Can I have FAITH in this Review?

Find

Appraise

Include

Total

Heterogeneity

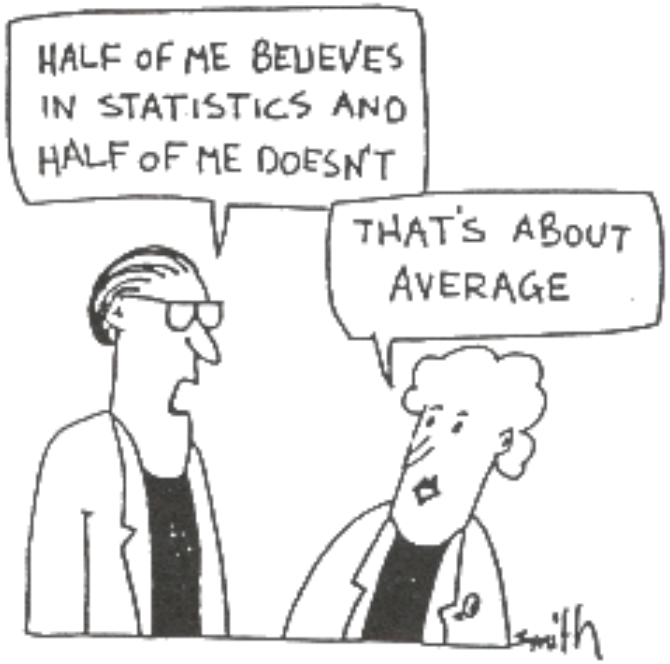
Paul Glasziou

Centre for Research in Evidence Based Practice

Bond University

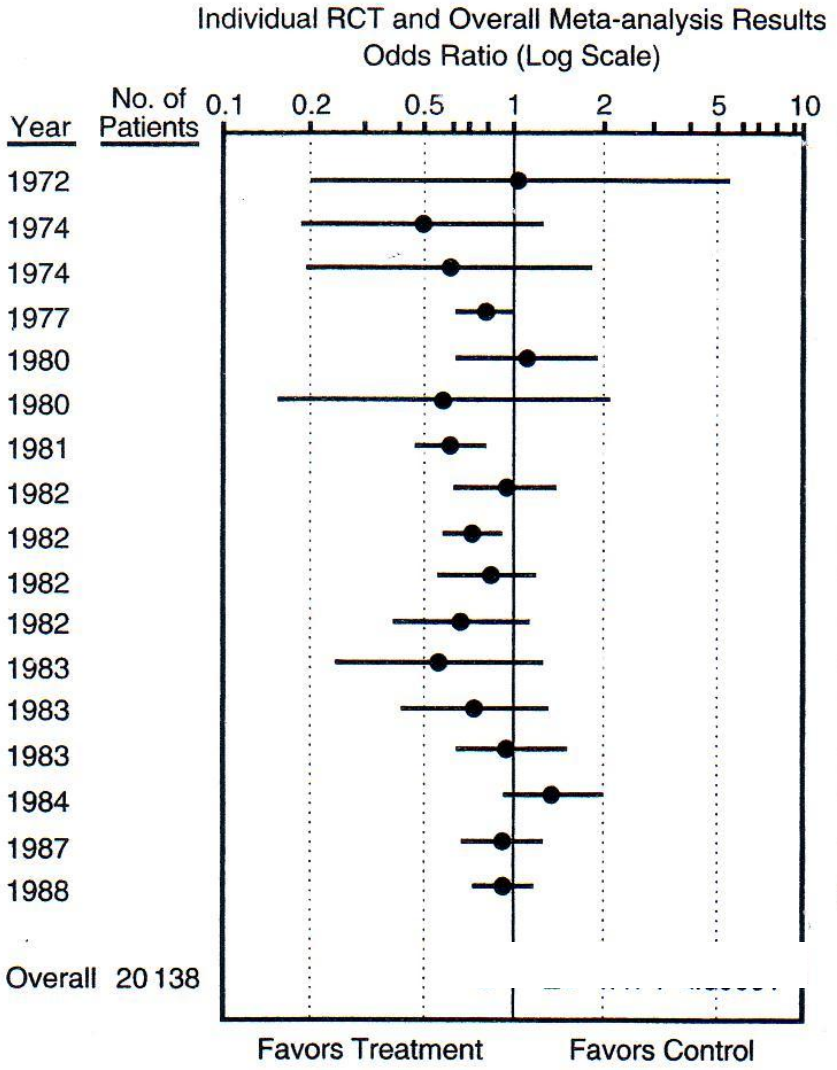
What do you do?

- For an acutely ill patient, you do a search
- You find several studies: some find that it works; some do not
- What do you do?

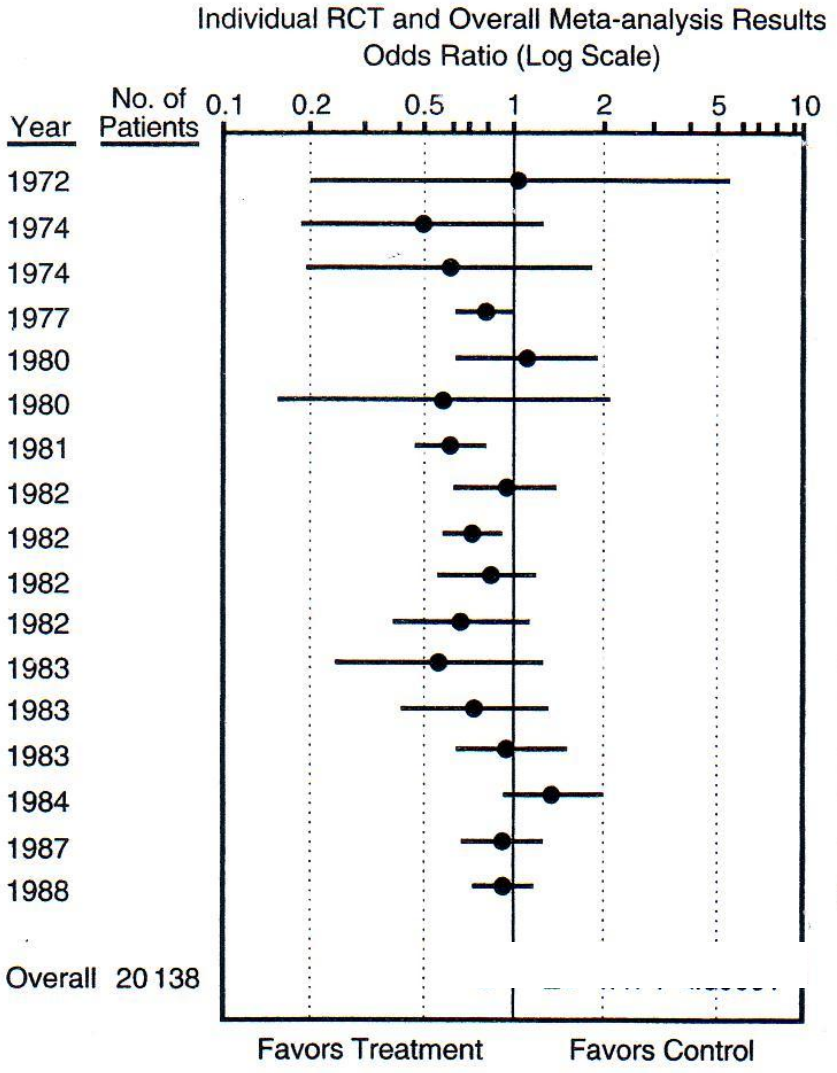


Ask somebody to find all studies, select the best, ...

Summary of the 17 studies

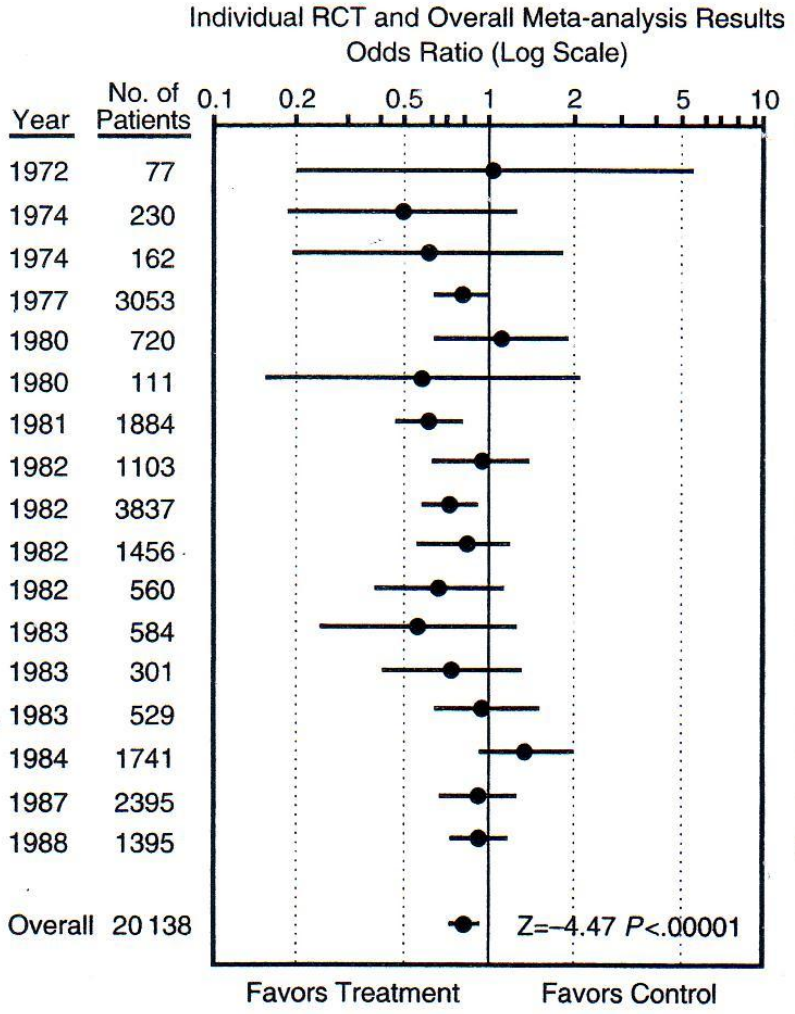


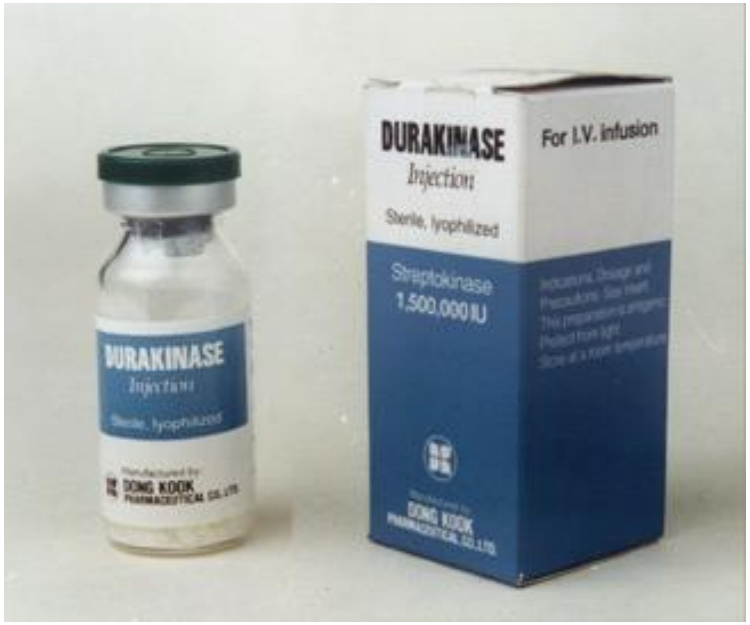
Summary of the 17 studies



A. Which is the smallest study?
B. Which is the largest study?
C. How many are statistically significant?
D. Which studies are "large enough"?

Summary of the 17 studies: streptokinase





How large should the study be?

EBM notebook

Was the study big enough? Two café rules

Why is a small study a problem?

When reading an article, we often wonder whether the study was large enough. If a study does not find a statistically significant effect (eg, at $p<0.05$), it may be because the study was too small or because there actually is no true effect. You should check whether the confidence intervals (CIs) show that the data are consistent with the effect being clinically important, even though the effect was not "statistically significant."

In this note we will provide you with 2 "café rules" (for when you are discussing studies over an espresso), and then point to the ideas behind them and some resources for more exact calculations.

HOW DO WE KNOW THE REQUIRED SAMPLE SIZE?

It is helpful to have an approximate idea of the sample size requirements for different types of studies. The first approximate rule is the 50:50 rule for studies looking at dichotomous ("present or absent") outcomes such as mortality, hospital admissions, or remissions.

What sample size is needed?

For disease X the usual mortality rate is 0%

What sample size is needed to detect a reduction in mortality?

- 100
- 1,000
- 100,000
- 1,000,000

Sample Size: Café Rule 1
The 50:50 Rule (proportions)

50 events are needed in the control group:

(For an 80% chance of finding a 50% reduction)

Control Rate	*Number Events*	*Control# (Rule 1)*	*Control# (Fisher exact)*
20%	50	250	215
10%	50	500	463
5%	50	1000	962

Glasziou P, Doll H. Was the study big enough? Two cafe rules. Evid Based Med. 2006;11(3):69-70.

What sample size is needed?

- There is usually a 12% mortality rate
  - You think your treatment will lower mortality by 50%
- What sample size is needed?

What sample size is needed?

- There is usually a 12% mortality rate
  - You think your treatment will lower mortality by 50%
- What sample size is needed?

- 12% means
  - 12/100 or 24/200 or 48/400
  - and 50 per 417
- Control + Treatment Groups = 834 in total

Systematic Review or meta-analysis?

- A **Systematic Review** is a review of a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review.

- Statistical methods (*meta-analysis*) may or may not be used to analyze and summarize the results of the included studies.

Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or *n*-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, *n*-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or *n*-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non -randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

*** Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.**

**** As always, a systematic review is generally better than an individual study.**

How to cite the Levels of Evidence Table

OCEBM Levels of Evidence Working Group*. "The Oxford 2011 Levels of Evidence".

Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653

** OCEBM Table of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson*

Is the review any good? FAITH check

- **Q**uestion – What is the PICO?

- **F**inding
  - Did they find most studies?
- **A**ppraisal
  - Did they
- **I**nclude
  - Did they include only good ones?
- **T**otal up
  - What to they all mean?
- **H**eterogeneity of PICOs, results

Why do I need to check the review?

Most reviews do not pass minimum criteria

A study of 158 reviews*

- Only 2 met all 10 criteria
- Median was only 1 of 10 criteria met

FAITH tool = 5 criteria

* McAlister Annals of Intern Med 1999

What it the review question (PICO)?

- Population
- Intervention
- Comparison
- Outcome(s)

Using Pedometers to Increase Physical Activity and Improve Health

A Systematic Review

Dena M. Bravata, MD, MS
Crystal Smith-Spangler, MD
Vandana Sundaram, MPH
Allison L. Gienger, BA

Context Without detailed evidence of their effectiveness, pedometers have recently become popular as a tool for motivating physical activity.

Objective To evaluate the association of pedometer use with physical activity and health outcomes among outpatient adults.

Do pedometers increase activity and improve health?

- **Find: what is your search strategy?**
  - **Databases?**
  - **Terms?**
  - **Other methods?**

Do yourself then
Get neighbour's help

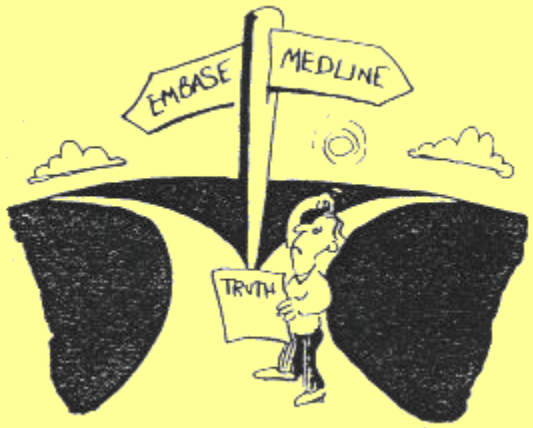
METHODS

Data Sources and Search Strategies

In collaboration with a professional librarian, we developed individualized search strategies for 7 databases: MEDLINE (January 1966 to February 2007); and EMBASE, Sport Discus, PsychINFO, Cochrane Library, Thompson Scientific (formerly known as Thompson ISI), and ERIC (January 1966 to May 2006). We used search terms such as *pedometer*, *activity monitor*, and *step counter*. We also reviewed the bibliographies of retrieved articles and relevant conference proceedings and contacted experts in exercise physiology for additional studies.

FIND: Did they find all Studies?

- Check for existing systematic review?
- Good initial search
  - Terms (text and MeSH)
  - At least 2 Databases: MEDLINE, EMBASE, CINAHL, CCTR, ...
- Plus a Secondary search
  - Check references of relevant papers & reviews and
  - Find terms (words or MeSH terms) you <u>didn't</u> use
  - Search again! (*snowballing*)



Is finding all published studies enough?

- Negative studies less likely to be published than 'Positive'
- How does this happen?
- Follow-up of 737 studies at Johns Hopkins

  (Dickersin, JAMA, 1992)
  - Positive SUBMITTED more than negative (2.5 times)

Registered vs Published Studies

Ovarian Cancer chemotherapy: single v combined

	Published
No. studies	16
Survival ratio	1.16
95% CI	1.06-1.27
P-Value	0.02

Simes, J. Clin Oncol, 86, p1529

Registered vs Published Studies

Ovarian Cancer chemotherapy: single v combined

	Published	Registered
No. studies	16	13
Survival ratio	1.16	1.05
95% CI	1.06-1.27	0.98-1.12
P-Value	0.02	0.25

Simes, J. Clin Oncol, 86, p1529

Which are biased? Which OK?

1. All positive studies
2. All studies conducted in the Northern Hemisphere
3. All studies published in BMJ, Lancet, JAMA or NEJM
4. All studies with more than 100 patients
5. All studies registered studies

Publication Bias: Solution

- All trials registered at inception,
  - The National Clinical Trials Registry: Cancer Trials
  - National Institutes of Health Inventory of Clinical Trials and Studies
  - International Registry of Perinatal Trials
- Meta-Registry of trial Registries
  - www.controlled-trials.com

Home
About WHO
Countries
Health topics
Publications
Data and statistics
Programmes and projects
International Clinical Trials Registry Platform
About us
Why register trials?
International Search Portal
Register network
Universal Trial Reference Number
Results reporting
News and events
Resources

International Clinical Trials Registry Platform (ICTRP)

WHO > Programmes and projects

Welcome to the WHO International Clinical Trials Registry Platform

The mission of the WHO Registry Platform is to ensure that a complete view of research is accessible to all those involved in health care decision making. This will improve research transparency and will ultimately strengthen the validity and value of the scientific evidence base.

The registration of all interventional trials is a scientific, ethical and moral responsibility.

What is a clinical trial?

A clinical trial is any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes. Interventions include but are not restricted to drugs, cells and other biological products, surgical procedures, radiologic procedures, devices, behavioural treatments, process-of-care changes, preventive care, etc

Functions

The Register Network | The International Search Portal

World Health Organization
INTERNATIONAL CLINICAL TRIALS REGISTRY PLATFORM SEARCH PORTAL

Search for trials

The Register Network

List of Registers

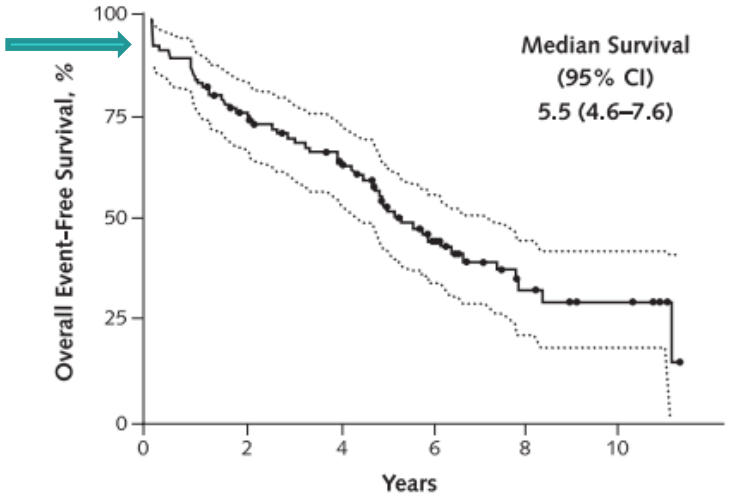
Frequently Asked Questions

What was the "Find" date?

Of 100 systematic reviews:

Median time to a change that would effect clinical decisions was 5.5 years.

Figure 2. **Overall survival time (95% CI) free of signals for updating.**



Systematic reviews at risk, *n*	100	73	59	34	14	6

The immediate decrease in survival at time zero reflects the 7 systematic reviews for which signals for updating had already occurred at the time of publication. The low number of reviews at risk after 10 years reflects the fact that the sample spanned 1995 to 2005 and censoring occurred on 1 September 2006. Thus, only reviews published before September 1996 and having no signals for updating could have more than 10 years of observation.

Shojania Ann Intern Med, 2007

Appraise & Include studies

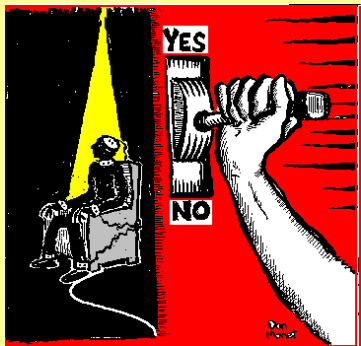
Did they check & select only the good quality studies?

Miscalculating NNT

Authors often miscalculate the NNT. What should we do about this?

Q1. Would the death penalty for miscalculating an NNT prevent future miscalculation? (FACT)

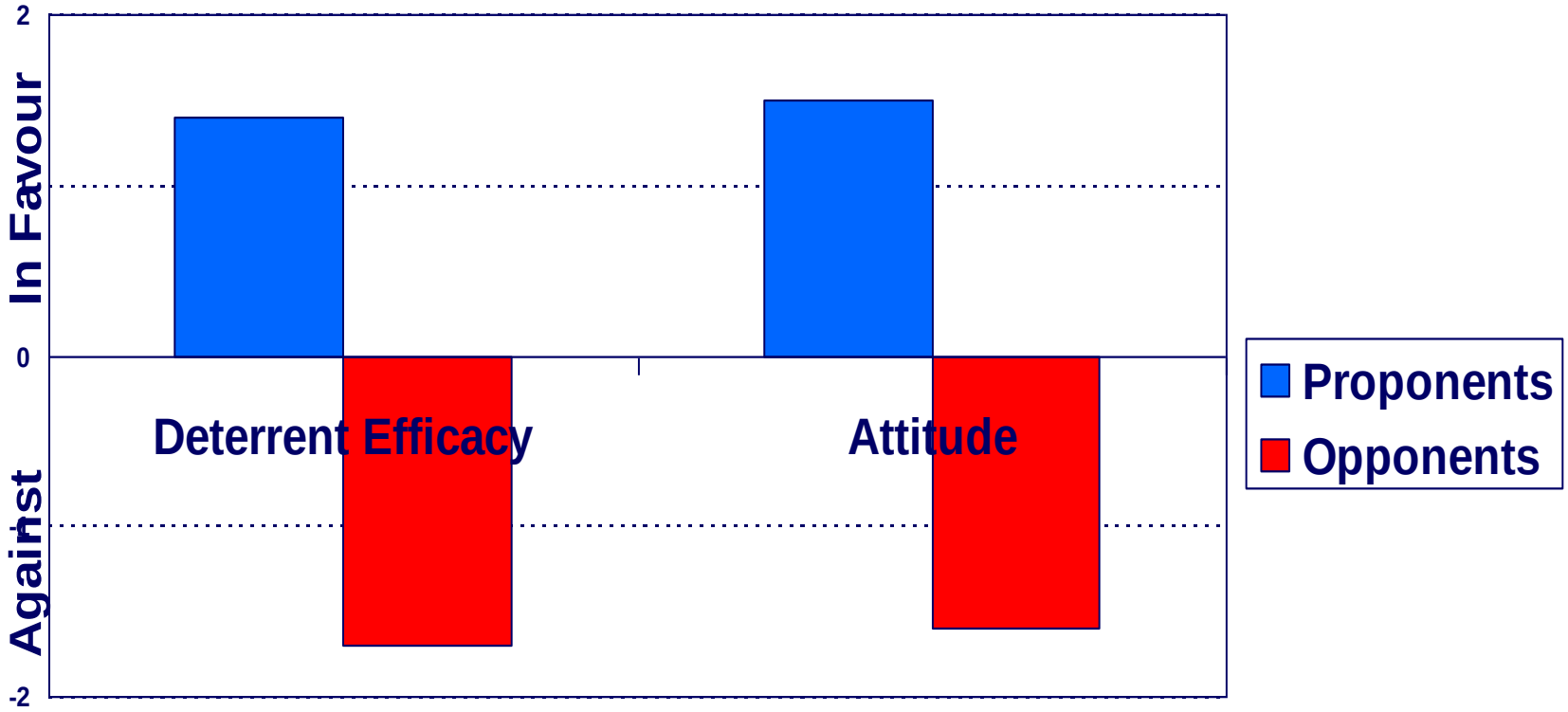
Q2. Should we have the death penalty for miscalculation of NNT? (VALUE)



Selective Criticism of Evidence

Biased appraisal increases polarization

Capital punishment: beliefs and contradictory studies



Lord et al, J Pers Soc Psy, 1979, p2098

Selective Criticism of Evidence

28 reviewers assessed one "study"
results randomly positive or negative

	"Positive"	*"Negative"*
Relevance	5.2	4.9
Methods	4.2	2.4
Presentation	4.3	2.6
Summary	3.2	1.8

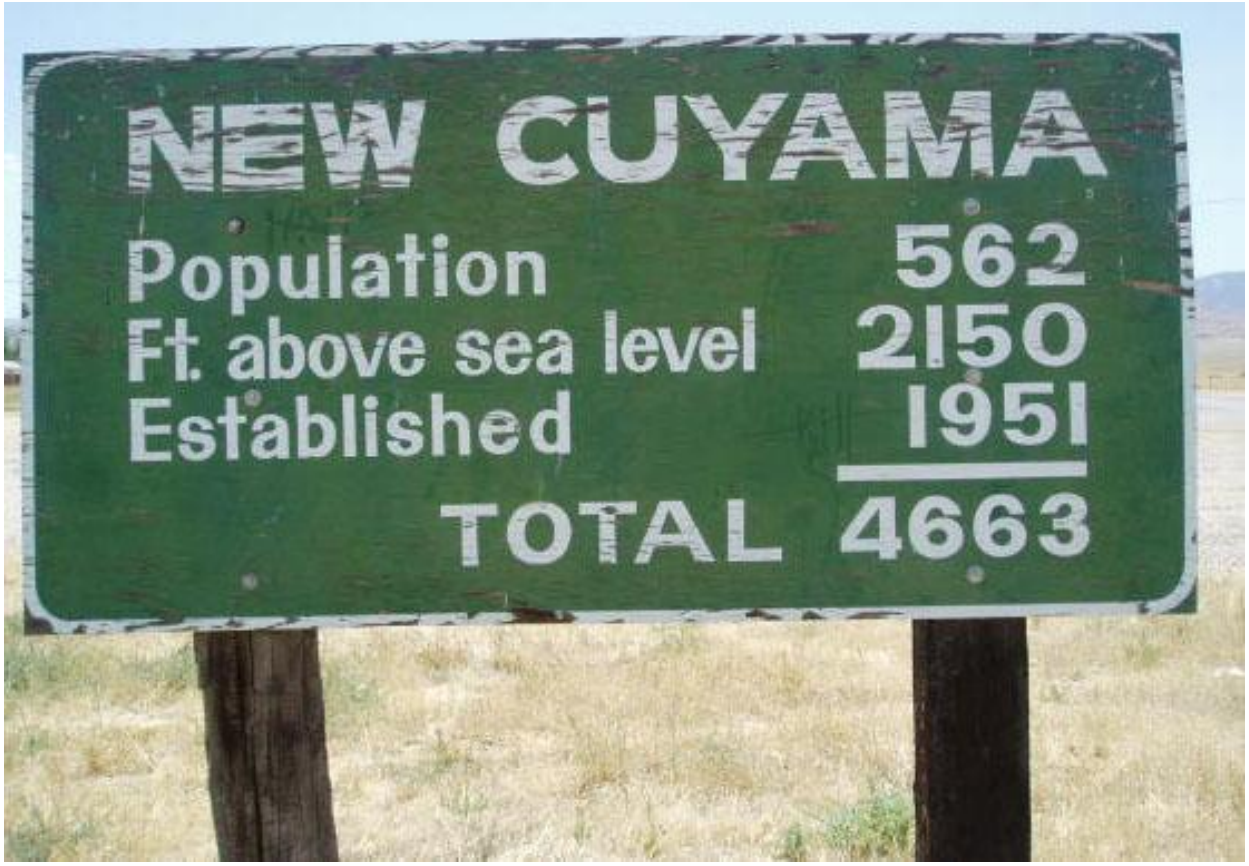
(Cog Ther Res, 1977, p161-75)

Assessment: How can you avoid biased selection of studies?

- Assessment and selection should be:
  - Standardized "Objective" OR
  - Blinded to Results

* assessment of quality blind to study outcome

Total up: pooling the results



Meta-analysis (Forest) plot

Study	No of patients	Treatment Mean (SD)	No of patients	Control Mean (SD)	Weighted mean difference (fixed effect) (95% CI)	Weight (%)
Dieppe 1980[8]	12	38.0 (29.0)	12	70.0 (30.0)		7.45
Gaffney 1995[9]	42	21.7 (20.7)	42	43.1 (28.7)		36.26
Jones 1998[17]	29	48.0 (30.0)	30	57.5 (30.0)		17.71
Ravaud 1999[11]	24	23.7 (26.2)	21	45.7 (26.6)		17.36
Smith 2003[12]	38	20.8 (30.0)	33	24.7 (30.0)		21.22
Total (95% CI)	145		138			100.00

Test for heterogeneity: χ^2=6.87, df=4, P=0.14, I^2=41.7%

Test for overall effect: z=5.01, P=0.00001

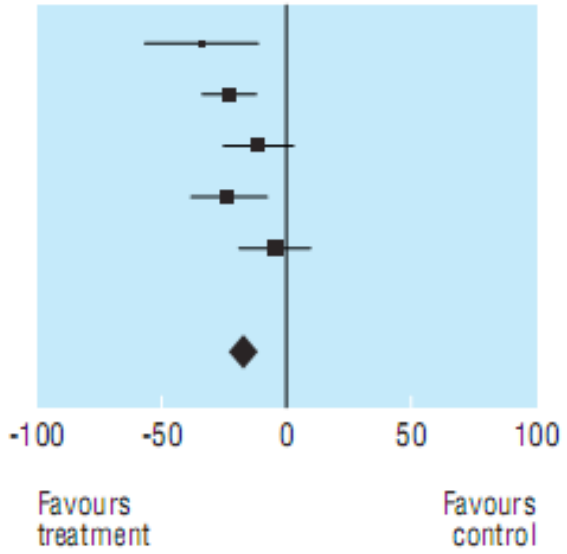
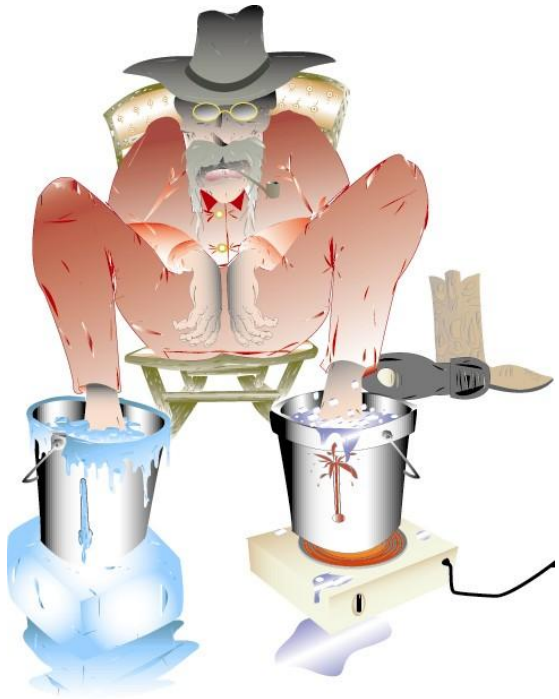


Fig 4 Visual analogue scale for pain up to two weeks after steroid injection in knee

Heterogeneity? Use in my patients

Is the AVERAGE effect similar across studies?

- If NO, then WHY?
  - Study methods (RAMbo - biases)
  - PICO (Patients, Intervention, …)
- If YES, then 2 questions
  - Effect in different individuals?
  - Which version of treatment?



Meta-analysis (Forest) plot

Study	No of patients	Treatment Mean (SD)	No of patients	Control Mean (SD)	Weighted mean difference (fixed effect) (95% CI)	Weight (%)
Dieppe 1980[8]	12	38.0 (29.0)	12	70.0 (30.0)		7.45
Gaffney 1995[9]	42	21.7 (20.7)	42	43.1 (28.7)		36.26
Jones 1998[17]	29	48.0 (30.0)	30	57.5 (30.0)		17.71
Ravaud 1999[11]	24	23.7 (26.2)	21	45.7 (26.6)		17.36
Smith 2003[12]	38	20.8 (30.0)	33	24.7 (30.0)		21.22
Total (95% CI)	145		138			100.00

Test for heterogeneity: χ^2=6.87, df=4, P=0.14, I^2=41.7%

Test for overall effect: z=5.01, P=0.00001

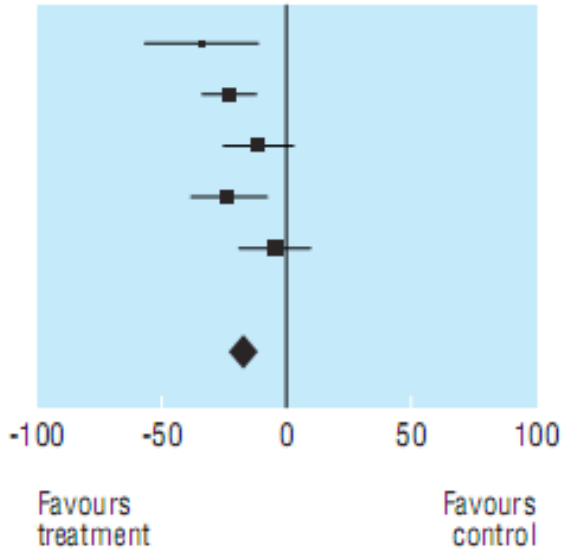
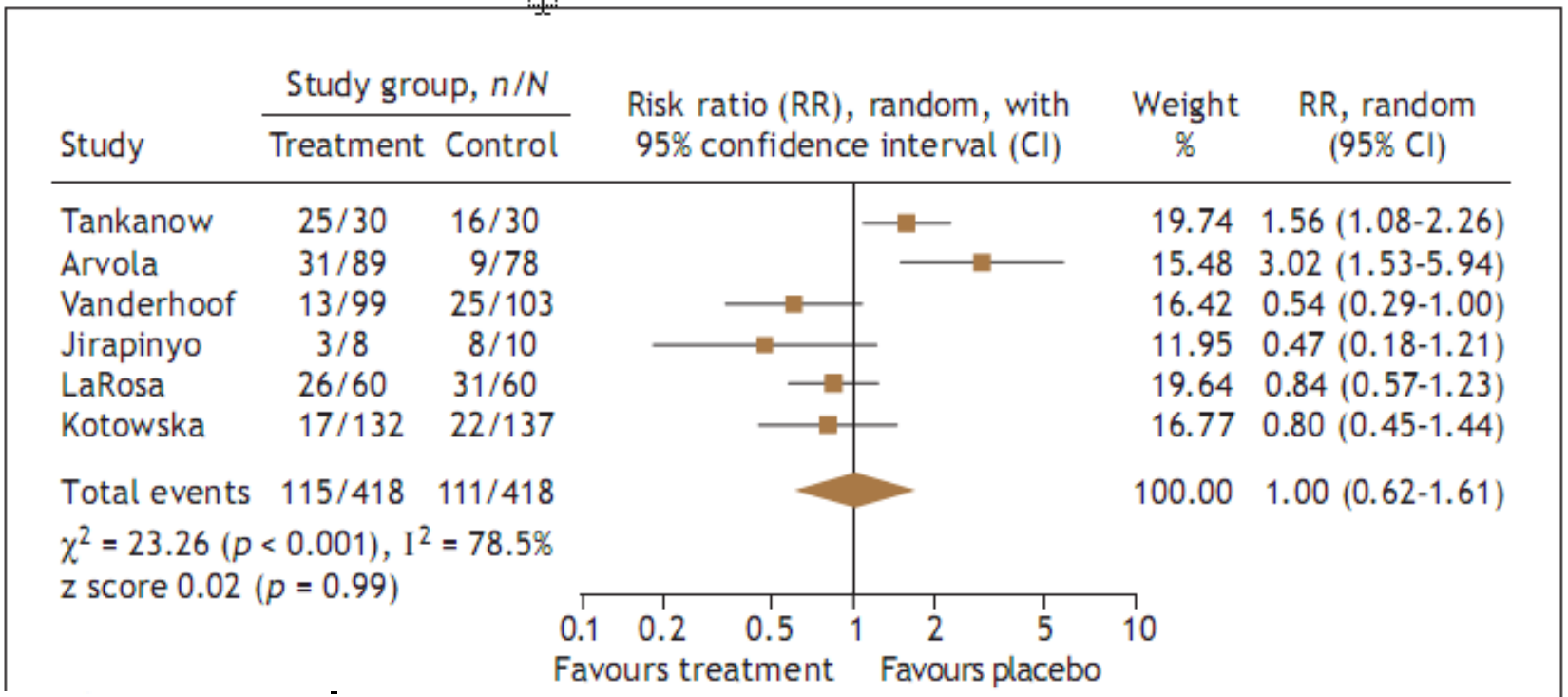


Fig 4 Visual analogue scale for pain up to two weeks after steroid injection in knee

Are the results similar across studies? 3 tests

1. Eyeball" test – do they look they same?
2. Test of "Null hypothesis" of no variation (p-value)
3. Proportion of variation not due to chance (I^2)

Are these trials different?



$\chi^2 = 23.26$ ($p < 0.001$), $I^2 = 78.5\%$

z score 0.02 ($p = 0.99$)

tention-to-treat analysis. The analysis
cs and placebo (z score) and statisti-

Conclusion
EBM and Systematic Review

- EBM (quick & dirty)
- Steps
  1. Ask Question
  2. Search
  3. Appraise
  4. Apply
- Time: 90 seconds
- < 20 articles
- This patient survives!

- Systematic Review
- Steps
  1. Ask Question
  2. Search ++++ x 2
  3. Appraise x 2
  4. Synthesize
  5. Apply
- Time: 6 months, team
- < 2,000 articles
- This patient is dead

Find a systematic review!! (and appraise it FAST)

Pros and cons of systematic reviews

- Advantages
  - Larger numbers & power
  - Robustness across PICOs
- Disadvantages
  - May conclude small biases are real effects

Is the review any good? FAITH check

- **Q**uestion – What is the PICO?

- **F**inding
  - Did they find most studies?
- **A**ppraisal
  - Did they
- **I**nclude
  - Did they include only good ones?
- **T**otal up
  - What to they all mean?
- **H**eterogeneity of PICOs, results

Using review results: what do I do with my patient?

- STUDY: meta-analysis of behavioural interventions for insomnia adults
  - ".. confirms the efficacy of behavioral interventions for person with chronic insomnia."

- PROBLEM: No regimens for 'behavioural intervention' described
  - Author asked: "what specific treatment regime (or regimes) would you recommend based on your review?"
  - Author response: "It was found that cognitive, behavioral and relaxation therapies all in general lead to similar improvements in sleep outcomes---although cognitive approaches might have been a bit better. The references for these studies are found in the article. "

Rx

"Behavioural Intervention"