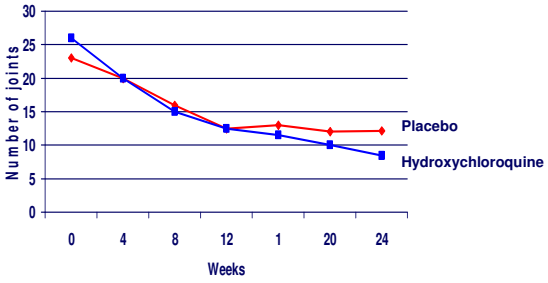
Rapid Critical Appraisal
of RCTs and Systematic Reviews

Dr Su May Liew
Centre for Evidence Based Medicine
University of Oxford
www.cebm.net



Hydroxychloroquine in Rheumatoid Arthritis



Step 3 in EBM: appraisal

1. Formulate an answerable question
2. Track down the best evidence
3. Critically appraise the evidence for:
   - Validity
   - Impact (size of the benefit)
   - Applicability
4. Integrate with clinical expertise and patient values
5. Evaluate our effectiveness and efficiency
   - keep a record; improve the process

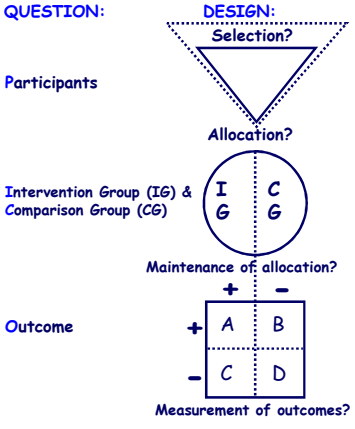


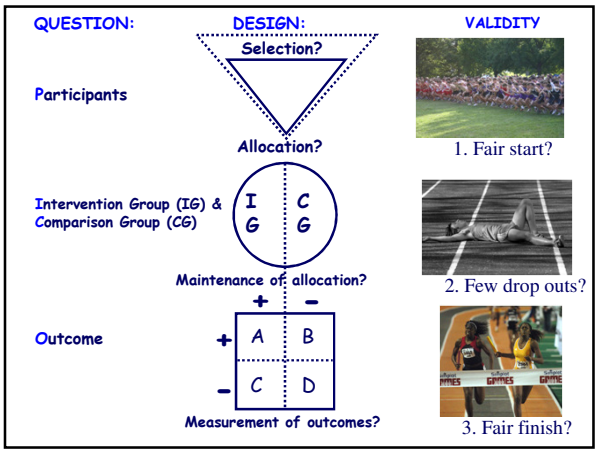
Searching for critical appraisal checklists randomized controlled trials .

11,100 articles (0.40 seconds) Google

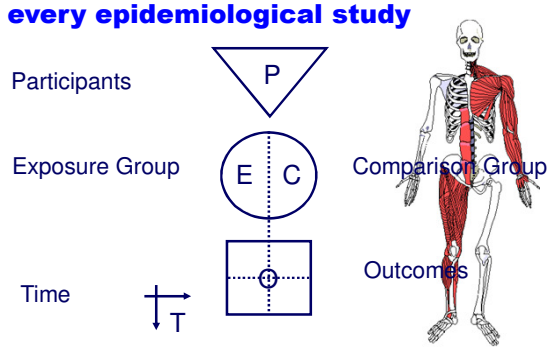
A CHECKLIST FOR APPRAISING RANDOMIZED CONTROLLED TRIALS

1. Was the objective of the trial sufficiently described?
2. Was a satisfactory statement given of the diagnostic criteria for entry to the trial?
3. Were concurrent controls used (as opposed to historical controls)?
4. Were the treatments well defined?
5. Was random allocation to treatments used?
6. Was the potential degree of blindness used?
7. Was there a satisfactory statement of criteria for outcome measures? Was a primary outcome measure identified?
8. Were the outcome measures appropriate?
9. Was a pre-study calculation of required sample size reported?
10. Was the duration of post-treatment follow-up stated?
11. Were the treatment and control groups comparable in relevant measures?
12. Were a high proportion of the subjects followed up?
13. Were the drop-outs described by treatment and control groups?
14. Were the side-effects of treatment reported?
15. How were the ethical issues dealt with?
16. Was there a statement adequately describing or referencing all statistical procedures used?
17. What tests were used to compare the outcome in test and control patients?
18. Were 95% confidence intervals given for the main results?
19. Were any additional analyses done to see whether baseline characteristics (prognostic factors) influenced the outcomes observed?
20. Were the conclusions drawn from the statistical analyses justified?





The PECOT acronym: the 5 parts of every epidemiological study



All epidemiological studies can be hung on the GATE frame

Was it a fair race?

1. Fair start?



2. Few drop outs



3. Fair finish?



Using the PICO to orient us

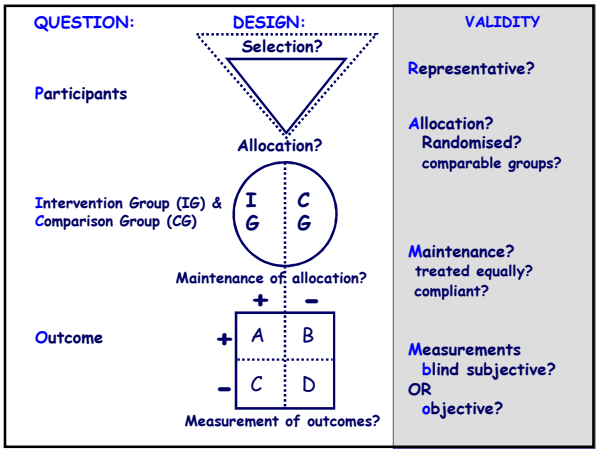
Clinical Question
In people who take long-haul flights does wearing graduated compression stockings prevent DVT?

ARTICLES

Frequency and prevention of symptomless deep-vein thrombosis in long-haul flights: a randomised trial

John H Scurr, Samuel J Machin, Sarah Bailey-King, Ian J Mackie, Sally McDonald, Philip D Coleridge Smith

Scurr et al, Lancet 2001; 357:1485-89



Use the **RAMMbo** to check validity



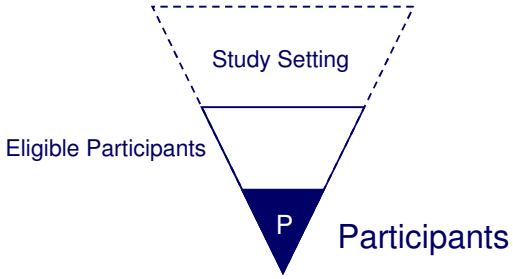
Was the Study valid?

1. **R**epresentativeness
   - Who did the subjects represent?
2. **A**llocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial's start?
3. **M**aintenance
   - Were the groups treated equally?
   - Were outcomes ascertained & analysed for most patients?
4. **M**easurements blinded OR objective
   - Were patients and clinicians "**b**linded" to treatment? OR
   - Were measurements **o**bjective & standardised?

Study statistics (p-values & confidence intervals)

User Guide. JAMA, 1993

Participants



Appraisal checklist - RAMMbo

Study biases

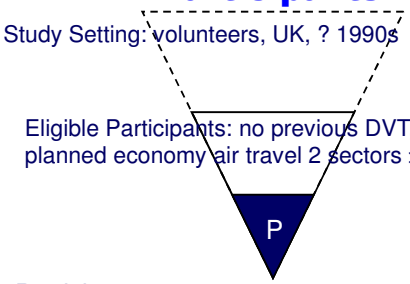
1. Recruitment
   - Who did the subjects represent?
2. **A**llocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial's start?
3. Maintainence
   - Were the groups treated equally?
   - Were outcomes ascertained & analysed for most patients?
4. Measurements
   - Were patients and clinicians "blinded" to treatment? OR
   - Were measurements objective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993

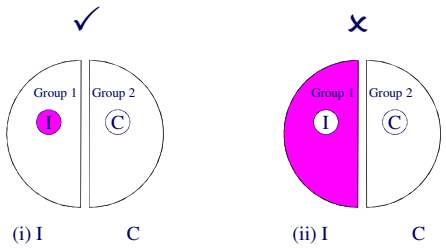
DVT in long-haul flights: Lancet 2001;357:1485-9

Participants



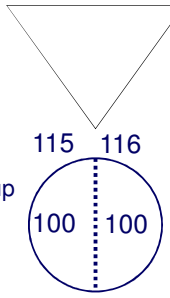
Participants: 200, mean age 61-62 years

Comparable Groups: the only difference should be the treatments



Is the difference between I and C because of (i) the intervention or (ii) because the groups were not comparable in the first place?

Exposure & Comparison Groups



Exposure or Intervention Group (EG): Below knee compression stockings

Comparison or Control Group (CG): no stockings

Fair Allocation to treatments

How do we get comparable groups?

- Was assignment to treatments randomised?
  - Was the allocation process tamper proof? OR
  - Were the groups similar at start of trial?

Benefits of Randomisation (and Allocation Concealment)

- Minimises confounding - known and unknown potential confounders are evenly distributed between study groups
  - reduces bias in those selected for treatment
  - guarantees treatment assignment will not be based on patients' prognosis

Fair Allocation – balance achieved?

Were the groups similar at the start?

- ✓ Usually Table 1 in Results section
- ✓ Do imbalances favour one treatment?

Results

Volunteers were excluded before randomisation if they did not fulfil the entry requirements or could not attend hospital for investigation both before and after travel (figure). Thus, 231 of 479 volunteers were randomised. 27 passengers were unable to attend for subsequent ultrasound investigation because of ill-health (three), change of travel plans, or inability to keep appointments (24). Two who

	No stockings	Stockings
Number	116	115
Age (years)	62 (56-68)	61 (56-66)
Number of women (%)	61 (53%)	81 (70%)
Number with varicose veins	41	45
Days of stay	17 (13-32)	16 (13-27)
Hours flying time	22 (18-36)	24 (19-35)
Haemoglobin (g/L)	142 (133-149)	140 (133-147)
WBC (×10⁹/L)	5·9 (5·0-7·3)	6·0 (5·0-6·9)
Packed cell volume	0·44 (0·42-0·47)	0·44 (0·41-0·46)
Platelets (× 10⁹/L)	240 (206-272)	242 (219-290)
Number FVL positive	7	4
Number PGM positive	1	3

Median (interquartile range) shown, unless otherwise indicated. WBC=white blood cells. FVL=factor V Leiden. PGM=prothrombin gene mutation.

Table 1: **Characteristics of study groups**

Allocation Concealment OR

Demonstrated baseline balance

BEST – most valid technique

- Central computer randomization



DOUBTFUL

- Envelopes, etc



NOT RANDOMISED

- Date of birth, alternate days, etc

RAMMbo

Appraisal checklist - RAMMbo

Study biases

1. Recruitment
   - Who did the subjects represent?
2. Allocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial's start?

3. Maintenance

- Were outcomes ascertained & analysed for most patients?
- Were the groups treated equally?

4. Measurements
   - Were patients and clinicians "blinded" to treatment? OR
   - Were measurements objective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993

ARTICLES

Frequency and prevention of symptomless deep-vein thrombosis in long-haul flights: a randomised trial

ally McDonald, Philip D Coleridge Smith

Randomisation
Volunteers were randomised by sealed envelope to one of two groups.

Passengers were randomly allocated to one of two groups: one group wore class-I below-knee graduated elastic compression stockings, the other group did not.

Scurr et al, Lancet 2001; 357:1485-89

Effects of non-equal treatment

Apart from actual intervention - groups should receive identical care!

- Trial of Vitamin E in pre-term infants (1948)
- Vit E "prevented" retrolental fibroplasia

  (By removal from 100% Oxygen to give the frequent doses of Vit E!)
- Rx: Give placebo in an identical regime, and a standard protocol



Equal treatment in DVT study?

	Number of Participants	
	No Stockings	Stockings
Aspirin	9	11
Hormone replacement therapy	8	16
Thyroxine	6	6
Antihypertensives, including diuretics	10	12
Antipeptic ulcer drugs	8	3

*Includes additions to usual drugs

Table 3: **All drugs taken by volunteers who attended for examination before and after air travel***

Scurr et al, Lancet 2001; 357:1485-89

How important are the losses?

- Equally distributed?
  - Stocking group: 6 men, 9 women - 15
  - No stocking group: 7 men, 9 women - 16
- Similar characteristics?
  - No information provided

Maintaining the Randomisation

- Principle 1 (Intention to treat)
  - Once a patient is randomised, s/he should be analysed in the group randomised to - even if they discontinue, never receive treatment, or crossover.
- Principle 2 (adequate followup)
  - "5-and-20 rule of thumb"
  - 5% probably leads to little bias
  - >20% poses serious threats to validity

Appraisal checklist - RAMMbo

Study biases

1. Recruitment
   - Who did the subjects represent?
2. Allocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial's start?
3. Maintainence
   - Were outcomes ascertained & analysed for most patients?
   - Were the groups treated equally?
4. **M**easurements
   - Were patients and clinicians "**b**linded" to treatment? OR
   - Were measurements **o**bjective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993

Follow-up in DVT study?

- 231 randomised (115 to stockings; 116 none)
- 200 analysed
  - 27 were unable to attend for subsequent ultrasound
  - 2 were excluded from analysis because they were upgraded to business class
  - 2 were excluded from analysis because they were taking anticoagulants

See figure on page 1486

Scurr et al, Lancet 2001; 357:1485-89

Measurement Bias - minimizing differential error

- Objective or
- Blinded
  - Participants?
  - Investigators?
  - Outcome assessors?
  - Analysts?
- Papers should report WHO was blinded and HOW it was done



Figure 1: The authors: double blinded versus single blinded



Figure 2: The authors blinded and masked

Schulz and Grimes. Lancet, 2002

ARTICLES

Frequency and prevention of symptomless deep-vein thrombosis in long-haul flights

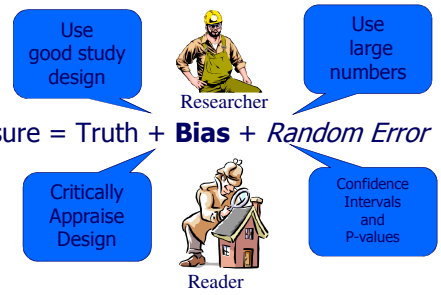
John H Scurr, Samuel J Machin, Se

Evaluation
Most passengers removed their stockings on completion of their journey. The nurse removed the stockings of those passengers who had continued to wear them. A further duplex examination was then undertaken with the technician unaware of the group to which the volunteer had been randomised

Methods

Duplex ultrasonography was used to assess the deep veins before and after travel.

Fundamental Equation of Error



Appraisal checklist - RAMMbo

Study biases

1. **R**ecruitment
   - Who did the subjects represent?
2. **A**llocation
   - Was the assignment to treatments randomised?
   - Were the groups similar at the trial's start?
3. **M**aintainence
   - Were the groups treated equally?
   - Were outcomes ascertained & analysed for most patients?
4. **M**easurements
   - Were patients and clinicians "**bl**inded" to treatment? OR
   - Were measurements **o**bjective & standardised?

Study statistics (p-values & confidence intervals)

Guyatt. JAMA, 1993

Two methods of assessing the role of chance

- P-values (Hypothesis Testing)
  - use statistical test to examine the 'null' hypothesis
  - associated with "p values" - if p<0.05 then result is statistically significant
- Confidence Intervals (Estimation)
  - estimates the range of values that is likely to include the true value

Relationship between p-values and confidence intervals - if the value corresponding to 'no effect' (RR of 1 or treatment difference of 0) falls outside the CI then the result is statistically significant

What is a fair test for treatments?

- ✓Why do we need a comparison?
  - PI**C**O
- ✓How can comparisons be fair?
  - RAMMbo
- How do we assess the role of chance?

P-values (Hypothesis Testing) - in DVT study

- Incidence of DVT
  - Stocking group - 0
  - No Stocking group - 0.12

Risk difference = 0.12 - 0 = 0.12 (P=0.001)

The probability that this result would only occur by chance is 1 in 1000 → statistically significant

Confidence Intervals (Estimation) - in DVT study

- Incidence of DVT
  - Stocking group - 0
  - No Stocking group - 0.12

Risk difference = 0.12 - 0 = 0.12

(95% CI, 0.058 - 0.20)

The true value could be as low as 0.058 or as high as 0.20 - *but is probably closer to 0.12*

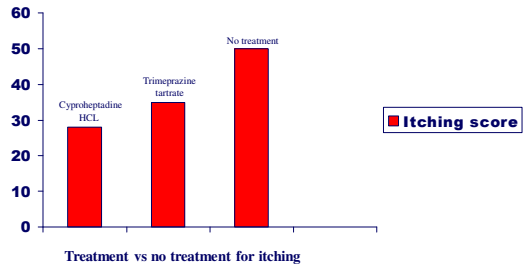
Since the CI does not include the 'no effect' value of '0' → the result is statistically significant

Causes of an "Effect" in a controlled trial

- Who would now consider wearing stockings on a long haul flight?

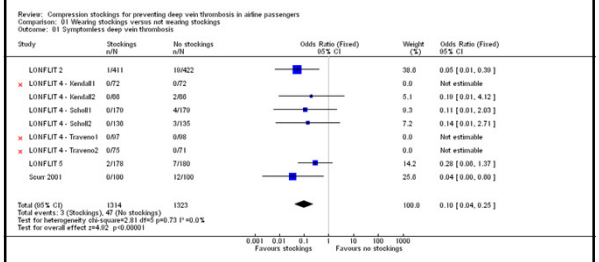
Placebo effect

Trial in patients with chronic severe itching



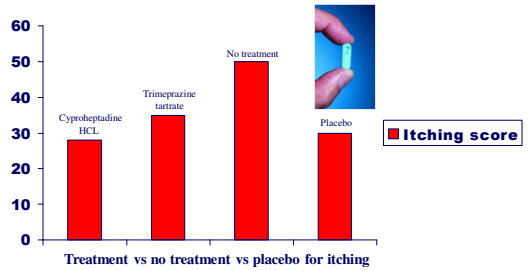
Treatment vs no treatment for itching

M Clarke, S Hopewell, E Juszczak, A Eisinga, M Kjeldstrøm
Compression stockings for preventing deep vein thrombosis in airline passengers
***Cochrane Database of Systematic Reviews* 2006 Issue 4**



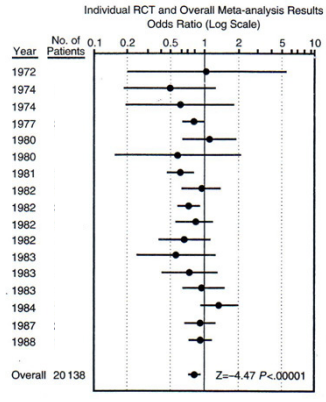
Placebo effect

Trial in patients with chronic severe itching



Treatment vs no treatment vs placebo for itching

Placebo effect - attributable to the expectation that the treatment will have an effect



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

What is your question?

Search for a systematic review

Does the PICO of the review fit that of your question?

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

* McAlister Annals of Intern Med 1999

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.

- Population
- Intervention
- Comparison
- Outcome(s)

Is the review any good? FAST appraisal

- **Q**uestion – What is the PICO?
- **F**inding
  - Did they find most studies?
- **A**ppraisal
  - Did they select good ones?
- **S**ynthesis
  - What to they all mean?
- **T**ransferability of results



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

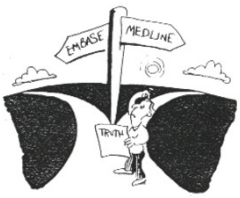
Data Sources and Search Strategies

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 | APPRAISE | SYNTHESISE | TRANSFERABLE

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 didn't use
  - Search again! (*snowballing*)



FIND | APPRAISE | SYNTHESISE | TRANSFERABLE

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

FIND | APPRAISE | SYNTHESISE | TRANSFERABLE

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*
  - Positive SUBMITTED more than negative (2.5 times)

*Dickersin, JAMA, 1992

FIND | APPRAISE | SYNTHESISE | TRANSFERABLE

Which are biased? Which OK?

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

FIND | APPRAISE | SYNTHESISE | TRANSFERABLE

Registered vs Published Studies

Ovarian Cancer chemotherapy: single v combined

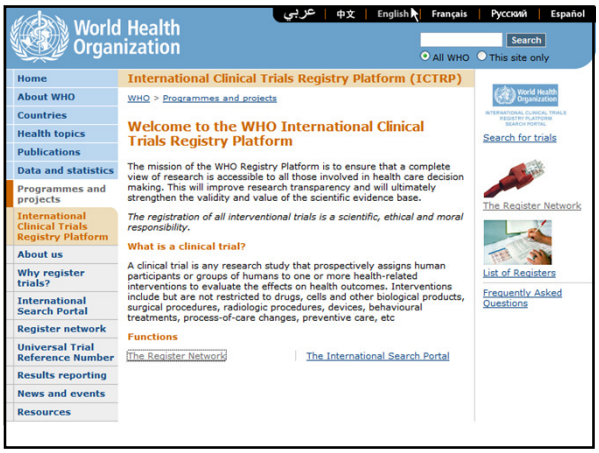
	Published
No. studies	16
Survival ratio	1.16
95% CI	1.06-1.27
P-Value	0.02

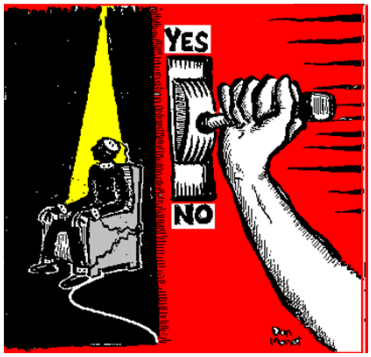
Simes, J. Clin Oncol, 86, p1529

FIND | APPRAISE | SYNTHESISE | TRANSFERABLE

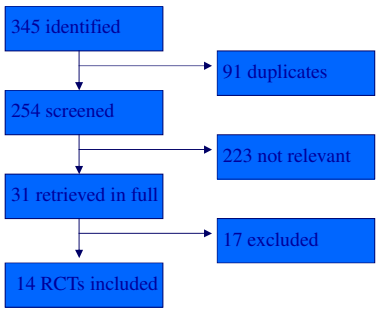
Publication Bias: Solution

- 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



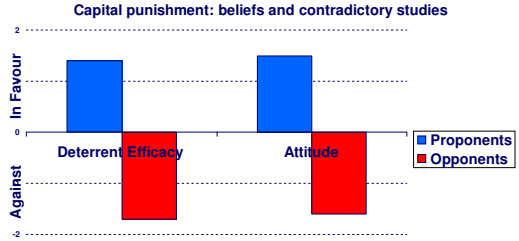


Flowchart



Selective Criticism of Evidence

Biased appraisal increases polarization



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

APPRAISE & select studies

Did they select only the good quality studies?



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)

FIND | **APPRAISE** | SYNTHESISE | TRANSFERABLE

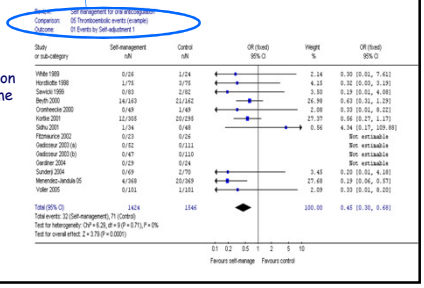
Assessment: How can you avoid biased selection of studies?

- Assessment and selection should be:
  - Standardized "Objective" OR
  - Blinded to Results
- Cochrane Handbook has appraisal 'Risk of Bias' guide

* assessment of quality blind to study outcome

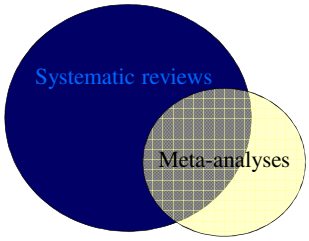
Review: Self management for oral anticoagulation
Comparison: 05 Thromboembolic events (example)
Outcome: 01 Events by Self-adjustment 1

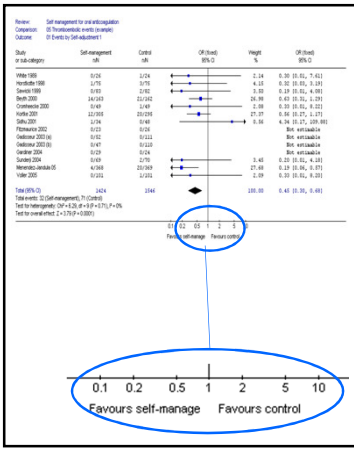
there's a label to tell you what the comparison is and what the outcome of interest is



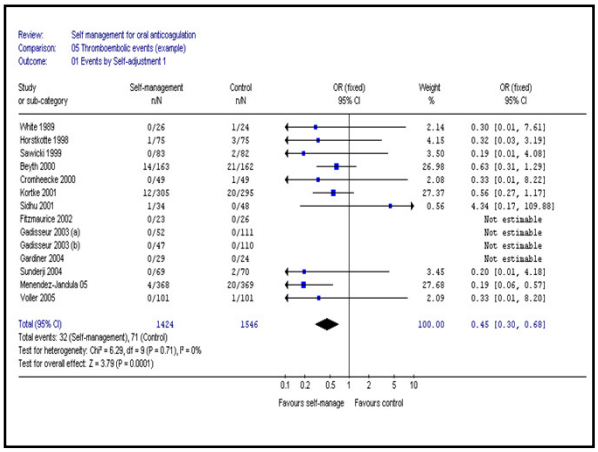
What is a meta-analysis?

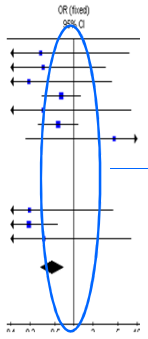
Optional part of a systematic review



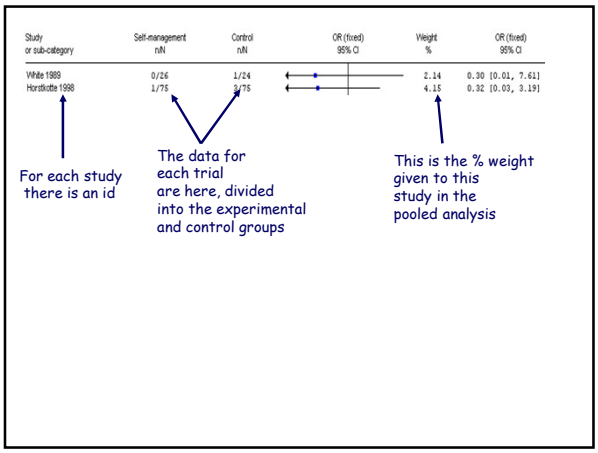


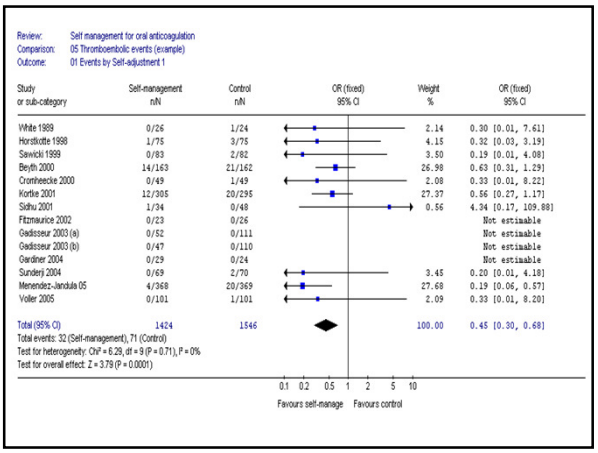
At the bottom there's a horizontal line. This is the scale measuring the treatment effect.

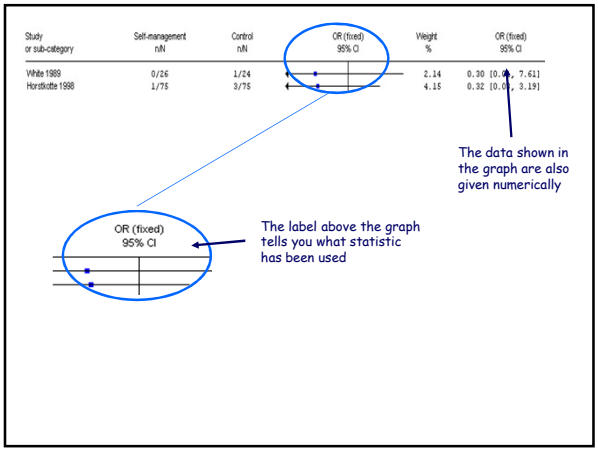




The vertical line in the middle is where the treatment and control have the same effect – there is no difference between the two



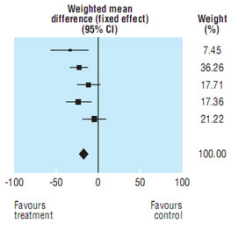


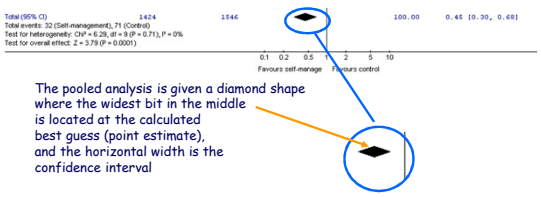


Meta-analysis (Forest) plot

The figure on the right is from Figure 3. See if you can answer the following questions about this plot.

1. How many studies are there?
2. How many studies favour treatment?
3. How many studies are statistically significant?
4. Which is the largest study?
5. Which is the smallest study?
6. What is the combined result?





The pooled analysis is given a diamond shape where the widest bit in the middle is located at the calculated best guess (point estimate), and the horizontal width is the confidence interval

Note on interpretation

If the confidence interval crosses the line of no effect, this is equivalent to saying that we have found no statistically significant difference in the effects of the two interventions

FIND | APPRAISE | **SYNTHESISE** | TRANSFERABLE

Meta-analysis (Forest) plot

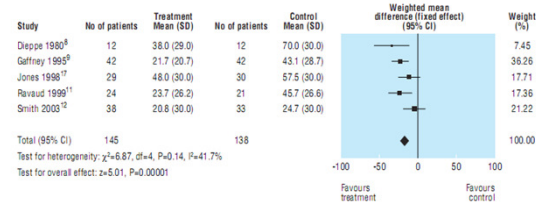


Fig 4 Visual analogue scale for pain up to two weeks after steroid injection in knee

Weighting studies

- More weight to the studies which give us more information
  - More participants
  - More events
  - More precision
- Weight is proportional to the precision

FIND | APPRAISE | SYNTHESISE | **TRANSFERABLE**

Meta-analysis (Forest) plot

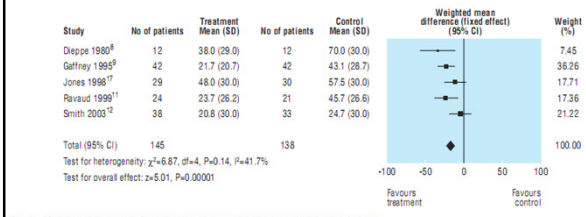
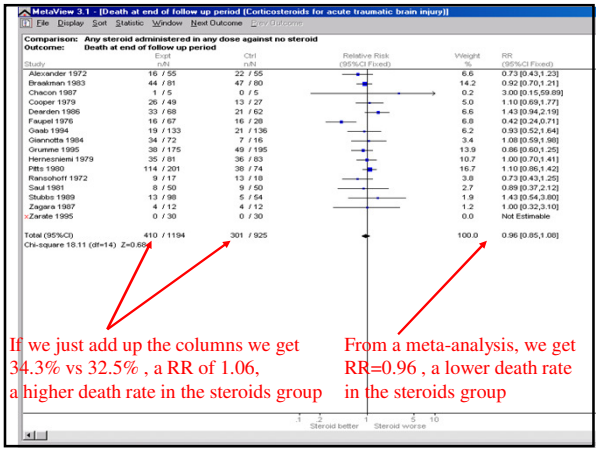


Fig 4 Visual analogue scale for pain up to two weeks after steroid injection in knee

Are the results similar across studies? 3 tests

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



If we just add up the columns we get 34.3% vs 32.5% , a RR of 1.06, a higher death rate in the steroids group

From a meta-analysis, we get RR=0.96 , a lower death rate in the steroids group

FIND | APPRAISE | SYNTHESISE | **TRANSFERABLE**

Are these trials different?

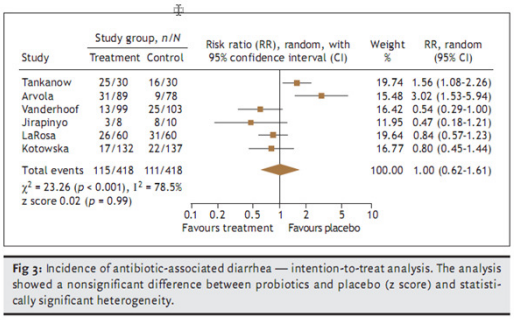


Fig 3: Incidence of antibiotic-associated diarrhea — intention-to-treat analysis. The analysis showed a nonsignificant difference between probiotics and placebo (z score) and statistically significant heterogeneity.

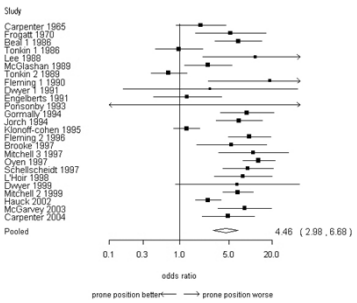
FIND | APPRAISE | SYNTHESISE | **TRANSFERABLE**

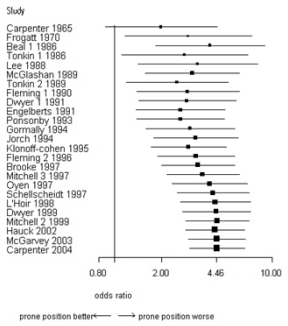
Transferable? Use in my patients

Is the AVERAGE effect similar across studies?

- If NO, then WHY?
  - Study methods - biases
  - PICO
- If YES, then 2 questions
  - Effect in different individuals?
  - Which version of treatment?

Risk of SIDS and sleeping position





Cumulative meta-analysis

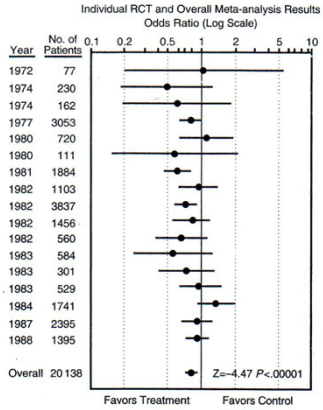
When did we know that sleeping position affected mortality?

Conclusion
EBM and Systematic Review

- EBM (quick & dirty)
  - Ask Question
  - Search
  - Appraise
  - Apply
- Time: 90 seconds
- < 20 articles
- This patient survives!

- Systematic Review
  - Ask Question
  - Search ++++ x 2
  - Appraise x 2
  - Synthesize
  - 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



