Conducting Systematic Reviews and Meta-analyses

Ian Scott
Director of Internal Medicine
Director, Clinical Services Evaluation Unit
Princess Alexandra Hospital
A/Prof Medicine, University of Queensland

Brisbane, Australia

3rd Asia Pacific Conference on Evidence-based Medicine
Hong Kong
Nov 26-28, 2004
The challenge of Archie Cochrane

“It is surely a great criticism of our profession that we have not organised a critical summary, by speciality or subspeciality, adapted periodically, of all relevant randomised controlled trials.”

1979
Systematic review: a definition

• A formal identification, assessment, and synthesis of all primary evidence relevant to a particular question

• *Meta-analysis*: synthesis of quantitative outcomes using statistical methods for standardising and analysing aggregated data

• **Main Steps**
  - Formulating a focussed question
  - Extensive search for all studies
  - Appraisal of quality and selection of studies
  - Systematic synthesis of selected studies
Not a light undertaking

Analysis of workload of 37 meta-analyses

- Av hours 1139 (216-2518)
  • 30 person-weeks FTE
- 588 hrs: protocol development, search and retrieval
- 144 hrs: statistical analysis
- 206 hrs: report writing
- 201 hrs: administration

\[ \text{No. hrs} = 721 + 0.243x - 0.0000123x^2 \]

where \(x=\text{no. potential citations before applying exclusion criteria}\)

Allen & Oilken JAMA 1999
Why do a systematic review?

- Increased statistical power to detect beneficial effect (or harm) overall or in pre-specified patient subgroups

- Testing of robustness and generalisability of pooled result
  - Similar results across different settings and design
  - Homogeneity vs heterogeneity

- In cases of heterogeneity, ability to generate new hypotheses (and trials) to assess reasons for heterogeneity

- Standardisation of clinical practice based on an assessment of all relevant trials
SYSTÉMATIQUE Review: Steps

Formulation of question(s)

Finding

Assessing

Synthesising

Applicability of Results
Choosing a good question (to reward the effort involved)

- **Burden of disease or condition**
  - Prevalence; duration
  - Severity: mortality, morbidity, QOL
  - Costs: to individual, society

- **Intervention**
  - Frequency of (or potential for) use
  - Replicability, cost

- **Ability to change practice**
  - Uncertainty (trials with different results)
  - Variation in practice
  - Level of established preference
  - Timing (or relevance/topicality)
  - Motivation to change

- **Feasibility of assessment**
  - Availability of data
  - Cost of undertaking search

- **Other**
  - Interest and motivation of reviewers
  - Ethical, political and social considerations

Adapted from Counsell Ann Intern Med 1997
SYSTEMATIC Review: Steps

Finding

Assessing

Synthesising
Planning study selection

• Begin with a well defined question
  - PICO

• Choose selection criteria that fit your question
  - Define each component
    • Standardised, clearly-defined terms or conventions
  - Settings
  - Study design
  - Currency of test or intervention
  - Dichotomous vs continuous outcome measures

• Construct and pre-test study selection form
  - include log of excluded studies and reasons

• Write a detailed protocol
  • TIP: try a pilot search with PubMed Clin Queries
An example

Do anticoagulants improve outcomes in patients with acute ischaemic stroke vs no treatment?

**Type of exposure**
- Do anticoagulants improve outcomes in patients with acute ischaemic stroke vs no treatment?

**Type of outcome**
- Death
- Disability
- Major h’nage
- Recurrent stroke
- DVT

**Type of person**
- Clinical Dx
- CT or MRI
- Age, sex, severity
- Within 48 hours

**Type of control**
- No anticoagulants

**Inclusion criteria**
- Unfractionated heparin
- LMW heparin
- Heparinoids
- Oral agents

**Search strategy**
- Index terms for anticoagulants
- Free-text terms for relevant agents
- Journals, conferences, books in coagulation, haematology
- Drug companies, trialists

**Type of study**
- RCTs only
- Unconfounded Placebo or open control

**Index and free-text terms**
- Index and free-text terms for stroke
- Journals, conferences geriatrics, neurology, stroke
- Trialists in stroke medicine
- Index and free-text terms for controlled trials
Efficient searching

• Using inclusive search strings
  - Synonyms, wildcard (* or $), exploded MeSH
• Getting clues on improving search string
  - Hidden studies from retrieved references; expert citations
• Using multiple databases, and multiple languages
• Using methodological terms and filters
  - Study type, years of publication
• When stuck: handsearching, contacting experts, accessing databases of ‘grey literature’
Optimal MEDLINE search strategy

#1 RANDOMIZED-CONTROLLED-TRIAL in PT
#2 CONTROLLED-CLINICAL-TRIAL in PT
#3 RANDOMIZED-CONTROLLED-TRIALS
#4 RANDOM ALLOCATION
#5 DOUBLE-BLIND-METHOD
#6 SINGLE-BLIND-METHOD
#7 #1 or #2 or #3 or #4 or #5 or #6
#8 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#9 #7 not #8
#10 CLINICAL-TRIAL in PT
#11 explode CLINICAL-TRIALS
#12 (clin* near trial*) in TI
#13 (clin* near trial*) in AB
#14 (singl* or doubl* or trebl* or tripl*) near (blind* or mask*)
#15 (#14 in TI) or (#14 in AB)
#16 PLACEBOS
#17 placebo* in TI
#18 placebo* in AB
#19 random* in TI
#20 random* in AB
#21 RESEARCH-DESIGN
#22 #10 or #11 or #12 or #13 or #15 or #16 or #17 or #18 or #19 or #20 or #21
#23 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#24 #22 not #23
#25 #24 not #9
#26 TG=COMPARATIVE-STUDY
#27 explode EVALUATION-STUDIES
#28 FOLLOW-UP-STUDIES
#29 PROSPECTIVE-STUDIES
#30 control* or prospectiv* or volunteer*
#31 (#30 in TI) or (#30 in AB)
#32 #26 or #27 or #28 or #29 or #31
#33 TG=ANIMAL not (TG=HUMAN and TG=ANIMAL)
#34 #32 not #33
#35 #34 not (#9 or #25)
#36 #9 or #25 or #35
Types of Evidence: therapy

- Primary - research reports
  - Randomised trials
  - Cohort studies
  - Case-control studies
  - Case reports

STRONGEST

WEAKEST
FINDING all Studies
Sources of bias

- MEDLINE -> 50% of published studies
- Other databases: EMBASE, CINHAL, ...
- Previous systematic reviews: Cochrane, DARE, CumSearch

PROBLEMS:
- Finding overpublished work
  - Duplicate publications common
- Finding unpublished work
  - Negative trials unpublished?
Sources of studies

Near patient testing: 75 articles

SciCitation  MEDLINE  Experts  HandSearch

Unique  Non-unique

McManus et al BMJ 1998
Duplicate publication bias

SR of 19 trials of efficacy of prophylactic IV ondansetron to prevent post-op vomiting
Tramer et al BMJ 1997
Outcome reporting bias

*Mandel 1987 NEJM*
Efficacy of 2-week course of amoxycillin +/- decongestants in 518 infants and children with otitis media + effusion
Outcome: rate of resolution of effusion at 4 weeks
Result: OR rate of resolution = 2.0 in amoxycillin group higher in patients with unilateral effusion

*Cantekin et al JAMA 1991*
Re-analysis according to measures used:
- Tympanometric evaluation (p=0.12)
- Measure of improved hearing (p=0.31)
- Otoscopic ‘judgement’ (p=0.014)

But sensitivity analysis of otoscopic judgement: NS
Delayed follow-up at 6 weeks after treatment:
- OR recurrence of effusion = 2-6 in antibiotic treated group
Resolution of effusion no different between groups (p=0.47)
## Publication bias

Registered vs Published Studies

<table>
<thead>
<tr>
<th>Published</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. studies</td>
<td>16</td>
</tr>
<tr>
<td>Survival ratio</td>
<td>1.16</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.06-1.27</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.02</td>
</tr>
</tbody>
</table>

SR of ovarian cancer chemotherapy trials (multiple vs single agents)
Simes Clin Oncol 1989
## Publication bias
Registered vs Published Studies

<table>
<thead>
<tr>
<th></th>
<th>Published</th>
<th>Registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. studies</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Survival ratio</td>
<td>1.16</td>
<td>1.05</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.06-1.27</td>
<td>0.98-1.12</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.02</td>
<td>0.25</td>
</tr>
</tbody>
</table>

SR of ovarian cancer chemotherapy trials (multiple vs single agents)
Simes Clin Oncol 1989
Publication Bias

Negative studies less likely to be published than 'positive':

Follow-up of 737 studies at Johns Hopkins (Dickersin, JAMA, 1992)

OR submission: pos vs neg studies = 2.5

Stern & Simes, BMJ 1996
Altruism and trust lie at the heart of research on human subjects. Altruistic individuals volunteer for research because they trust that their participation will contribute to improved health for others and that researchers will minimise risks to participants. In return for the altruism and trust that make clinical research possible, the research enterprise has an obligation to conduct research ethically and to report it honestly. Honest reporting begins with revealing the existence of all clinical studies, even those that reflect unfavourably on a research sponsor’s product.

Unfortunately, selective reporting of trials does occur, and it distorts the body of evidence available for clinical decision-making. Researchers (and journal editors) are generally most enthusiastic about the publication of trials that show either a large effect of a new treatment (positive trials) or equivalence of two approaches to treatment (non-inferiority trials). Researchers (and journals) typically are less excited about trials that show that a new treatment is inferior to standard treatment (negative trials) and even less interested in trials that are neither clearly positive nor clearly negative, since inconclusive trials will not in themselves change practice. Irrespective of their scientific interest, trial results that place financial interests at risk are particularly likely to remain unpublished and hidden from public view. The interests of the sponsor or authors notwithstanding, anyone should be able to learn of any trial’s existence and its important characteristics.

The case against selective reporting is particularly compelling for research that tests interventions that could enter mainstream clinical practice. Rather than a single trial, it is usually a body of evidence, consisting of many studies, that changes medical practice. When research sponsors or investigators conceal the presence of selected trials, these studies cannot influence the thinking of patients, clinicians, other researchers, and experts who write practice guidelines or decide on insurance-coverage policy. If all trials are registered in a public repository at their inception, every trial’s existence is part of the public record and the many stakeholders in clinical research can explore the full range of clinical
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
  - etc (see Cochrane Handbook)

- **Unethical NOT to make results available**
  - Whether published or not, data submitted to registry
Language bias

n=40 pairs of articles

38% + articles in German vs 67% + articles in English

Egger et al Lancet 1997
Trials in LOE have:

\[
\begin{align*}
0.2 & \quad 0.5 & \quad 1 & \quad 2 & \quad 5 \\
0.2 & \quad 0.5 & \quad 1 & \quad 2 & \quad 5 \\
\end{align*}
\]

\[
\begin{align*}
n=42 & \quad n=8 \\
n=50 & \quad n=48 \\
\end{align*}
\]

Language bias

Oxman-Guyatt quality score

\[
\begin{align*}
n=50 & \quad n=32 & \quad n=48 \\
Lang. Restricted SR & \quad Lang. Incl/LE SR & \quad Lang. Incl/LOE SR \\
CM SR & \quad CAM SR \\
\end{align*}
\]

Moher et al Health Technol Assess 2003
SYSTEMATIC Review: Steps

- Finding
- Assessing
- Synthesising
The need for a standardised assessment procedure

- Need to avoid bias in how studies are:
  - Selected for inclusion
  - Appraised for quality
Selective Criticism of Evidence

28 reviewers assessed one “study”
results randomly positive or negative

<table>
<thead>
<tr>
<th></th>
<th>&quot;Positive&quot;</th>
<th>&quot;Negative&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relevance</strong></td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td><strong>Presentation</strong></td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>3.2</td>
<td></td>
</tr>
</tbody>
</table>

Cog Ther Res 1977: 161-75)
Assessment of Quality and Selection of Studies

- **Quality is difficult to define:**
  - Trial methodology (internal validity); clinical relevance; quality of reporting; generalisability

- **Quality varies:**
  Standardized assessment (?blind*) using composite quality scales (40)
  Group/rank by quality features

- **Select a threshold**
  e.g. all randomized, double blind trials with high follow-up
## Composite Quality Appraisal 
**Scales for Interventions**

### Iain Chalmers - 1989

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Valid randomization</td>
<td>a. All patients included (ITT)</td>
<td>a. Double blind</td>
</tr>
<tr>
<td>b. Not “blinded”</td>
<td>b. Withdrawals not described (ITT)</td>
<td>b. Outcome assessment blind</td>
</tr>
<tr>
<td>c. No randomization mentioned</td>
<td>c. “Efficacy” analysis or &gt; 15% lost</td>
<td>c. Not blinded</td>
</tr>
</tbody>
</table>

### Jadad - 1996

<table>
<thead>
<tr>
<th>1. Method of Allocation</th>
<th>2. Blinding</th>
<th>3. Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Described as randomised</td>
<td>Described as double blind</td>
<td>Attrition described for each group (incl. patients lost or excluded)</td>
</tr>
<tr>
<td>Allocation sequences appropriately generated</td>
<td>Control Rx indistinguishable</td>
<td></td>
</tr>
</tbody>
</table>
Critical Appraisal Checklists

• Review of trial “quality scales”
  - 25 scales and 9 checklists
  - contain 3 to 57 items
  - take 10 to 45 minutes

• How should they be used?
  ✍️ Weighting
  ⏎ Sub-groups / Sensitivity analyses
  ⏎ Meta-regression with quality features

Problems with Quality Scales

SR of trials (n=17) of LMW heparin vs standard heparin for thromboprophylaxis in general surgical patients

*Nurmohamed et al* Lancet 1992

Re-examined by

*Juni et al* JAMA 1999
**Effect of study features on effect size in 250 trials**

<table>
<thead>
<tr>
<th></th>
<th>Ratio of odds ratios</th>
<th>p-value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inadequate randomisation</strong></td>
<td>0.95</td>
<td>0.58</td>
<td>Similar effects</td>
</tr>
<tr>
<td><strong>Inadequate allocation concealment</strong></td>
<td>0.70</td>
<td>&lt;0.001</td>
<td>Exaggerated effects</td>
</tr>
<tr>
<td><strong>Inappropriate Exclusions</strong></td>
<td>1.07</td>
<td>0.32</td>
<td>Similar effects</td>
</tr>
<tr>
<td><strong>No double-blinding</strong></td>
<td>0.83</td>
<td>0.01</td>
<td>Exaggerated effects</td>
</tr>
</tbody>
</table>

Schultz et al, JAMA 1995; 273:408-12
Component approach

<table>
<thead>
<tr>
<th>No of trials</th>
<th>RR 95%CI</th>
<th>Ratio RR</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>All trials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Trials stratified by:

Concealment of allocation
- Adequate: 6
- Inadequate/unclear: 11

Blinding of outcomes
- Yes: 11
- No: 6

Intention-to-treat analysis
- Yes: 7
- No: 10

Juni, Altman, Egger 2001
### Scales or components?

<table>
<thead>
<tr>
<th></th>
<th>Non-Cochrane reviews</th>
<th>Cochrane reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial quality assessed</strong></td>
<td>n=204</td>
<td>n=36</td>
</tr>
<tr>
<td>Scales</td>
<td>52 (67%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Components</td>
<td>20 (26%)</td>
<td>33 (92%)</td>
</tr>
</tbody>
</table>

Moher et al Health Technol Assess 1999

**Recommendation:** Apply sensitivity analysis to the components of study quality that you feel are important in the context of your meta-analysis.
Executing Study
Selection and Appraisal

- Ensure methods are reliable, impartial and explicit
- Follow protocol and record progress
- Review each study independently and in duplicate
- Appraise studies according to:
  - Rigor (sensitivity analysis)
  - Reasons for differences among study results
  - Need to provide readers with sufficient information with which to judge applicability of review to clinical practice
- Consider blinding to study outcomes for controversial topics
- Correspond with study authors to confirm study characteristics
- Minimise effects of publication bias by including abstracts, unpublished reports
SYSTEMATIC Review: Steps

Finding

Assessing

Synthesising
General objectives of synthesising data

• Combine the statistical results of multiple trials to provide a weighted average effect of the intervention under consideration

  - More weight given to trials providing more information - larger sample size, more events
  - Average effect = summary point estimate
  - Precision = 95% CI
Choosing a summary measure of effect

- **Dichotomous**
  - Risk difference: $I_E - I_C$ (-ve RD = benefit)
  - Relative risk: $I_E / I_C$ (RR <1 = benefit)
  - Odd ratio: $\text{Odds}_E / \text{odds}_C$ (OR <1 = benefit)
  - Hazard ratio: $P^+_E / P^+_C$ (HR <1 = benefit)

- **Continuous**
  - Mean difference: $x_E - x_C$ (-ve MD = benefit)
  - Standardised MD: $x^s_E - x^s_C$ (-ve MD = benefit) - effect size
  - Weighted MD (average or pooled)
  - Standardised weighted MD (average or pooled standardised difference)
## Data extraction and result presentation

<table>
<thead>
<tr>
<th>Trial</th>
<th>N</th>
<th>Intervention</th>
<th>Population and other content-specific items</th>
<th>Random(^n) procedure</th>
<th>Blinding</th>
<th>F/U</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>324</td>
<td>20mg daily</td>
<td>Information on participants, methods and outcomes</td>
<td>Central computer</td>
<td>Double</td>
<td>90% at 2 yr</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>25mg bd</td>
<td></td>
<td>Envelope</td>
<td>Single</td>
<td>80% at 5 yr</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>987</td>
<td>10-25mg</td>
<td>Not stated</td>
<td>None</td>
<td>98% at 1 yr</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Glasziou et al 2001
Graphical presentation

<table>
<thead>
<tr>
<th>Study</th>
<th>Surgical</th>
<th>Medical</th>
<th>Odds ratio (95% CI)</th>
<th>Pooled odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clagett</td>
<td>0/15</td>
<td>0/14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MACE</td>
<td>3/36</td>
<td>0/35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VA</td>
<td>8/211</td>
<td>2/233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAS</td>
<td>19/825</td>
<td>3/834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L'AURC</td>
<td>2/128</td>
<td>0/109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32/1215</td>
<td>5/1225</td>
<td></td>
<td>4.51 (2.36 to 8.64)</td>
</tr>
</tbody>
</table>

Log scale

Carotid endarterectomy better
Fixed and random effects estimates

• **Fixed effect model**
  - Assumes single ‘true’ value
  - Between-study variation simply due to chance
  - Within-study error only affects precision of effect estimate
  - Weights $1/d^2$

• **Random effect model**
  - Assumes ‘true’ values varies according to population and circumstances
  - Assumes studies are representative (or random) sample of total population of situations to which results will be applied
  - Between-study variance ($\tau^2$) must also be considered
  - Weights $1/(d^2 + \tau^2)$

Neither model completely correct
Fixed vs random effect model

Relative Risk With 95% CI for Low-Risk Fractures After Treatment With Alendronate

<table>
<thead>
<tr>
<th>Study</th>
<th>Effect Size</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone</td>
<td>0.46</td>
<td>(0.16, 1.37)</td>
</tr>
<tr>
<td>Chesnut</td>
<td>1.62</td>
<td>(0.23, 11.53)</td>
</tr>
<tr>
<td>Greenspan</td>
<td>1.00</td>
<td>(0.11, 9.35)</td>
</tr>
<tr>
<td>Hosking</td>
<td>3.47</td>
<td>(1.22, 9.84)</td>
</tr>
<tr>
<td>Liberman</td>
<td>1.72</td>
<td>(0.66, 4.45)</td>
</tr>
<tr>
<td>Liberman</td>
<td>1.96</td>
<td>(0.35, 11.15)</td>
</tr>
<tr>
<td>McClung</td>
<td>0.44</td>
<td>(0.07, 2.90)</td>
</tr>
<tr>
<td>Black</td>
<td>1.06</td>
<td>(0.73, 1.55)</td>
</tr>
<tr>
<td>Cumming</td>
<td>0.76</td>
<td>(0.59, 0.97)</td>
</tr>
</tbody>
</table>

Pooled Fixed-Effects Estimate: 0.90 (0.74, 1.09)

Pooled Random-Effects Estimate: 1.05 (0.72, 1.53)

Guyatt 2000
Are the studies combinable?

• **Need to assess:**
  - *Clinical diversity*
    - Variation in intervention
    - Timing, intensity, co-interventions
    - Variation in population
    - Stage or severity of disease
  - *Methodological diversity*
    - Variation in study methods (biases)
    - Variation in outcome measure
      - Different measures, timing
  - *Statistical heterogeneity*

• **Are variations in results between studies consistent with chance?**
  
  Test of heterogeneity have low power
  
  • Between-study variance ($\tau^2$) using same unit of measure
  • $\chi^2$ (Breslow-Day or Cochran Q)
    - values where $\chi^2 > df$ with $p < 0.1$ indicate heterogeneity
Exploring heterogeneity
Subgroup analysis

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Outcome</th>
<th>Antibiotic versus Placebo</th>
<th>Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Expt</td>
<td>Ctrl</td>
<td>Peto OR (95%CI Fixed)</td>
</tr>
<tr>
<td>Pain at 24 hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burke91</td>
<td>53 / 112</td>
<td>56 / 117</td>
<td>38.6</td>
</tr>
<tr>
<td>Thalin85</td>
<td>58 / 159</td>
<td>58 / 158</td>
<td>49.7</td>
</tr>
<tr>
<td>vanBuchem81</td>
<td>13 / 47</td>
<td>11 / 40</td>
<td>11.8</td>
</tr>
<tr>
<td>Subtotal (95%CI)</td>
<td>124 / 318</td>
<td>125 / 315</td>
<td>100.0</td>
</tr>
<tr>
<td>Chi-square 0.00 (df=2) Z=0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Pain at 2-7 days |
| Burke91 | 20 / 111 | 29 / 114 | 21.4 | 0.65 [0.34, 1.22] |
| Halsted88 | 17 / 62  | 7 / 27   | 8.3  | 1.08 [0.39, 2.97] |
| Kaleida91 | 19 / 488 | 38 / 492 | 29.9 | 0.50 [0.29, 0.85] |
| Mygind81  | 15 / 72  | 29 / 77  | 17.3 | 0.45 [0.22, 0.90] |
| Thalin85  | 15 / 158 | 25 / 158 | 19.5 | 0.57 [0.29, 1.10] |
| vanBuchem81 | 4 / 38  | 3 / 46   | 3.6  | 1.68 [0.36, 7.87] |
| Subtotal (95%CI) | 90 / 929 | 131 / 914 | 100.0 | 0.59 [0.44, 0.79] |
| Chi-square 4.20 (df=5) Z=3.53 |
Exploring heterogeneity
Meta-regression

L’Abbe plot
All-cause mortality in trials of mild to moderate hypertension
Hoes et al
J Hypertens 1995
Exploring heterogeneity
Meta-regression

Galbraith plot
OR of IHD in 28 trials
of cholesterol reduction
Law et al BMJ 1994
Exploring heterogeneity
Publication bias
• Drug X has been studied in randomised trials of myocardial infarction
• Of 24, 4 showed a significant benefit; 20 showed no significant benefit

Do the negative trials render the pooled result statistically insignificant?

• Binomial n = 24, p = 0.05, P(r>3) = 0.03
• Drug X is streptokinase (circa 1982)
Sensitivity analysis
Methodological issues

Criteria (no. trials)

Statistical model
- Fixed effects (33)
- Random effects (33)

Concealment of allocation
- Yes (7)
- No/unclear (26)

Blinding
- Double-blind (25)
- Other (8)

Trial size
- <25 deaths (11)
- 25-99 deaths (11)
- <99 deaths (11)

Follow-up
- <1 year (5)
- 1-2 years (24)
- > 2 years (4)

Excluding trial ceased early
- BHAT (32)
- APSI (32)
- BHAT and APSI (31)

All-cause mortality of β-blockers in secondary prevention after MI
*Freemantle et al. BMJ 1999*
Subgroup analysis

Warning:
Differences between subgroups are observational, indirect and across trials and may exist due to confounding
- Direct within-study comparisons are more reliable

ISAT $\beta$-blocker post-AMI

- Present (11 trials)
  - p = 0.009 (1985)
- Absent (14 trials)
  - p = 0.16 (1999)

Hypothetical study

- Group 1
  - Present (12 trials)
  - Absent (21 trials)
- Group 2
- All patients

Yusuf et al Prog Cardiovasc Dis 1985
Freemantle et al BMJ 1999
Glasziou et al 2001
Subgroup analyses - Are the differences real?

• Is subgroup difference suggested by comparisons within rather than between studies?

• Was the subgroup analysis pre-specified or post-hoc?

• Was the subgroup effect one of a small number of hypothesised effects tested?

• Is the magnitude of effect large and statistically significant?

• Is the effect consistent across studies?

• Is there indirect evidence supporting observed effect?
Individual patient data

Contacting trialists and requesting individual patient data for pooling rather than just trial summary estimates

• **Advantages:**
  - Increase statistical power
  - Analyse by time to event – survival analysis
  - More flexible analysis of patient subgroups and outcomes
  - Better identification of trials, more balanced interpretation of results, wider endorsement and dissemination, future collaborations
  - Ability to add small amounts of new or additional data

• **Disadvantages:**
  - More time consuming and costly
  - Reviewers need wider range of skills
  - Inability to include IPD from all relevant trials
Reporting Results

• Keep the needs of the user in mind
• Put the results in context:
  - What was already known
  - Why was this SR done
  - What has this SR added
  - What are the summary results for all measured outcomes (benefit/harm) in terms of RR, OR, RD and NNT
    • Evidence tables with measures of precision
    • Balance sheet
  - What are the implications for
    • Clinical practice
    • Future research
“... doing a meta-analysis is easy, doing one well is hard.”

Ingram Olkin
Interactive

After hearing this talk what is your attitude now towards undertaking a systematic review

1. All too hard - give up, don’t start!
2. Better understand the process - helpful starting point
3. Already in the process of doing a SR and this has provided helpful hints
4. This is doable - rearing to go!
5. Unsure