... multiple eligible outcome measurements (eg, scales, definitions, time points) within the outcome domain?	N/PN	Y/PY	NI
5.3 ... multiple eligible analyses of the data?	N/PN	Y/PY	NI
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the predicted direction bias due to selection of the reported results?			
Overall bias			
Risk-of-bias judgment (low/high/some concerns)			
Optional: What is the overall predicted direction of bias for this outcome?			

Y=yes; PY=probably yes; PN=probably no; N=no; NA=not applicable; NI=no information.
*Signalling questions for bias due to deviations from intended interventions relate to the effect of assignment to intervention.

Cochrane Risk of Bias (RoB) 2.0 Tool

Sterne et al. BMJ 2019

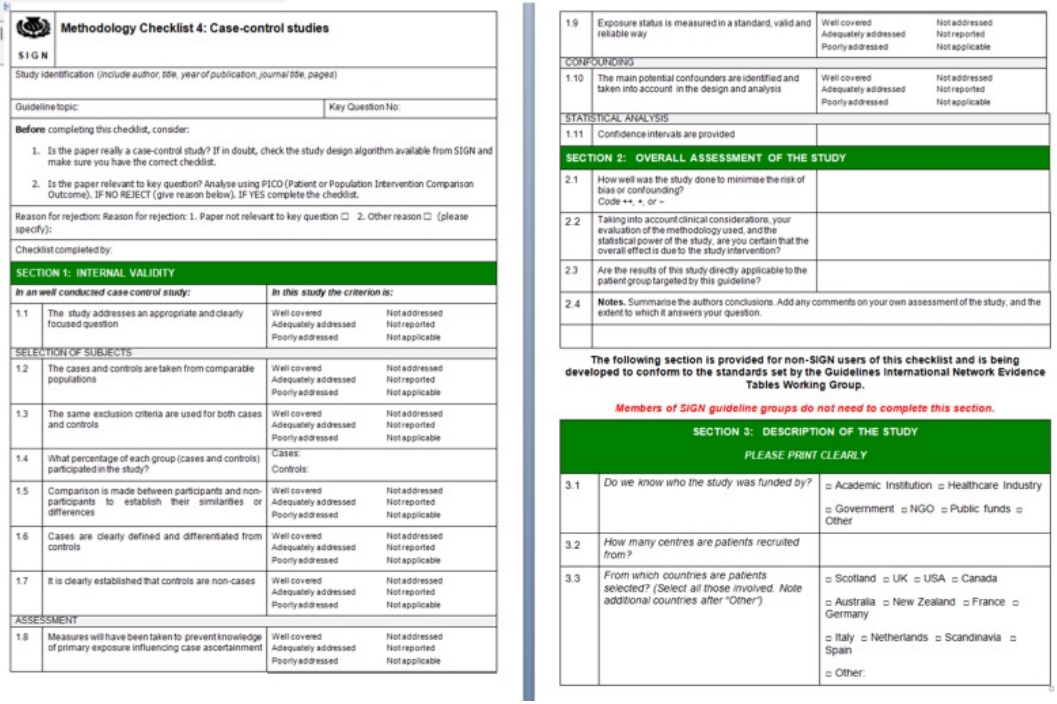
Quality Assessment – Observational Studies

Table 1 | Bias domains included in ROBINS-I

Domain	Explanation
Pre-intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias due to confounding	Baseline confounding occurs when one or more prognostic variables (factors that predict the outcome of interest) also predicts the intervention received at baseline ROBINS-I can also address time-varying confounding, which occurs when individuals switch between the interventions being compared and when post-baseline prognostic factors affect the intervention received after baseline
Bias in selection of participants into the study	When exclusion of some eligible participants, or the initial follow-up time of some participants, or some outcome events is related to both intervention and outcome, there will be an association between interventions and outcome even if the effects of the interventions are identical This form of selection bias is distinct from confounding—A specific example is bias due to the inclusion of prevalent users, rather than new users, of an intervention
At intervention	**Risk of bias assessment is mainly distinct from assessments of randomised trials**
Bias in classification of interventions	Bias introduced by either differential or non-differential misclassification of intervention status Non-differential misclassification is unrelated to the outcome and will usually bias the estimated effect of intervention towards the null Differential misclassification occurs when misclassification of intervention status is related to the outcome or the risk of the outcome, and is likely to lead to bias
Post-intervention	**Risk of bias assessment has substantial overlap with assessments of randomised trials**
Bias due to deviations from intended interventions	Bias that arises when there are systematic differences between experimental intervention and comparator groups in the care provided, which represent a deviation from the intended intervention(s) Assessment of bias in this domain will depend on the type of effect of interest (either the effect of assignment to intervention or the effect of starting and adhering to intervention).
Bias due to missing data	Bias that arises when later follow-up is missing for individuals initially included and followed (such as differential loss to follow-up that is affected by prognostic factors); bias due to exclusion of individuals with missing information about intervention status or other variables such as confounders
Bias in measurement of outcomes	Bias introduced by either differential or non-differential errors in measurement of outcome data. Such bias can arise when outcome assessors are aware of intervention status, if different methods are used to assess outcomes in different intervention groups, or if measurement errors are related to intervention status or effects
Bias in selection of the reported result	Selective reporting of results in a way that depends on the findings and prevents the estimate from being incl synthesis)

ROBINS-I

SIGN



SIGN

Methodology Checklist 4: Case-control studies

Study identification (*include author, title, year of publication, journal title, pages*)

Guideline topic: | Key Question No:

Before completing this checklist, consider:

1. Is the paper really a case-control study? If in doubt, check the study design algorithm available from SIGN and make sure you have the correct checklist.
2. Is the paper relevant to key question? Analyse using PICO (Patient or Population Intervention Comparison Outcome). If NO REJECT (give reason below). IF YES complete the checklist.

Reason for rejection: Reason for rejection: 1. Paper not relevant to key question ☐ 2. Other reason ☐ (please specify):

Checklist completed by:

SECTION 1: INTERNAL VALIDITY

	In an well conducted case control study:	*In this study the criterion is:*
1.1	The study addresses an appropriate and clearly focused question	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
	SELECTION OF SUBJECTS	
1.2	The cases and controls are taken from comparable populations	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
1.3	The same exclusion criteria are used for both cases and controls	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
1.4	What percentage of each group (cases and controls) participated in the study?	Cases: Controls:
1.5	Comparison is made between participants and non-participants to establish their similarities or differences	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
1.6	Cases are clearly defined and differentiated from controls	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
1.7	It is clearly established that controls are non-cases	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
	ASSESSMENT	
1.8	Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
1.9	Exposure status is measured in a standard, valid and reliable way	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
	CONFOUNDING	
1.10	The main potential confounders are identified and taken into account in the design and analysis	Well covered / Adequately addressed / Poorly addressed / Not addressed / Not reported / Not applicable
	STATISTICAL ANALYSIS	
1.11	Confidence intervals are provided	

SECTION 2: OVERALL ASSESSMENT OF THE STUDY

2.1	How well was the study done to minimise the risk of bias or confounding? *Code ++, +, or –*	
2.2	Taking into account clinical considerations, your evaluation of the methodology used, and the statistical power of the study, are you certain that the overall effect is due to the study intervention?	
2.3	Are the results of this study directly applicable to the patient group targeted by this guideline?	
2.4	**Notes.** Summarise the authors conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question.	

The following section is provided for non-SIGN users of this checklist and is being developed to conform to the standards set by the Guidelines International Network Evidence Tables Working Group.

Members of SIGN guideline groups do not need to complete this section.

SECTION 3: DESCRIPTION OF THE STUDY

PLEASE PRINT CLEARLY

3.1	*Do we know who the study was funded by?*	☐ Academic Institution ☐ Healthcare Industry ☐ Government ☐ NGO ☐ Public funds ☐ Other
3.2	*How many centres are patients recruited from?*	
3.3	*From which countries are patients selected? (Select all those involved. Note additional countries after "Other")*	☐ Scotland ☐ UK ☐ USA ☐ Canada ☐ Australia ☐ New Zealand ☐ France ☐ Germany ☐ Italy ☐ Netherlands ☐ Scandinavia ☐ Spain ☐ Other:

www.cochrane.org

Sterne et al. BMJ 2016

http://www.sign.ac.uk/methodology/checklists.html

A bad example

International Journal of STD & AIDS 1999; **10**: 8–16

REVIEW ARTICLE

Circumcision and HIV infection: review of the literature and meta-analysis

R S Van Howe MD FAAP

Department of Pediatrics, Marshfield Clinic, Lakeland Center, USA

Summary: Thirty-five articles and a number of abstracts have been published in the medical literature looking at the relationship between male circumcision and HIV infection. Study designs have included geographical analysis, studies of high-risk patients, partner studies and random population surveys. Most of the studies have been conducted in Africa. A meta-analysis was performed on the 29 published articles where data were available. When the raw data are combined, a man with a circumcised penis is at greater risk of acquiring and transmitting HIV than a man with a non-circumcised penis (odds ratio (OR)=1.06, 95% confidence interval (CI)=1.01–1.12). Based on the studies published to date, recommending routine circumcision as a prophylactic measure to prevent HIV infection in Africa, or elsewhere, is scientifically unfounded.

$OR_{c:nc} = 1.06$
$OR_{nc:c} = 0.94$

- **Limitations** – not systematic, varying study designs combined (CC, cohort, Xsectional) & **faulty analysis** - simply added the numbers in each cell of the 2-by-2 tables. This ignores the size and variability in each study, and confounding.

General Challenges

- **Key challenge** not computational but cognitive
  - defining the question (particularly comparators and outcomes)
  - judging the validity and applicability of identified studies (PRISMA, MOOSE, Cochrane bias tool)
  - clinical heterogeneity, pool or not?, "**compare apples & oranges**"
  - appropriate analysis & data interpretation
- **Biases**
  - Missing studies (**publication bias**)
  - Quality of individual studies ('**GIGO**')
- **Requires effort, substance-area, & methodological expertise** to sensibly perform & discuss results
- Understand when to say **no** to meta-analysis (see **bold** above)

Reanalysis

Original authors reported circumcised risk 1.06(1.01-1.12) or 0.94 (0.89-0.99) risk for uncircumcised (this scale)

FARRELL AND EGGER. INTL J OF STD AND AIDS. 2000 -> COCHRANE REVIEW 2013

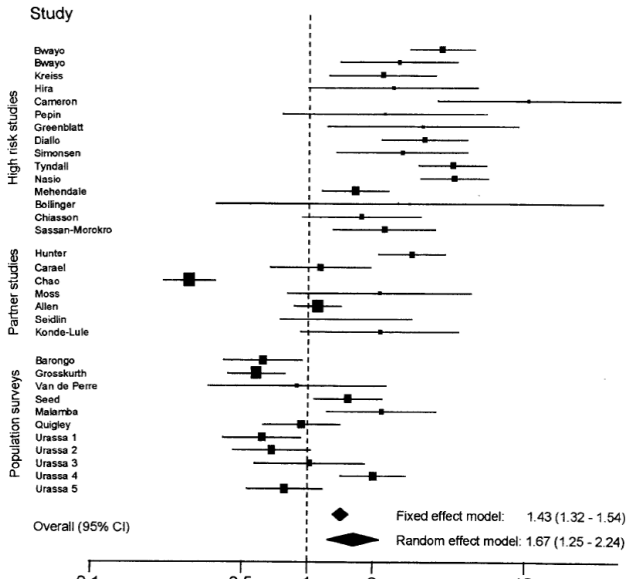


Table 1. Re-analysis of van Howe's data from 33 observational studies examining the association between circumcision and HIV infection. The results from meta-analyses using fixed and random effect models are compared to van Howe's results that were based on a simple pooling of the data

		Meta-analysis		Results presented by van Howe[1]	
Type of studies	No. of studies	Odds ratio (95% CI) (fixed effect model)	Odds ratio (95% CI) (random effect model)	Odds ratio (95% CI) (simple pooling of data)	*P* from test of heterogeneity
Studies in high-risk groups	15	2.97 (2.59–3.42)	3.00 (2.34–3.84)	1.18 (1.09–1.28)	<0.0001
Partner studies	7	0.99 (0.85–1.15)	1.29 (0.62–2.69)	1.42 (1.26–1.59)	<0.0001
Studies in general population groups	11	0.97 (0.85–1.09)	0.96 (0.71–1.30)	0.86 (0.77–0.97)	<0.0001
All studies	33	1.43 (1.32–1.54)	1.67 (1.25–2.24)	0.94 (0.89–0.99)	<0.0001

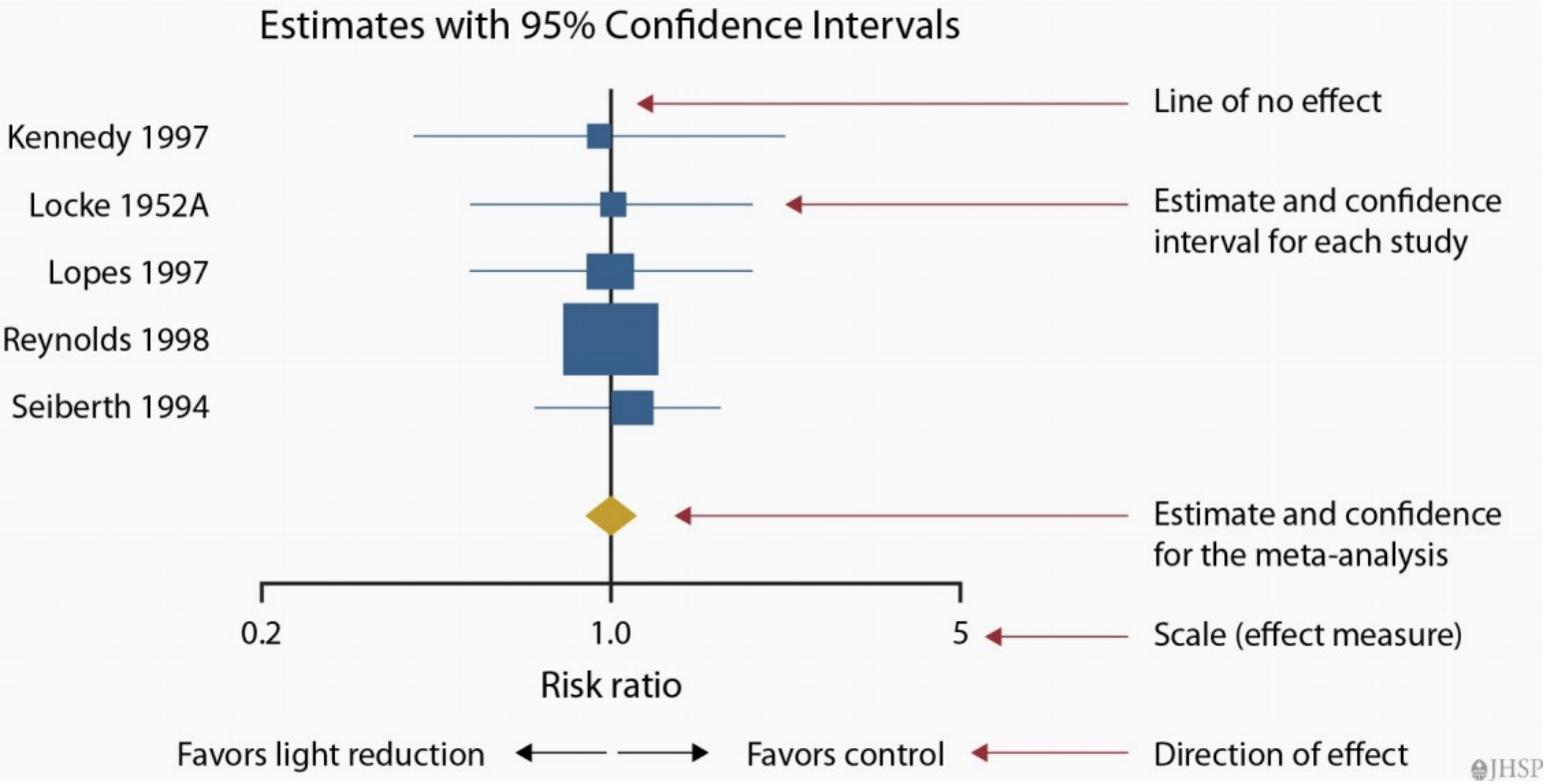
Combined odds ratios comparing the probability of HIV infection among men with intact foreskins with circumcised men are shown. An odds ratio above one thus indicates that lack of circumcision increases the risk of HIV infection. Conversely, an odds ratio below one indicates that an intact foreskin protects against HIV infection. CI=confidence interval

Meta-analysis: quantitative evidence synthesis

» Each “observation” is a study.

» To perform a meta-analysis we compute an effect size and variance for each study, and then compute a weighted mean of these effect sizes.

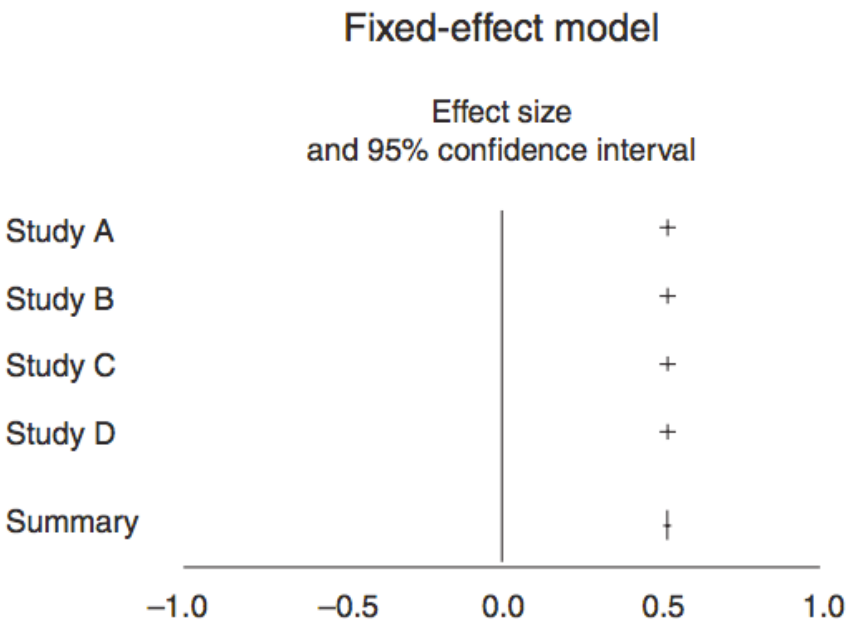
» To compute the weighted mean we generally assign more weight to the more precise studies, but the rules for assigning weights depend on our assumptions about the distribution of true effects.

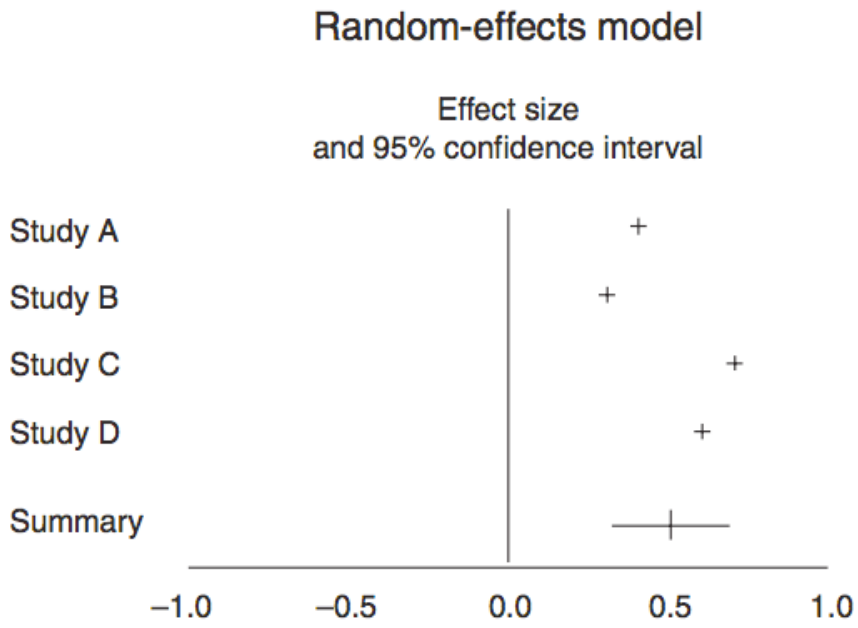


Meta-analysis: Fixed vs. Random Effects

» **Fixed Effect model:**
  – Assume studies are identical; goal is to compute the **common effect** size for the identified population and not to generalize to other populations.
  – Only source of uncertainty is the within-study (sampling or estimation) error
  – Methods include: inverse variance, Mantel Haenszel, Peto (different weights)

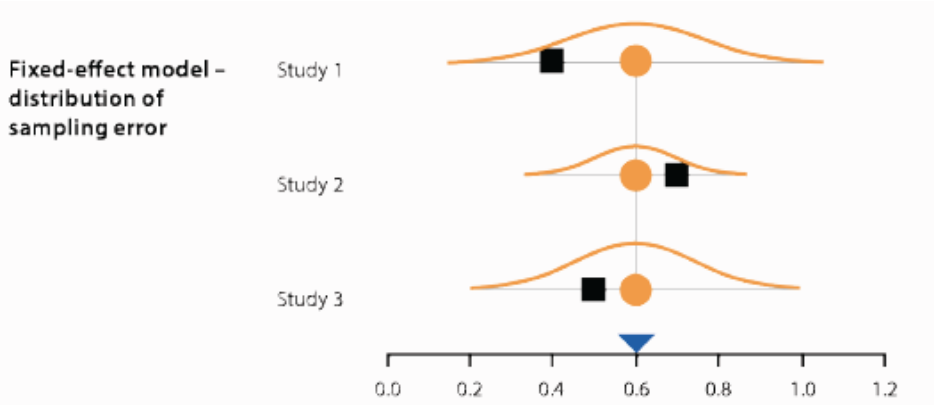
» **Random Effect model:**
  – Also incorporates additional between-study variation.
  – **Don't assume a common effect**, but estimates a mean of a distribution of a random sample of all possible studies that could have been included (includes both within and between study variation
  – Methods include: DerSimonian & Laird, Bayesian





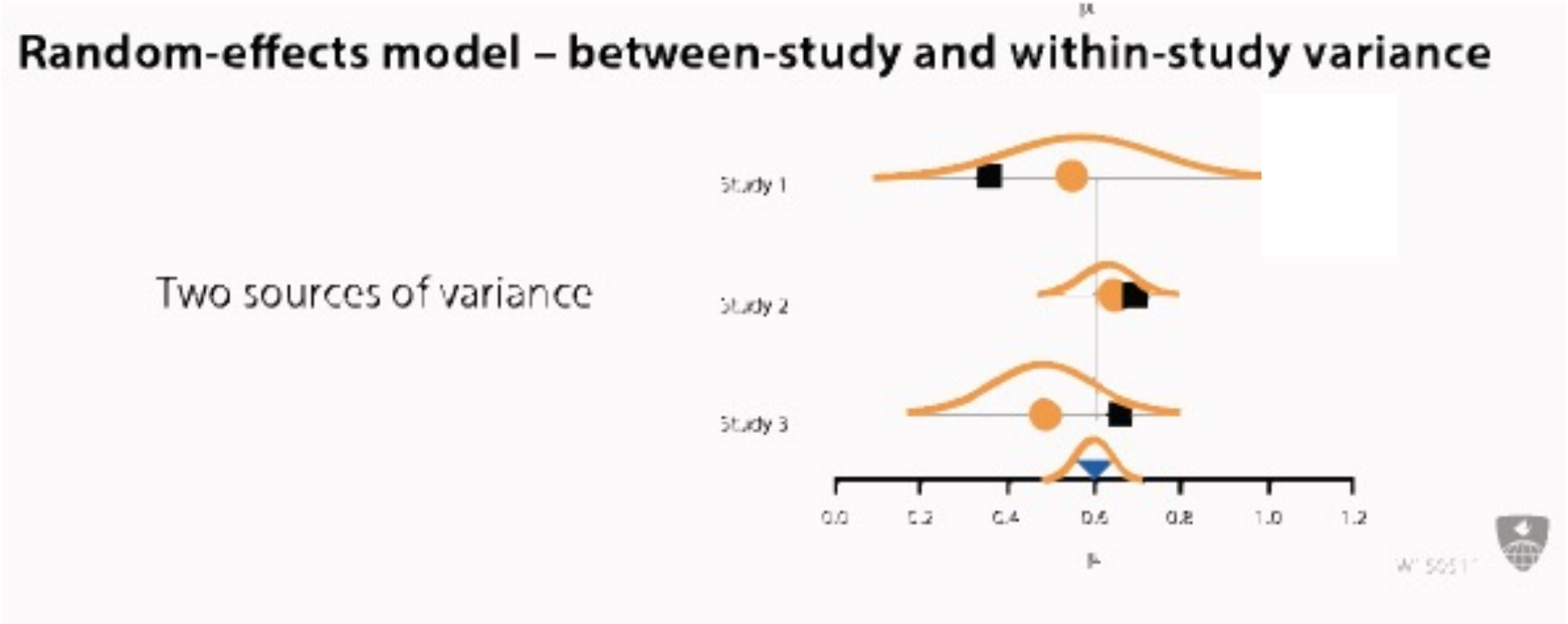
Example of fixed-effects meta-analysis

Study	Effect size Y	Variance Within V_Y
Saint	−0.366	0.185
Kelly	−0.288	0.290
Pilbeam	−0.384	0.156
Lane	−1.322	0.058
Wright	−0.417	0.282
Day	−0.159	0.160
Sum		



- Fixed effect assumes estimated effect comes from a single homogenous population, differences are sampling errors
- Meta-analytical result is our estimate of this common effect
- Each study average weighted by its precision AKA inverse variance (i.e bigger studies more precise, smaller variance get more weight)

Example of random-effects meta-analysis

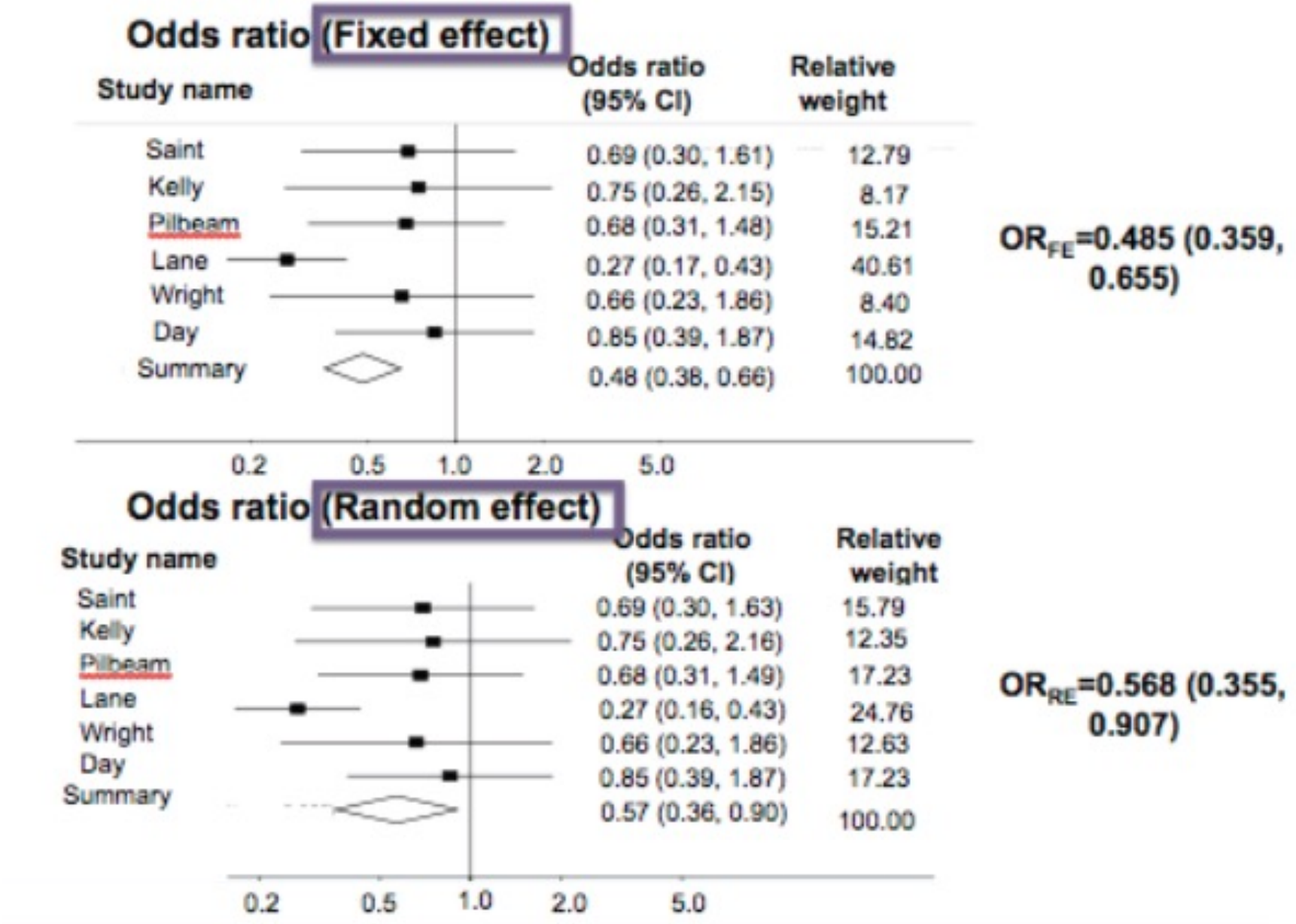


True effects in the studies are assumed to have been sampled from a distribution of true effects

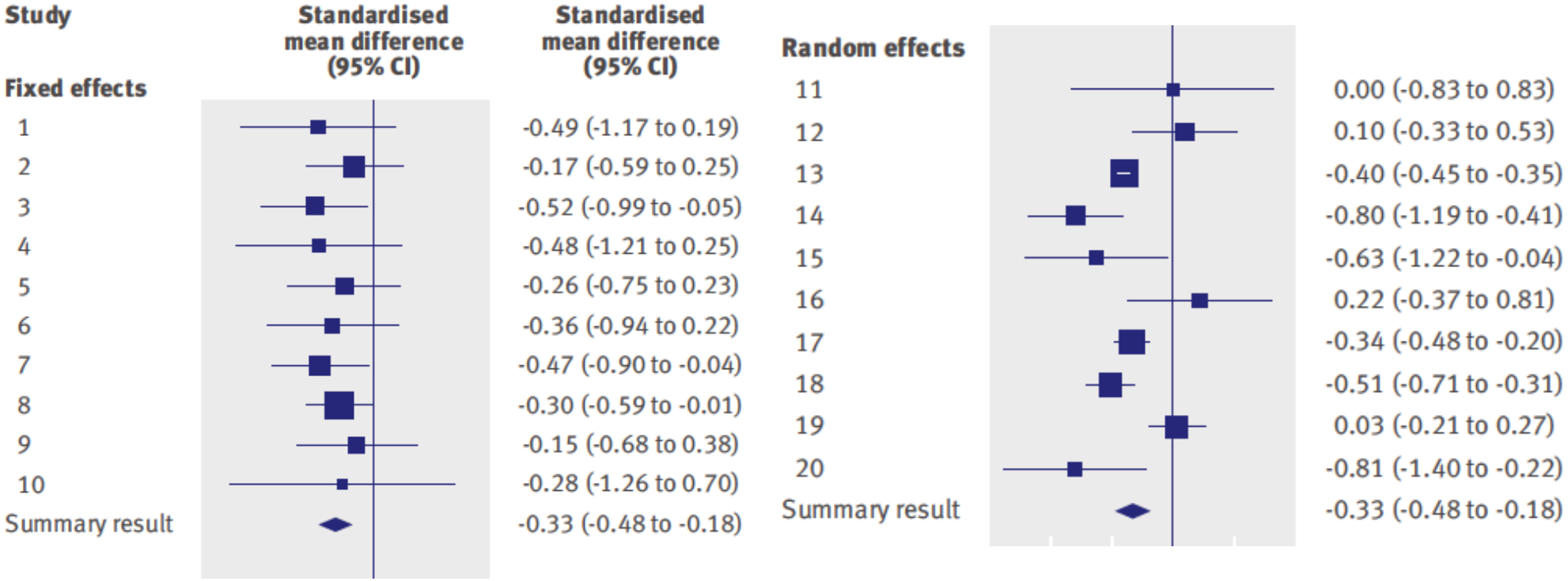
The summary effect is our estimate of the mean of all relevant true effects, and the null hypothesis is that the mean of these effects is 0 (or 1 for ratio)

The confidence interval for the random effects estimate indicates our uncertainty about the location of the center of the random effects distribution, not its width

Fixed vs. Random Effects



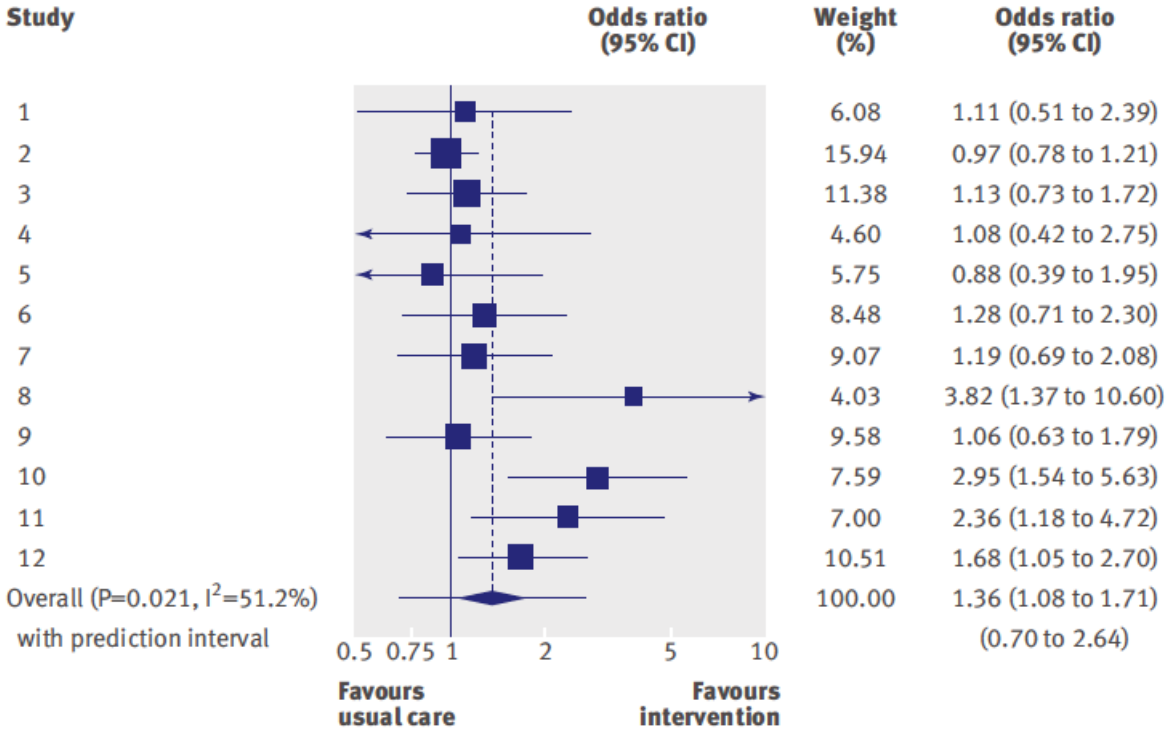
Difference between fixed & random effects models



- Plots of two distinct hypothetical meta-analyses -> same summary estimate (centre of diamond) and its 95% CI (width of diamond).
- **Fixed** effect MA (left) the summary result provided the best **estimate** of an **assumed common treatment effect**
- **Random** effects MA (right) the summary result gives the **average from the distribution of treatment effects across studies**

BMJ | 30 APRIL 2011 | VOLUME 342

Prediction intervals

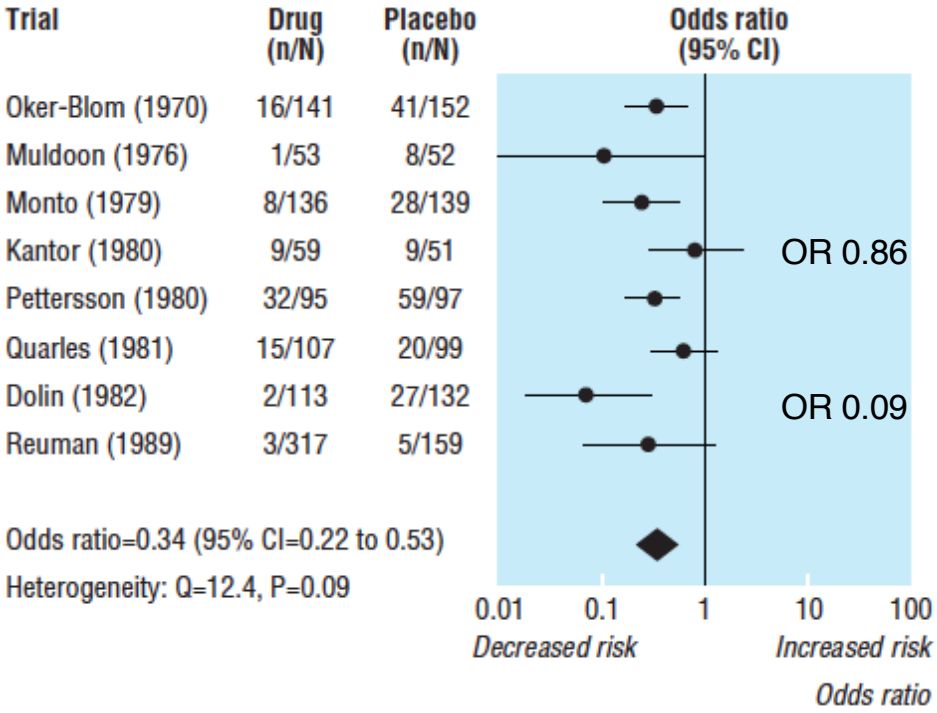


Random effects meta-analysis of 12 trials that examine the effect of inpatient rehabilitation designed for geriatric patients versus usual care on improving functional outcome (36% improvement)

- Prediction interval is centred at the summary estimate, and its width accounts for the uncertainty of the summary estimate, the estimate of between study standard deviation in the true treatment effects (often denoted by the Greek letter τ), and the uncertainty in the between study standard deviation estimate itself
- Provides more realistic accounting of the uncertainty

Testing for heterogeneity in MA

- Assessment of the consistency of effects across studies is essential
- **Heterogeneity tests null hypothesis all studies are evaluating the same effect**
- Usual test statistic (**Cochran's Q**) = $\sum w_i^* (ES_i - meanES)^2$
- The test has low power at with few studies and too sensitive with many studies.
- Heterogeneity is expected (diversity in doses, populations, etc) no point in simply testing for heterogeneity when what matters is the extent to which it affects the conclusions of the MA



The treatment effects in the eight trials seem inconsistent: the reduction in odds vary from 14% to 91%, but the test of heterogeneity yields a P value of 0.09, conventionally interpreted as non-significant.

I^2 statistic

- **I^2** is the percentage of observed total variation across studies that is due to real heterogeneity rather than chance (**preferred test statistic**)
- **$I^2 = 100\% \times (Q - df)/Q$**, where Q is Cochran's heterogeneity statistic and df the degrees of freedom (df= #ES-1).
- I^2 lies between 0% and 100%. A value of **0% indicates no observed heterogeneity**, and **larger values show increasing heterogeneity**

Advantages of I^2

- Focuses attention on the effect of any heterogeneity on the meta-analysis
- Interpretation is intuitive—the percentage of total variation across studies due to heterogeneity
- Can be accompanied by an uncertainty interval
- Simple to calculate and can usually be derived from published meta-analyses
- Does not inherently depend on the number of studies in the meta-analysis
- May be interpreted similarly irrespective of the type of outcome data (eg dichotomous, quantitative, or time to event) and choice of effect measure (eg odds ratio or hazard ratio)
- Wide range of applications

Summary points

Inconsistency of studies' results in a meta-analysis reduces the confidence of recommendations about treatment

Inconsistency is usually assessed with a test for heterogeneity, but problems of power can give misleading results

A new quantity I^2, ranging from 0-100%, is described that measures the degree of inconsistency across studies in a meta-analysis

I^2 can be directly compared between meta-analyses with different numbers of studies and different types of outcome data

I^2 is preferable to a test for heterogeneity in judging consistency of evidence

BMJ 2003;327:557–60

Testing for heterogeneity

Topic	Outcome/analysis	Effect measure	No of studies	Heterogeneity test			I^2 (95% uncertainty interval)*
				Q	df	P	
Tamoxifen for breast cancer[16]	Mortality	Peto odds ratio	55	55.9	54	0.40	3 (0 to 28)
Streptokinase after myocardial infarction[17]	Mortality	Odds ratio	33	39.5	32	0.17	19 (0 to 48)
Selective serotonin reuptake inhibitors for depression[13]	Drop-out	Odds ratio	135	179.9	134	0.005	26 (7 to 40)
Magnesium for acute myocardial infarction[18]	Death	Odds ratio	16	40.2	15	0.0004	63 (30 to 78)
Magnetic fields and leukaemia[19]	All studies	Odds ratio	6	15.9	5	0.007	69 (26 to 87)
Amantadine[11]	Prevention of influenza	Odds ratio	8	12.44	7	0.09	44 (0 to 75)

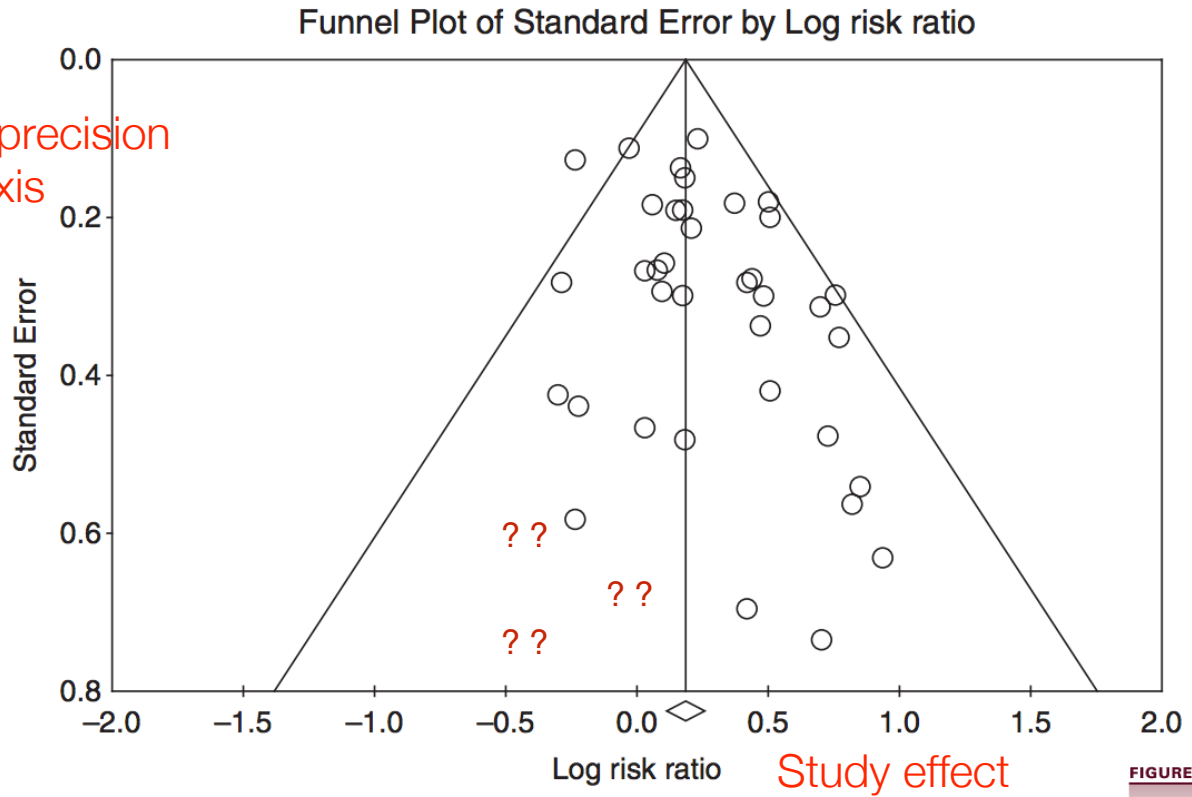
- Studies 1 & 2 I^2 values of 3% and 19% respectively consistent with Q results. These indicate little variability between studies that cannot be explained by chance.
- Study 3 I^2 shows a small effect although the **Q test** for is highly significant (**too sensitive due to many studies**)
- Studies 4 & 5 are consistent between Q and I^2
- Study 6 I^2 suggests moderate inconsistency despite negative **Q test** for heterogeneity (**too insensitive due to few studies**)

Assessing biases in meta-analysis

- Some biases are peculiar to meta-analysis.
- Positive results are more likely to be
  - Published (publication bias)
  - Published quickly (time lag bias)
  - Published in English (language bias)
  - Published more than once
  - Be cited by others (citation bias)
- Will be present to some extent in all meta analyses
- Need to assess magnitude of the problem

Detecting Publication Bias: Funnel plot

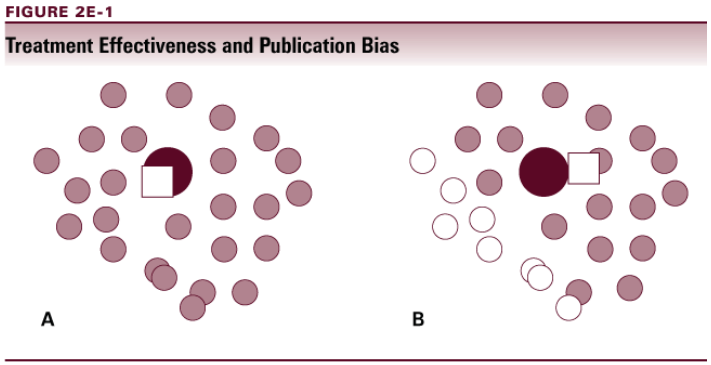
Study precision on y-axis



Study effect size on x-axis

Figure 30.2 Passive smoking and lung cancer – funnel plot.

In the absence of bias, plot of effects vs. precision should form an 'inverted' funnel, with more variation among smaller studies.



A, The black circle represents the underlying truth. The white square represents the pooled estimate from a systematic review of all the evidence. The small shaded circles represent the results of individual studies. B, The white circles represent the results of studies that the reviewers failed to identify because the studies were not published. Note the error in the pooled estimate represented by the gap between the pooled estimate (white square) and the underlying truth (black circle).

How to read a funnel plot: look at lower left corner where small negative studies should appear, if empty, think **publication bias**

What is Publication Bias?

Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias

Kerry Dwan[1*], Douglas G. Altman[2], Juan A. Arnaiz[3], Jill Bloom[4], An-Wen Chan[5], Eugenia Cronin[6], Evelyne Decullier[7], Philippa J. Easterbrook[8], Erik Von Elm[9,10], Carrol Gamble[1], Davina Ghersi[11], John P. A. Ioannidis[12,13], John Simes[14], Paula R. Williamson[1]

- "Studies that report **positive or significant results are more likely to be published** and outcomes that are statistically significant have higher odds of being fully reported." - PLoS ONE, August 2008;3:e3081
- Studies that are never published are obviously much less likely to be included in a meta-analysis. If the missing studies are a random subset of all relevant studies -> less information, wider confidence intervals, and less powerful tests, but will have no bias.
- However, if the missing studies are systematically different than the ones we were able to locate, then our sample will be biased, generating a biased picture of the cumulative evidence.

Example – Meta-analysis searching thoroughly

Rosiglitazone for type 2 diabetes mellitus (Review)

Richter B, Bandeira-Echtler E, Bergerhoff K, Clar C, Ebrahim SH

July 2007

THE COCHRANE COLLABORATION®

The Cochrane Library (issue 1, 2007);

MEDLINE - OVID interface (until April 2007);

EMBASE - OVID interface (until April 2007)

18 published RCTs 3888 patients included

- No evidence that patient-oriented outcomes (mortality, morbidity, QoL are positively or negatively influenced by rosiglitazone

NEJM - Rosiglitazone MA

The NEW ENGLAND JOURNAL of MEDICINE

N Engl J Med. 2007 Jun 14;356(24):2457-71. Epub 2007 May 21

Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes

Steven E. Nissen, M.D., and Kathy Wolski, M.P.H.

Nissen used data released under a legal settlement by Avandia's maker, GlaxoSmithKline, included results of 42 studies, 26 still unpublished

Table 4. Rates of Myocardial Infarction and Death from Cardiovascular Causes.

Study	Rosiglitazone Group	Control Group	Odds Ratio (95% CI)	P Value
	no. of events/total no. (%)			
Myocardial infarction				
Small trials combined	44/10,280 (0.43)	22/6105 (0.36)	1.45 (0.88–2.39)	0.15
DREAM	15/2,635 (0.57)	9/2634 (0.34)	1.65 (0.74–3.68)	0.22
ADOPT	27/1,456 (1.85)	41/2895 (1.44)	1.33 (0.80–2.21)	0.27
Overall			1.43 (1.03–1.98)	0.03
Death from cardiovascular causes				
Small trials combined	25/6,557 (0.38)	7/3700 (0.19)	2.40 (1.17–4.91)	0.02
DREAM	12/2,365 (0.51)	10/2634 (0.38)	1.20 (0.52–2.78)	0.67
ADOPT	2/1,456 (0.14)	5/2854 (0.18)	0.80 (0.17–3.86)	0.78
Overall			1.64 (0.98–2.74)	0.06

CONCLUSIONS

Rosiglitazone was associated with a significant increase in the risk of myocardial infarction and with an increase in the risk of death from cardiovascular causes that had borderline significance. Our study was limited by a lack of access to original

Critical reading - asking basic questions

Cardiovascular Thrombotic Events in Controlled, Clinical Trials of Rofecoxib

Marvin A. Konstam, MD; Matthew R. Weir, MD; Alise Reicin, MD; Deborah Shapiro, DrPh; Rhoda S. Sperling, MD; Eliav Barr, MD; Barry J. Gertz, MD, PhD

Background—In comparing aspirin, nonselective nonsteroidal antiinflammatory agents (NSAIDs), and cyclooxygenase (COX)-2 inhibitors, variation in platelet inhibitory effects exists that may be associated with differential risks of cardiovascular (CV) thrombotic events. Among the randomized, controlled trials with the COX-2 inhibitor rofecoxib, one study demonstrated a significant difference between rofecoxib and its NSAID comparator (naproxen) in the risk of CV thrombotic events. A combined analysis of individual patient data was undertaken to determine whether there was an excess of CV thrombotic events in patients treated with rofecoxib compared with those treated with placebo or nonselective NSAIDs.

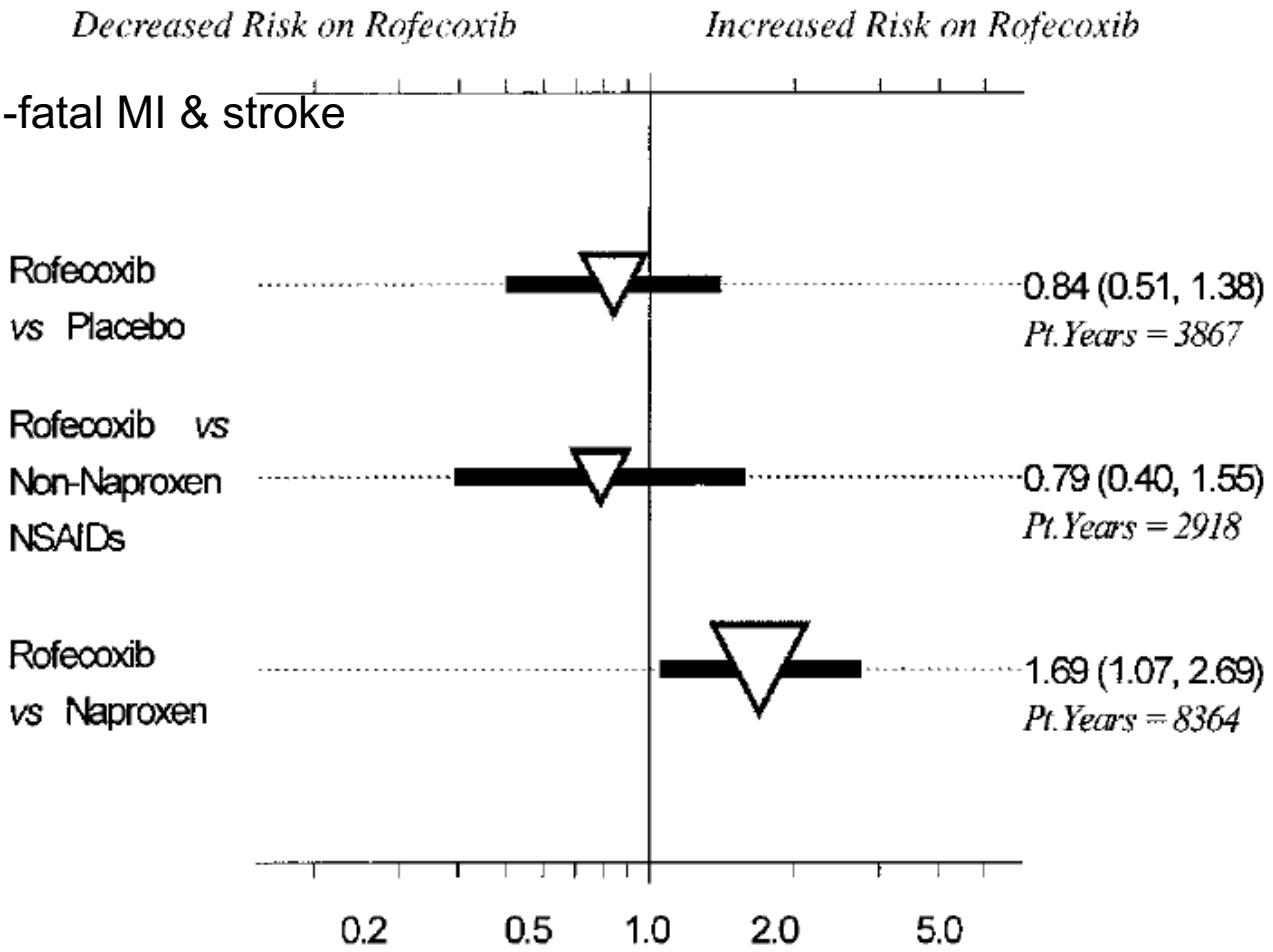
Methods and Results—CV thrombotic events were assessed across 23 phase IIb to V rofecoxib studies. Comparisons were made between patients taking rofecoxib and those taking either placebo, naproxen (an NSAID with near-complete inhibition of platelet function throughout its dosing interval), or another nonselective NSAIDs used in the development program (diclofenac, ibuprofen, and nabumetone). The major outcome measure was the combined end point used by the Antiplatelet Trialists' Collaboration, which includes CV, hemorrhagic, and unknown deaths; nonfatal myocardial infarctions; and nonfatal strokes. More than 28 000 patients, representing >14 000 patient-years at risk, were analyzed. The relative risk for an end point was 0.84 (95% CI: 0.51, 1.38) when comparing rofecoxib with placebo; 0.79 (95% CI: 0.40, 1.55) when comparing rofecoxib with non-naproxen NSAIDs; and 1.69 (95% CI: 1.07, 2.69) when comparing rofecoxib with naproxen.

Conclusions—This analysis provides no evidence for an excess of CV events for rofecoxib relative to either placebo or the non-naproxen NSAIDs that were studied. Differences observed between rofecoxib and naproxen are likely the result of the antiplatelet effects of the latter agent. (***Circulation***. **2001;104:2280-2288.**)

The data - Your interpretation?

Conclusions—This analysis provides no evidence for an excess of CV events for rofecoxib relative to either placebo or the non-naproxen NSAIDs that were studied. Differences observed between rofecoxib and naproxen are likely the result of the antiplatelet effects of the latter agent. (*Circulation*. 2001;104:2280-2288.)

Mortality, non-fatal MI & stroke



•23 RCTs of > 28,000 patients, > 14,000 pt. yrs

Circulation. 2001;104:2280-2288

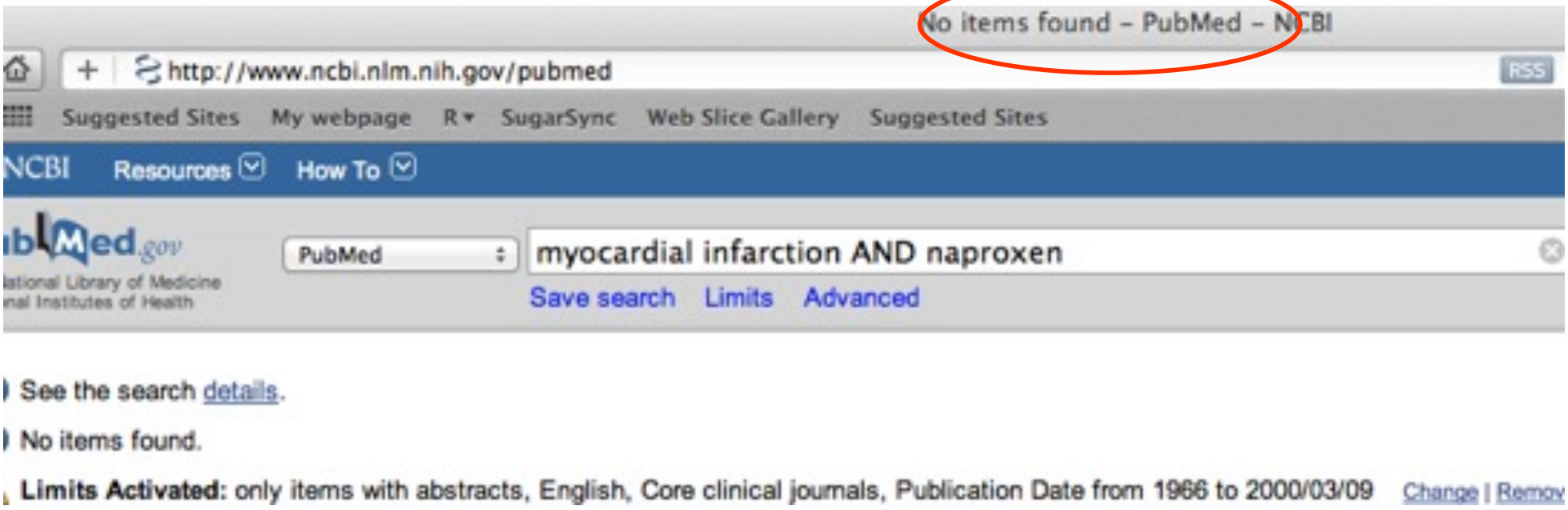
It helps to read the small print

Received October 2, 2001; accepted October 3, 2001

From the Division of Cardiology, New England Medical Center, Boston, Mass (M.A.K.); the Nephrology Division, University of Maryland Hospital, Baltimore (M.R.W.); and Merck Research Laboratory, Merck, Whitehouse Station, NJ (A.R., D.S., R.S.S., E.B., B.J.G.).

Drs Konstam and Weir have been paid consultants to Merck and Co and Pharmacia, and Dr Konstam also has been a paid consultant to Pfizer Inc. Neither has been compensated for work on this article. Drs Reicin, Shapiro, Sperling, Barr, and Gertz are employees of Merck Research Laboratories, Merck and Co, Inc. As such, they receive financial compensation that includes stock ownership and stock options.

This article originally appeared Online on October 15, 2001 (*Circulation.* 2001;104:r15–r23).



Cumulative MA - knowing sooner

ORIGINAL INVESTIGATION

HEALTH CARE REFORM

Pooled Analysis of Rofecoxib Placebo-Controlled Clinical Trial Data

Arch Intern Med. 2009;169(21):1976-1984

Lessons for Postmarket Pharmaceutical Safety Surveillance

http://archinte.jamanetwork.com/article.aspx?articleid=1108572

Joseph S. Ross, MD, MHS; David Madigan, PhD; Kevin P. Hill, MD, MHS; David S. Egilman, MD, MPH; Yongfei Wang, MS; Harlan M. Krumholz, MD, SM

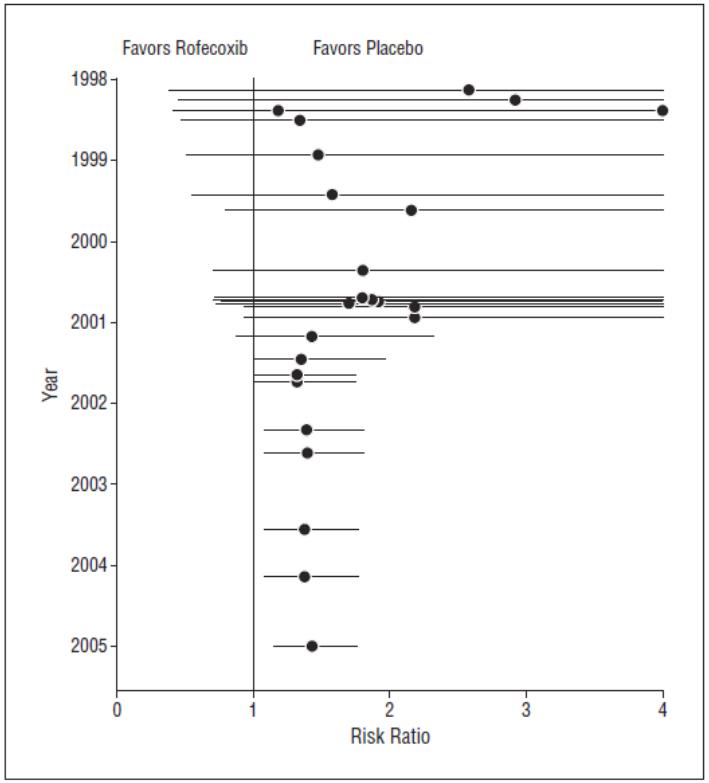
30 trials - 17,256 pts

Table 1. Randomized, Placebo-Controlled Rofecoxib Trials of 4 Weeks' Duration or Longer Conducted by Merck & Co Inc Included in Analyses

Source	Trial No.	Indication Studied	Intervention: Rofecoxib Dose, mg	Intervention: Control	Planned Duration, wk	LPO
Ehrich et al,[14] 1999	010	Osteoarthritis	25 and 125	Placebo	6	February 8, 1996
Ehrich et al,[15] 2001	029	Osteoarthritis	12.5, 25, and 50	Placebo	6	February 5, 1997
Saag et al,[16] 2000	033	Osteoarthritis	12.5 and 25	Placebo	6	November 18, 1997
Day et al,[17] 2000	040	Osteoarthritis	12.5 and 25	Placebo	6	January 1, 1998
Laine et al,[18] 1999	044	Osteoarthritis	25 and 50	Placebo	24	February 18, 1998
Hawkey et al,[19] 2000	045	Osteoarthritis	25 and 50	Placebo	24	February 18, 1998
Truitt et al,[20] 2001	058	Osteoarthritis	12.5 and 25	Placebo	6	April 1, 1998
Unpublished	083	Osteoarthritis	25	Placebo	64	February 9, 2000
Kivitz et al,[21] 2004	085	Osteoarthritis	12.5	Placebo	6	March 3, 1999
Weaver et al,[22] 2006	090	Osteoarthritis	12.5	Placebo	6	May 17, 1999
Smugar et al,[23] 2006	112	Osteoarthritis	12.5 and 25	Placebo	6	September 8, 2000
Smugar et al,[23] 2006	116	Osteoarthritis	25	Placebo	6	June 22, 2000
Laine et al,[24] 2004	136	Osteoarthritis	25	Placebo	12	February 5, 2002
Birbara et al,[25] 2006	219	Osteoarthritis	12.5	Placebo	6	November 28, 2003
Birbara et al,[25] 2006	220	Osteoarthritis	12.5	Placebo	6	November 24, 2003
Unpublished	017	Rheumatoid arthritis	125 and 175	Placebo	6	May 21, 1997
Schnitzer et al,[26] 1999	068	Rheumatoid arthritis	25 and 50	Placebo	8	September 10, 1998
Truitt et al,[27] 2001 (abstract only)	096	Rheumatoid arthritis	12.5 and 25	Placebo	12	July 21, 2000
Geusens et al,[28] 2002	097	Rheumatoid arthritis	25 and 50	Placebo	12	June 6, 2000
Hawkey et al,[29] 2003	098 and 103	Rheumatoid arthritis	50	Placebo	12	July 6, 2000
Thal et al,[30] 2005	078	Alzheimer disease	25	Placebo	208	April 23, 2003
Reines et al,[31] 2004	091	Alzheimer disease	25	Placebo	52	November 30, 2000
Unpublished	126	Alzheimer disease	25	Placebo	52	May 30, 2001
Nickel et al,[32] 2003	118	Chronic nonbacterial prostatitis	25 and 50	Placebo	6	July 26, 2000
Katz et al,[33] 2003	120 and 121	Low back pain	25 and 50	Placebo	4	June 27, 2000
Bresalier et al[3] 2005, and Baron et al,[34] 2008	122	Colorectal adenomas	25	Placebo	156	September 30, 2004
Unpublished	125	Migraine prophylaxis	25	Placebo	12	June 29, 2001
Unpublished	129	Familial adenomatous polyposis	25	Placebo	24	May 14, 2002

Abbreviation: LPO, last patient out (clinical trial completion date).

Cumulative MA



DRUG WITHDRAWN FROM THE MARKET SEPT 2004

Financial ties and spin

Yank et al. (BMJ, 2007) studied 124 meta-analyses of anti-hypertensives on clinical outcomes and compared study results and conclusions among industry-sponsored vs. non-sponsored studies:

Table 5 | Proportion of meta-analyses with favourable results or conclusions, and proportion with poor concordance between results and conclusions, by financial ties*

Financial ties	No (%) with favourable results	No (%) with favourable conclusions	No (%) with poor concordance between results and conclusions*
One drug company (n=49)	27 (55)	45 (92)	18 (37)
All other (n=75):	49 (65)	55 (73)	6 (8)
Multiple drug companies (n=14)	8 (57)	11 (79)	3 (21)
No statement (n=25)	14 (56)	17 (68)	3 (12)
Both drug and non-profit (n=9)	6 (67)	6 (67)	0 (0)
Non-profit (n=27)	21 (78)	21 (78)	0 (0)

*Poor concordance for each row was determined by the calculation: [number of meta-analyses with favourable conclusions]-[number of meta-analyses with favourable results].

» Industry sponsored studies more discordance btw results & conclusions i.e. **more spin**

Summary

- Need full data disclosure for meta-analysis
- Need care for both the statistical and substantive components
- There are limitations to aggregate meta-analysis
- Enlarging the scope by individual patient data meta-analysis or network meta-analysis (includes direct & indirect comparisons) may also be useful but resource intense
- All seem to agree SR/MA important **before** undertaking a new study
- What about **after** the study - should the new results not be interpreted in the context of what is already known?

Thank you!

"Everybody gets so much information all day long that they lose their common sense" – Gertrude Stein