Performing, Reporting, and Interpreting a Meta-Analysis of Trials

Robin Christensen
BSc, MSc, Phd (Biostatistician)

Professor of Clinical Epidemiology, adj.
Head of Musculoskeletal Statistics Unit,
The Parker Institute
Disclosure: Potential Conflicts of Interest (COI)

RC: has received grant support and/or provided expert advice and/or presentations for Abbott/AbbVie, Amgen, HealthCare Pharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Cambridge Weight Plan, Eli Lilly, Genzyme, Hospira, Ipsen, Laboratoires Expanscience, MSD, Mundipharma, Norpharma, Novartis, Pfizer, Roche, Sobi, and Wyeth;

RC reports being involved in health-care initiatives and research that could benefit from wide uptake of comparative effectiveness research (including Cochrane Collaboration, OMERACT, IDEOM, the GRADE Working Group, RADS in Denmark).
IMPORTANT READING!

(Meta-Analysis: Need to Know)

Annals of Internal Medicine

The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration

Alessandro Liberati, MD, DrPH; Douglas G. Altman, DSc; Jennifer Tetzlaff, BSc; Cynthia Mulrow, MD, MSc; Peter C. Gøtzsche, MD, DrMedSci, MSc; John P.A. Ioannidis, MD; Mike Clarke, BA, DPhil; P.J. Devereaux, MD, BSc, PhD; Jos Kleiinen, MD, PhD; and David Moher, PhD
How to Read a Systematic Review and Meta-analysis and Apply the Results to Patient Care

Mohammad Hassan Murad, MD, MPH; Victor M. Montori, MD, MSc; John P. A. Ioannidis, MD, DSc; Roman Jaeschke, MD, MSc; P. J. Devereaux, MD, PhD; Kameshwar Prasad, MD, DM, FRCPE; Ignacio Neumann, MD, MSc; Alonso Carrasco-Labra, DDS, MSc; Thomas Agoritsas, MD; Rose Hatala, MD, MSc; Maureen O. Meade, MD; Peter Wyer, MD; Deborah J. Cook, MD, MSc; Gordon Guyatt, MD, MSc

Systematic review and/or Meta-analysis

• Systematic Review:
  - The application of scientific strategies that limit bias to the systematic assessment, critical appraisal and synthesis of all relevant studies on a specific topic

• Meta-Analysis:
  - Statistical method(s) to combine (means and standard errors) and summarize the results of several studies; not necessarily systematic – “simple” estimation across

“A good meta-analysis can only be based on a thorough systematic review!”
Systematic review and/or Meta-analysis

Identification
- # of records identified through database searching
- # of additional records identified through other sources
  
  # of records after duplicates removed

Screening
  
  # of records screened
  
  # of records excluded

Eligibility
  
  # of full-text articles assessed for eligibility
  
  # of full-text articles excluded, with reasons

Included
  
  # of studies included in qualitative synthesis

# of studies included in quantitative synthesis (meta-analysis)
Evidence-Based Medicine, v. 1.0

Category

1A  Meta-analysis of RCTs

1B  At least one RCT

2A  At least one controlled study without randomisation

2B  At least one quasi-experimental study

3   Descriptive studies, such as comparative, correlation or case-control studies

4   Expert committee reports or opinions and/or clinical experience of respected authorities

Modif. from Gordon Guyatt (1991)
Evidence-Based Medicine, v. 1.0

EBM v. 1.0 is out – We now communicate evidence in the form of confidence in the estimates ("GRADE")
What is “GRADE”? 

#1: It is an ACRONYM (G-R-A-D-E) 

#2: Supports and captures previous “evidence initiatives” 

#3: Replaces the previous ‘Evidence-Based Medicine’ (EBM) paradigm 

#4: GRADE is EBM v. 2.0 

#5: GRADE is “The new shit” ..........

Robin.Christensen@Regionh.dk
Began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care.

A transparent and structured process for developing and presenting summaries of evidence.

GRADE provides guideline developers with a comprehensive and transparent framework for carrying out the steps.

Gordon Guyatt  
Andy Oxman  
Holger Schünemann
Organizations that endorse the use of GRADE
GRADE Motive and Outline:

Guidelines should inform clinicians what
*The quality of the underlying evidence is and whether
*Recommendations are strong or weak (conditional)

Quality of the Evidence (reflect our confidence) | Strength of Recommendation (Weak/Strong - For/Against)
GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

BMJ | 26 APRIL 2008 | VOLUME 336
“I figure there’s a 40% chance of showers, and a 10% chance we know what we’re talking about.”
Evidence Synthesis (eg. Meta-Analysis) from:

Randomized Controlled Trial(s) - - - - - - High Quality

Observational studies - - - - - - - - - - Low Quality
Evidence Synthesis (eg. Meta-Analysis) from:

- Randomized Controlled Trial(s) - - - - - - High Quality

- Moderate Quality

- Observationel studies - - - - - - - - - - Low Quality

- Very Low Quality
Clear definition of different grades of quality of evidence — reflecting our confidence

- **High quality**: Further research is very *unlikely to change* our confidence in the estimate of effect.

- **Moderate quality**: Further research *could* have an impact on our confidence in the estimate of effect and may change the estimate.

- **Low quality**: Further research is very likely to have an important impact on our confidence in the estimate of effect and is *likely to change the estimate*.

- **Very Low quality**: Any estimate of effect is very *uncertain*. 
Rating the quality of the evidence
(judging the confidence in the estimates)
How do you get started?
How do you get started?

How do you eat an elephant?

One bite at a time.

One PICO at a time.
How to get a “GRADE start”

GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables

Gordon Guyatt\textsuperscript{a,b,*}, Andrew D. Oxman\textsuperscript{c}, Elie A. Akl\textsuperscript{m}, Regina Kunz\textsuperscript{d}, Gunn Vist\textsuperscript{c}, Jan Brozek\textsuperscript{a}, Susan Norris\textsuperscript{e}, Yngve Falck-Ytter\textsuperscript{f}, Paul Glasziou\textsuperscript{g}, Hans deBeer\textsuperscript{h}, Roman Jaeschke\textsuperscript{b}, David Rind\textsuperscript{i}, Joerg Meerpohl\textsuperscript{j,k}, Philipp Dahm\textsuperscript{l}, Holger J. Schünemann\textsuperscript{a,b}
KISS: PICO

- Population
- Intervention
- Comparator
- Outcome
PICO

• **Population:** Patients with shoulder pain

• **Intervention:** *image-guided* glucocorticoid injections

• **Comparator:** *(other)* glucocorticoid injections

• **Outcome:** Patient-relevant outcomes
Framing the PICO questions

GRADE guidelines: 2. Framing the question and deciding on important outcomes

Gordon H. Guyatt\textsuperscript{a,}*\textsuperscript{,} Andrew D. Oxman\textsuperscript{b}, Regina Kunz\textsuperscript{c}, David Atkins\textsuperscript{d}, Jan Brozek\textsuperscript{a}, Gunn Vist\textsuperscript{b}, Philip Alderson\textsuperscript{e}, Paul Glasziou\textsuperscript{f}, Yngve Falck-Ytter\textsuperscript{g}, Holger J. Schünemann\textsuperscript{a}
Meta-Analysis

• A meta-analysis combines the results of several studies that address a set of related research hypotheses

• Enables a generalization to the population of studies

• Higher statistical power to detect an effect than in any single study

• ”Easier” to read 1 overall paper instead of all the individual papers available ........
Effect Size measures

Binary (dichotomous) Outcomes:

- Risk Difference
- Risk Ratio (RR, “Relative Risk”); calculated on the “Log scale”
- Odds Ratio (OR); calculated on the “Log scale”

Continuous Outcomes:

- Mean Difference
- Standardized Mean Difference
Effect Size measures

**Binary (dichotomous) Outcomes:**

- Risk Difference
- Risk Ratio (RR, "Relative Risk"); calculated on the "Log scale"

*Odds Ratio (OR); calculated on the "Log scale"

**Continuous Outcomes:**

- Mean Difference
- Standardized Mean Difference
How to decide on statistical “effect sizes”
Meta-Analysis: an example ("single PICO")

- Study #1: \( I_1 - C_1 = ES_1 \)
- Study #2: \( I_2 - C_2 = ES_2 \)
- Study #3: \( I_3 - C_3 = ES_3 \)

Meta-Analysis: \( \sum ES_i \Rightarrow \text{Average } ES_{1-3} \)
Meta-Analysis:

an example (“single PICO”)

• Study #1:
  \[ I_1 - C_1 = ES_1 \]

• Study #2:
  \[ I_2 - C_2 = ES_2 \]

• Study #3:
  \[ I_3 - C_3 = ES_3 \]

• Meta-Analysis:
  \[ \sum_{i} ES_i \Rightarrow \text{Average } ES_{1-3} \]
Meta-Analysis:
What’s on your screen?
Random vs Fixed effects models

- A **fixed-effect meta-analysis** estimates a single effect that is assumed to be common to every study, while a **random-effects meta-analysis** estimates the mean of a distribution of effects.

- The key **difference** is that the **random-effects model** captures the true effect(s) across studies that we *a priori* do not think will show us the same overall mean;

- The reason for preferring the **random-effects model per default** (rather than the fixed-effect analysis), is that if the effects sizes are homogenous, then the random-effects model reduces to the fixed-effect model! (i.e. *No case = No Problem*)
A practical ‘rule of thumb’ from Dr. Robs:

If the 95% Confidence Limits from the random effects model “cover” the point estimate from fixed......

....don’t worry about the consequences.
Why Random effects is the obvious default choice

RESEARCH METHODS & REPORTING

Interpretation of random effects meta-analyses

Richard D Riley, Julian P T Higgins, Jonathan J Deeks

Cite this as: BMJ 2011;342:d549
doi: 10.1136/bmj.d549
(Down-) Rating the quality of evidence

- Study limitations (RoB)
- Imprecision (95% CI)
- Inconsistency of results ($I^2$)
- Indirectness of evidence (PICO)
- Publication bias likely (Funnel plot)

**RCTs**: High Quality Evidence
- Moderate Quality Evidence
**Observational studies**: Low QE
- Very-Low Quality Evidence
The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials

Flaws in the design, conduct, analysis, and reporting of randomised trials can cause the effect of an intervention to be underestimated or overestimated. The Cochrane Collaboration’s tool for assessing risk of bias aims to make the process clearer and more accurate.

Julian P T Higgins senior statistician¹, Douglas G Altman director², Peter C Gøtzsche director³, Peter Jüni head of division⁴, David Moher senior scientist⁵⁶, Andrew D Oxman senior researcher⁷, Jelena Savović postdoctoral fellow⁸, Kenneth F Schulz vice president⁹, Laura Weeks research associate⁵, Jonathan A C Sterne professor of medical statistics and epidemiology⁸, Cochrane Bias Methods Group, Cochrane Statistical Methods Group

Higgins JPT, BMJ 2011
Standard Consideration of Study Quality

- High-quality clinical guidelines pay careful attention to the methodologic quality of the studies that form the basis of their recommendations
- Randomization (i.e., sequence generation)
- Concealment of treatment allocation
- Blinding (incl. Participants, Therapists, and Assessors)
- Completeness of follow-up (incl. transparent reporting)
- Appropriate Intention-to-Treat Population
- Selective outcome reporting
- Industry involvement (?)
Stratified analysis

(huge inspiration from Jüni et al)

Trials: Inadequate

Trials: Unclear

Trials: Adequate

= very serious risk of bias

Favors Placebo  Risk Ratio (RR [95% CI])  Favors Drug

0.01  0.10  1.00  10.00  100.00
How do we evaluate internal validity?

GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)

Gordon H. Guyatt\textsuperscript{a,\textdagger}, Andrew D. Oxman\textsuperscript{b}, Gunn Vist\textsuperscript{b}, Regina Kunz\textsuperscript{c}, Jan Brozek\textsuperscript{a}, Pablo Alonso-Coello\textsuperscript{d}, Victor Montori\textsuperscript{e}, Elie A. Akl\textsuperscript{f}, Ben Djulbegovic\textsuperscript{g.h,i}, Yngve Falck-Ytter\textsuperscript{j}, Susan L. Norris\textsuperscript{k}, John W. Williams Jr.\textsuperscript{l}, David Atkins\textsuperscript{m}, Joerg Meerpohl\textsuperscript{n.o}, Holger J. Schünemann\textsuperscript{a}
Interpreting a meta-analysis:

Imprecision; a wide (95%) confidence interval on the summary analysis - and/or few events observed
Meta-Analysis:
Imprecise, is the diamond “all over the place”?
GRADE guidelines 6. Rating the quality of evidence—imprecision

Gordon H. Guyatt\textsuperscript{a,b,*}, Andrew D. Oxman\textsuperscript{c}, Regina Kunz\textsuperscript{d,e}, Jan Brozek\textsuperscript{a}, Pablo Alonso-Coello\textsuperscript{f}, David Rind\textsuperscript{g}, PJ Devereaux\textsuperscript{a}, Victor M. Montori\textsuperscript{h}, Bo Freyschuss\textsuperscript{i}, Gunn Vist\textsuperscript{c}, Roman Jaeschke\textsuperscript{b}, John W. Williams Jr.\textsuperscript{j}, Mohammad Hassan Murad\textsuperscript{h}, David Sinclair\textsuperscript{k}, Yngve Falck-Ytter\textsuperscript{l}, Joerg Meerpohl\textsuperscript{m,n}, Craig Whittington\textsuperscript{o}, Kristian Thorlund\textsuperscript{a}, Jeff Andrews\textsuperscript{p}, Holger J. Schünemann\textsuperscript{a,b}
Interpreting a meta-analysis:

Inconsistency [$I^2 (%)$] and Heterogeneity [Q-test]
Heterogeneity and Inconsistency

• Inconsistency refers to an unexplained heterogeneity of results.

• Inconsistency of studies’ results in a meta-analysis reduces the confidence in the estimate.

• Heterogeneity is usually assessed with a test for homogeneity ($Q$-test), but problems of power can give misleading results.

• The quantity $I^2$ (ranging from 0-100%) describes the degree of inconsistency across studies in a meta-analysis.

Higgins J et al, 2003 BMJ VOLUME 327
Inconsistency: statistical interpretation

Inconsistency ($I^2 \%$) rather than Heterogeneity (Q-test)

Measuring inconsistency in meta-analyses
Julian P T Higgins, Simon G Thompson, Jonathan J Deeks, Douglas G Altman

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

Inconsistency, do you feel that the included studies are telling you different “stories”?
What is Inconsistency?
Interpreting a meta-analysis:

## Indirectness [trust worthy inference from PICO]

<table>
<thead>
<tr>
<th>We want to treat 'P'</th>
<th>We (only) have data on 'P*'</th>
</tr>
</thead>
<tbody>
<tr>
<td>'I' would be without nursing; not in-hospital</td>
<td>'I+' included extra attention &amp; rescue drugs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>'C' would probably never be a Placebo</th>
<th>'C' - Placebo has never been replaced with a &quot;competitor&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>'O' of patient importance?</td>
<td>'Surrogate Outcomes' = no implications.....</td>
</tr>
</tbody>
</table>
Meta-Analysis:

Applicability, Are the studies directly relevant for the patients you’ll be treating?

?
What is Indirectness?

GRADE guidelines: 8. Rating the quality of evidence—indirectness

Gordon H. Guyatt\textsuperscript{a,b,*}, Andrew D. Oxman\textsuperscript{c}, Regina Kunz\textsuperscript{d,e}, James Woodcock\textsuperscript{f}, Jan Brozek\textsuperscript{a}, Mark Helfand\textsuperscript{g}, Pablo Alonso-Coello\textsuperscript{h}, Yngve Falck-Ytter\textsuperscript{i,j}, Roman Jaeschke\textsuperscript{b}, Gunn Vist\textsuperscript{c}, Elie A. Akl\textsuperscript{k}, Piet N. Post\textsuperscript{l}, Susan Norris\textsuperscript{m}, Joerg Meerpohl\textsuperscript{n,o}, Vijay K. Shukla\textsuperscript{p}, Mona Nasser\textsuperscript{q}, Holger J. Schünemann\textsuperscript{a,b},

The GRADE Working Group\textsuperscript{l}

Journal of Clinical Epidemiology 64 (2011) 1303–1310
Interpreting a meta-analysis:

Publication bias; statistically non-significant results never published - i.e., not included in the analysis.
Meta-Analysis:

Publication bias: *Do we have reason to suspect that some trials are left out (e.g. studies with v different outcome)*?
What is Publication Bias?

ORIGINAL ARTICLE

GRADE guidelines: 5. Rating the quality of evidence—publication bias

Gordon H. Guyatt a,b,*, Andrew D. Oxman c, Victor Montori d, Gunn Vist c, Regina Kunz c, Jan Brozek a, Pablo Alonso-Coello f, Ben Djulbegovic g,h,i, David Atkins j, Yngve Falck-Ytter k, John W. Williams Jr. l, Joerg Meerpohl m,n, Susan L. Norris o, Elie A. Akl p, Holger J. Schünemann a
Health Care Question (PICO)

- Deciding on important outcomes (≤ 7 major outcomes)
- Systematic review (PICO) (RCTs & Observational)
- Scrutiny of eligible literature
- Evidence synthesis and/or Meta-analysis
- Generate an estimate (95%CI) for each outcome

Rating the quality of evidence

- Study limitations (RoB)
- Imprecision (95% CI)
- Inconsistency of results (I²)
- Indirectness of evidence (PICO)
- Publication bias likely (Funnel plot)
- Large magnitude of effect
- Dose response
- Confounders likely minimize the effect

RCTs: High Quality Evidence
  - Moderate QE
  - Very-Low QE

Observational studies: Low QE
Introduction to Network Meta-Analysis

Robin Christensen

BSc, MSc, Phd (Biostatistician)

Professor of Clinical Epidemiology, Adj.

Head of Musculoskeletal Statistics Unit,

The Parker Institute
Topics

• What is comparative effectiveness research?

• We use both Direct, Indirect, and Network evidence to evaluate patient-important health outcomes

• A Network meta-analysis combines direct and indirect data (while respecting the original randomizations)

• When is a Network Meta-Analysis to be trusted? (how do we judge the quality of evidence)
Comparative effectiveness research

• Research comparing the benefits and harms of different strategies to prevent, diagnose, treat and monitor health conditions

• Placebo-controlled trials do not answer questions about the comparative benefits and harms of treatment

• Placebo is only ethical when patients who receive it will not be subject to risk of serious or irreversible harm

• FDA (2001) exclude placebo if an “existing treatment is known to prevent irreversible morbidity"
Grading the quality of the evidence

- In the context of evidence synthesis, quality reflects our confidence that the estimates of the effect are correct.

- The lower the quality, the more likely further research would change our confidence in the estimate.

- 'Direct evidence' from RCTs begin as high-quality evidence: "we are very confident that the true effect lies within the 95% confidence interval."

- QoE: Compromised by imprecision, inconsistency, publication bias, indirectness of study results.
Systematic review and/or Meta-analysis

Identification

- # of records identified through database searching
- # of additional records identified through other sources

Screening

- # of records after duplicates removed
- # of records screened
  - # of records excluded

Eligibility

- # of full-text articles assessed for eligibility
  - # of full-text articles excluded, with reasons

Included

- # of studies included in qualitative synthesis
- # of studies included in quantitative synthesis (meta-analysis)
Meta-analysis and/or network meta-analysis?

- Identification:
  - # of records identified through database searching
  - # of additional records identified through other sources
  - # of records after duplicates removed

- Screening:
  - # of records screened
  - # of records excluded

- Eligibility:
  - # of full-text articles assessed for eligibility
  - # of full-text articles excluded with reasons

- # of studies included in qualitative synthesis
  - # of studies included in quantitative synthesis (meta-analysis)

**NOT KOSHER**
PICO

- **Patients**
- **Intervention**
- **Comparator**
- **Outcome**
PICO
(the logic works best for direct evidence!)

- **Patients**
  - Obese individuals (BMI > 30; & motivated!)

- **Intervention**
  - Drug #A OR Drug#B

- **Comparator**
  - Drug #C

- **Outcome(s)**
  - >10% Weight Loss
Indirect and direct evidence! (?)

Drug#A

? trials

Drug#B

? trials

Placebo

Drug#C

? trials
A network of RCTs:
Indirect evidence only (i.e., no H2H comp.)

A
4 trials
(1077 vs. 923)

B
3 trials
(948 vs. 699)

C
4 trials
(800 vs. 800)

Placebo
Generic Example (no real data): PICO

- Defining the patient Population
  - Active: 'A'

- Interventions of interest
  - Active: 'B'
  - Active: 'C'

- Comparators (relevant control groups)
  - Control: 'P'

- Outcome measures of interest

  e.g., a binary outcome
### Summarizing Evidence:

Indirect evidence from forest plot

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Drug</th>
<th>Placebo</th>
<th>Risk Ratio IV, Random, 95% CI</th>
<th>Risk Ratio IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.1.1 Drug A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study#1 (Drug A)</td>
<td>21</td>
<td>46</td>
<td>3.13 [1.47, 6.65]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug A)</td>
<td>74</td>
<td>231</td>
<td>3.13 [2.04, 4.82]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug A)</td>
<td>210</td>
<td>500</td>
<td>2.10 [1.71, 2.57]</td>
<td></td>
</tr>
<tr>
<td>Study#4 (Drug A)</td>
<td>131</td>
<td>300</td>
<td>2.85 [1.91, 4.24]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1077</td>
<td>923</td>
<td>2.53 [2.01, 3.17]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>436</td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Tau² = 0.02; Chi² = 4.32, df = 3 (P = 0.23); I² = 30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 8.01 (P &lt; 0.00001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.2 Drug B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study#1 (Drug B)</td>
<td>313</td>
<td>500</td>
<td>2.48 [1.99, 3.11]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug B)</td>
<td>138</td>
<td>298</td>
<td>2.20 [1.71, 2.83]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug B)</td>
<td>94</td>
<td>150</td>
<td>2.61 [1.91, 3.56]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>948</td>
<td>699</td>
<td>2.41 [2.08, 2.79]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>545</td>
<td>162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Tau² = 0.00; Chi² = 0.84, df = 2 (P = 0.66); I² = 0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 11.70 (P &lt; 0.00001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3 Drug C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study#1 (Drug C)</td>
<td>12</td>
<td>110</td>
<td>2.38 [0.87, 6.52]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug C)</td>
<td>21</td>
<td>120</td>
<td>1.93 [0.97, 3.82]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug C)</td>
<td>65</td>
<td>320</td>
<td>1.35 [0.96, 1.89]</td>
<td></td>
</tr>
<tr>
<td>Study#4 (Drug C)</td>
<td>40</td>
<td>250</td>
<td>1.61 [1.01, 2.58]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>800</td>
<td>800</td>
<td>1.53 [1.20, 1.96]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>138</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Tau² = 0.00; Chi² = 1.76, df = 3 (P = 0.62); I² = 0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 3.40 (P = 0.0007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>2825</td>
<td>2422</td>
<td>2.23 [1.93, 2.58]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>1119</td>
<td>404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Tau² = 0.02; Chi² = 17.42, df = 10 (P = 0.07); I² = 43%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 10.84 (P &lt; 0.00001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for subgroup differences: Chi² = 11.00, df = 2 (P = 0.004), I² = 81.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Direct and indirect comparisons

• One trial compared treatment A to standard treatment C and another study compared treatment B to standard treatment C

\[ \Delta_{AC} = p_A - p_C \]
\[ \Delta_{BC} = p_B - p_C \]

• We want to obtain an estimate A vs. B?

• When there is no direct evidence, the adjusted indirect comparison may provide useful information on the relative efficacy of competing interventions.

Adjusted indirect comparisons

- A summary measure of the indirect comparison can be computed by subtracting the estimate from studies comparing ‘A versus C’ and from the studies comparing ‘B versus C’:

\[ \Delta_{AB} \approx \Delta_{AC} - \Delta_{BC} \]
Adjusted indirect comparisons

The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials


Department of Clinical Epidemiology and Biostatistics,
McMaster University, Hamilton, Ontario,
Canada, L8N 3Z5

J Clin Epidemiol Vol. 50, No. 6, pp. 683–691, 1997
Copyright © 1997 Elsevier Science Inc.
### Indirect Evidence from Forest Plot

#### 1.1.1 Drug A

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Drug Events</th>
<th>Drug Total</th>
<th>Placebo Events</th>
<th>Placebo Total</th>
<th>Weight</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study#1 (Drug A)</td>
<td>21</td>
<td>46</td>
<td>7</td>
<td>48</td>
<td>3.2%</td>
<td>3.13 [1.47, 6.65]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug A)</td>
<td>74</td>
<td>231</td>
<td>23</td>
<td>225</td>
<td>7.8%</td>
<td>3.13 [2.04, 4.82]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug A)</td>
<td>210</td>
<td>500</td>
<td>100</td>
<td>500</td>
<td>16.4%</td>
<td>2.10 [1.71, 2.57]</td>
<td></td>
</tr>
<tr>
<td>Study#4 (Drug A)</td>
<td>131</td>
<td>300</td>
<td>23</td>
<td>150</td>
<td>8.6%</td>
<td>2.85 [1.91, 4.24]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>1077</td>
<td>923</td>
<td></td>
<td></td>
<td>36.0%</td>
<td>2.53 [2.01, 3.17]</td>
<td></td>
</tr>
</tbody>
</table>

Total events: 436, 153

Heterogeneity: \( \text{Tau}^2 = 0.02; \chi^2 = 4.32, \text{df} = 3 \) \( (P = 0.23); \quad \text{I}^2 = 30\%

Test for overall effect: \( Z = 8.01 \) \( (P < 0.00001) \)

#### 1.1.2 Drug B

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Drug Events</th>
<th>Drug Total</th>
<th>Placebo Events</th>
<th>Placebo Total</th>
<th>Weight</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study#1 (Drug B)</td>
<td>313</td>
<td>500</td>
<td>63</td>
<td>250</td>
<td>15.4%</td>
<td>2.48 [1.99, 3.11]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug B)</td>
<td>138</td>
<td>298</td>
<td>63</td>
<td>299</td>
<td>14.1%</td>
<td>2.20 [1.71, 2.83]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug B)</td>
<td>94</td>
<td>150</td>
<td>36</td>
<td>150</td>
<td>11.5%</td>
<td>2.61 [1.91, 3.56]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>948</td>
<td>699</td>
<td></td>
<td></td>
<td>40.9%</td>
<td>2.41 [2.08, 2.79]</td>
<td></td>
</tr>
</tbody>
</table>

Total events: 545, 162

Heterogeneity: \( \text{Tau}^2 = 0.00; \chi^2 = 0.84, \text{df} = 2 \) \( (P = 0.66); \quad \text{I}^2 = 0\%

Test for overall effect: \( Z = 11.70 \) \( (P < 0.00001) \)

#### 1.1.3 Drug C

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Drug Events</th>
<th>Drug Total</th>
<th>Placebo Events</th>
<th>Placebo Total</th>
<th>Weight</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
<th>Risk Ratio, IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study#1 (Drug C)</td>
<td>12</td>
<td>110</td>
<td>5</td>
<td>109</td>
<td>1.9%</td>
<td>2.38 [0.87, 6.52]</td>
<td></td>
</tr>
<tr>
<td>Study#2 (Drug C)</td>
<td>21</td>
<td>120</td>
<td>11</td>
<td>121</td>
<td>3.8%</td>
<td>1.93 [0.97, 3.82]</td>
<td></td>
</tr>
<tr>
<td>Study#3 (Drug C)</td>
<td>65</td>
<td>320</td>
<td>48</td>
<td>318</td>
<td>10.4%</td>
<td>1.35 [0.96, 1.89]</td>
<td></td>
</tr>
<tr>
<td>Study#4 (Drug C)</td>
<td>40</td>
<td>250</td>
<td>25</td>
<td>252</td>
<td>6.9%</td>
<td>1.61 [1.01, 2.68]</td>
<td></td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>800</td>
<td>800</td>
<td></td>
<td></td>
<td>23.0%</td>
<td>1.53 [1.20, 1.96]</td>
<td></td>
</tr>
</tbody>
</table>

Total events: 138, 89

Heterogeneity: \( \text{Tau}^2 = 0.00; \chi^2 = 1.76, \text{df} = 3 \) \( (P = 0.62); \quad \text{I}^2 = 0\%

Test for overall effect: \( Z = 3.40 \) \( (P = 0.0007) \)

Total (95% CI): 2825, 2422

Total events: 1119, 404

Heterogeneity: \( \text{Tau}^2 = 0.02; \chi^2 = 17.42, \text{df} = 10 \) \( (P = 0.07); \quad \text{I}^2 = 43\%

Test for overall effect: \( Z = 10.84 \) \( (P < 0.00001) \)

Test for subgroup differences: \( \chi^2 = 11.00, \text{df} = 2 \) \( (P = 0.004); \quad \text{I}^2 = 81.8\%

\( P = 0.004 \)
### Adjusted indirect comparisons

1: Organize your data logically

2: Transform: back to their original scale

3: Subtract log-estimates;
   variance of a difference = sum of variances

4: Transform your estimates (Exp[x])

<table>
<thead>
<tr>
<th>DIRECT COMPARISON</th>
<th>RR</th>
<th>LCL95%</th>
<th>UCL95%</th>
<th>InRR</th>
<th>SE[InRR]</th>
<th>Var[InRR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A vs. Placebo</td>
<td>2.53</td>
<td>2.01</td>
<td>3.17</td>
<td>0.9282</td>
<td>0.1151</td>
<td>0.0132</td>
</tr>
<tr>
<td>B vs. Placebo</td>
<td>2.41</td>
<td>2.08</td>
<td>2.79</td>
<td>0.8796</td>
<td>0.0747</td>
<td>0.0056</td>
</tr>
<tr>
<td>C vs. Placebo</td>
<td>1.53</td>
<td>1.20</td>
<td>1.96</td>
<td>0.4253</td>
<td>0.1264</td>
<td>0.0160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADJUSTED INDIRECT COMPARISON</th>
<th>*RR</th>
<th>*LCL95%</th>
<th>*UCL95%</th>
<th>*InRR</th>
<th>*SE[lnRR]</th>
<th>*Var[lnRR]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A vs C</td>
<td>1.65</td>
<td>1.18</td>
<td>2.31</td>
<td>0.5030</td>
<td>0.1709</td>
<td>0.0292</td>
</tr>
<tr>
<td>A vs B</td>
<td>1.05</td>
<td>0.80</td>
<td>1.37</td>
<td>0.0486</td>
<td>0.1372</td>
<td>0.0188</td>
</tr>
<tr>
<td>B vs C</td>
<td>1.58</td>
<td>1.18</td>
<td>2.10</td>
<td>0.4544</td>
<td>0.1468</td>
<td>0.0215</td>
</tr>
</tbody>
</table>
Fine, Got it......

(Why don't we always use the adjusted indirect [Bucher] approach then?????)
The adjusted indirect comparison is *not* valuable when we have no common comparator .........

Tarp S, et al (submitted)
How to Use an Article Reporting a Multiple Treatment Comparison Meta-analysis

Edward J. Mills, PhD, MSc
John P. A. Ioannidis, MD, DSc
Kristian Thorlund, PhD, MSc
Holger J. Schünemann, MD, PhD, MSc
Milo A. Puhan, MD, PhD
Gordon H. Guyatt, MD, MSc

JAMA. 2012;308(12):1246-1253

RESEARCH METHODS & REPORTING

BMJ 2013;346:f2914 doi: 10.1136/bmj.f2914 (Published 14 May 2013)

Demystifying trial networks and network meta-analysis

Networks of randomized clinical trials can be evaluated in the context of a network meta-analysis, a procedure that permits inferences into the comparative effectiveness of interventions that may or may not have been evaluated directly against each other. This approach is quickly gaining popularity among clinicians and guideline developers because it permits methodological advances.
Unlike a contrast-based approach, a network meta-analysis apply “arm-based modeling” while still respecting the randomization.
Didn’t get it; What’s next?

http://cmim.cochrane.org/
network-meta-analysis-toolkit
Evaluating the Quality of Evidence from a Network Meta-Analysis

Georgia Salanti¹, Cinzia Del Giovane², Anna Chaimani¹, Deborah M. Caldwell³, Julian P. T. Higgins³,⁴*
RESEARCH METHODS & REPORTING

A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis

Milo A Puhan¹, Holger J Schünemann², Mohammad Hassan Murad³, Tianjing Li⁴, Romina Brignardello-Petersen⁵, Jasvinder A Singh⁶, Alfons G Kessels⁷, Gordon H Guyatt², for the GRADE Working Group
Minimal guidance exists on how to rate the quality of evidence supporting treatment effect estimates obtained from NMA.

GRADE present a four-step approach to rate the quality of evidence in each of the direct, indirect, and NMA estimates based on methods developed by the GRADE working group.

The quality of evidence supporting NMA estimates varies from high to very low across comparisons, and ratings given to a whole network are uninformative and likely to mislead.
# Estimates of effects and quality ratings for comparison of drugs to prevent osteoporotic hip fractures

<table>
<thead>
<tr>
<th>Comparison</th>
<th>DIRECT</th>
<th>INDIRECT</th>
<th>NETWORK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% confidence interval)</td>
<td>Quality of evidence</td>
<td>Odds ratio (95% credible interval)</td>
</tr>
<tr>
<td>Teriparatide v placebo</td>
<td>—</td>
<td>—</td>
<td>0.42 (0.10 to 1.82)</td>
</tr>
<tr>
<td>Denosumab v placebo</td>
<td>—</td>
<td>—</td>
<td>0.50 (0.27 to 0.86)</td>
</tr>
<tr>
<td>Raloxifene v placebo</td>
<td>0.84 (0.63 to 1.13)</td>
<td>Moderate‡</td>
<td>0.96 (0.53 to 1.78)</td>
</tr>
<tr>
<td>Zoledronate v placebo</td>
<td>—</td>
<td>—</td>
<td>0.50 (0.33 to 0.74)</td>
</tr>
<tr>
<td>Risedronate v placebo</td>
<td>0.17 (0.05 to 0.59)</td>
<td>Low*,‡‡</td>
<td>0.54 (0.36 to 0.75)</td>
</tr>
<tr>
<td>Ibandronate v placebo</td>
<td>—</td>
<td>—</td>
<td>0.49 (0.21 to 1.20)</td>
</tr>
<tr>
<td>Alendronate v placebo</td>
<td>—</td>
<td>—</td>
<td>0.45 (0.27 to 0.68)</td>
</tr>
<tr>
<td>Vitamin D v placebo</td>
<td>1.25 (0.82 to 1.89)</td>
<td>Low*,‡</td>
<td>1.08 (0.61; 1.91)</td>
</tr>
<tr>
<td>Vitamin D+calcium v placebo</td>
<td>0.83 (0.73 to 0.96)</td>
<td>Moderate*</td>
<td>0.54 (0.29 to 0.94)</td>
</tr>
<tr>
<td>Calcium v placebo</td>
<td>—</td>
<td>—</td>
<td>1.14 (0.82 to 1.59)</td>
</tr>
<tr>
<td>Denosumab v teriparatide</td>
<td>—</td>
<td>—</td>
<td>1.17 (0.24 to 5.54)</td>
</tr>
<tr>
<td>Raloxifene v teriparatide</td>
<td>—</td>
<td>—</td>
<td>2.05 (0.47 to 9.47)</td>
</tr>
</tbody>
</table>

*BMJ 2014;349:g5630 doi: 10.1136/bmj.g5630 (Published 24 September 2014)*
Network meta-analysis—highly attractive but more methodological research is needed

Tianjing Li¹, Milo A Puhan¹, Swaroop S Vedula ¹, Sonal Singh², Kay Dickersin¹* and for
The Ad Hoc Network Meta-analysis Methods Meeting Working Group
Network Meta-Analyses (in brief)

Use both direct (head-to-head) RCT evidence and indirect evidence to compare the relative effectiveness of all included interventions.

Quality may be difficult to interpret: the number of interventions evaluated may be large and the approaches may be complex.

Aspects to consider:
- Are primary studies sufficiently homogeneous? (similar in their populations, study designs, and outcomes)
- Is the direct evidence sufficiently similar to the indirect evidence to consider combining
Network Meta-Analysis
- (still) A call for transparency

Robin.Christensen@Regionh.dk