Introduction to using and producing Cochrane reviews

Protocol development II: Data collection and analysis
- Risk of bias assessment
- Introduction to Meta-analysis

14 November 2017 13.00-14.30

Outline I

Protocol development II:
- Risk of bias*

*With an example of a Cochrane protocol
Steps of a systematic review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Source: Cochrane Training [http://training.cochrane.org/]

See Chapter 8 of the Handbook

Components of a protocol

- Title
- Main text
  - Background
  - Objectives
  - Methods
    - Criteria for considering studies for this review
    - Searching methods for identification of studies
    - Data collection and analysis
      - Selection of studies
      - Data extraction and management
      - Assessment of treatment effect
What is bias?

**Systematic error or deviation from the truth**

- systematic reviews depend on included studies
  - incorrect studies = misleading reviews
  - should I believe the results?
- assess each study for risk of bias
  - can’t measure the presence of bias
  - may overestimate or underestimate the effect
  - look for methods shown to minimise risk

Bias is not the same as

**Imprecision**
- random error due to sampling variation
- reflected in the confidence interval

**Quality**
- bias can occur in well-conducted studies
- not all methodological flaws introduce bias

**Reporting**
- good methods may have been used but not well reported

Source: Cochrane Training (http://training.cochrane.org/)
Cochrane ‘Risk of bias’ assessment

• 7 evidence-based domains (RCT)
• review authors’ judgement
  ✓ Low risk of bias
  ✗ High risk of bias
  ? Unclear
• support for judgement
  • evidence/quotes from the paper or other sources
  • review author’s explanation

Source: Cochrane Training (http://training.cochrane.org/)

Domains to address

• Random sequence generation
• Allocation concealment
• Blinding of participants and personnel
• Blinding of outcome assessment
• Incomplete outcome data
• Selective reporting
• Other bias

⚠️ more than one author to assess risk of bias
  how will disagreements be resolved?

Source: Cochrane Training (http://training.cochrane.org/)
### Source of Bias (RCTs)

#### Selection
- Random allocation generation
- Allocation concealment

#### Performance
- Blinding – participants, personnel

#### Detection
- Blinding – outcome assessors

#### Attrition
- Incomplete outcome data

#### Reporting
- Selective outcome reporting

---

### Examples of Summary Descriptions: The Risk of Bias Table

*Figure R.6a: Example of a ‘Risk of bias’ table for a single study (fictional)*

<table>
<thead>
<tr>
<th>Entry</th>
<th>Judgement</th>
<th>Support for judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td>Low risk.</td>
<td>Quote: “patients were randomly allocated.” Comment: Probably done, since earlier reports from the same investigators clearly describe use of random sequences (Cartwright 1990).</td>
</tr>
<tr>
<td>Blinding of participants and personnel (performance bias)</td>
<td>Low risk.</td>
<td>Quote: “double blind, double dummy”; “High and low dose tablets or capsules were indistinguishable in all aspects of their outward appearance. For each drug an identically matched placebo was available (the success of blinding was evaluated by examining the drugs before distribution).” Comment: Probably done.</td>
</tr>
<tr>
<td>Blinding of outcome assessment (detection bias) (all-cause mortality)</td>
<td>Low risk.</td>
<td>Obtained from medical records; review authors do not believe this will introduce bias.</td>
</tr>
</tbody>
</table>

Source: Cochrane Training (http://training.cochrane.org/)
Random sequence generation

Low risk – unpredictable
- random number table
- computer random number generator
- stratified or block randomisation
- minimisation
- low tech - coin toss, shuffling cards or envelopes, throwing dice, drawing lots

High risk – predictable
- quasi-random – date of birth, day of visit, ID or record number, alternate allocation
- non-random – choice of clinician or participant, test results, availability

See Section 8.9 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)

Allocation concealment

Low risk – unpredictable
- central allocation (phone, web, pharmacy)
- sequentially numbered, sealed, opaque envelopes
- sequentially numbered, identical drug containers

High risk – predictable
- random sequence known to staff in advance
- envelopes or packaging without all safeguards
- non-random, predictable sequence

See Section 8.10 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)
Blinding of participants & personnel

Low risk ✔
- blinding, and unlikely that the blinding could have been broken
- no blinding or incomplete blinding, but outcome unlikely to be influenced

High risk ✗
- no blinding, incomplete or broken blinding, and outcome likely to be influenced

See Section 8.11 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)

Blinding of outcome assessment

Low risk ✔
- blinding, and unlikely that the blinding could have been broken
- no blinding, but measurement unlikely to be influenced

High risk ✗
- no blinding or broken blinding, and measurement likely to be influenced

See Section 8.12 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)
Incomplete outcome data

Low risk
• no missing data
• reasons for missing data not related to outcome
• missing data balanced across groups, and reasons similar
• proportion missing or plausible effect size not enough to have a clinically relevant effect

High risk
• reasons related to outcome, and imbalance in numbers or reasons
• proportion missing or plausible effect size enough to have a clinically relevant effect
• ‘as-treated’ analysis with substantial departure from allocation
• inappropriate use of imputation

See Section 8.13 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)

Selective reporting

Low risk
• protocol is available and all pre-specified outcomes of interest to the review reported in the pre-specified way
• protocol not available but it is clear that all pre-specified and expected outcomes of interest are reported

Unclear risk
• most studies will be judged in this category

High risk
• outcomes not reported as pre-specified or expected
  • e.g. missing, added, subsets, unexpected measurements or methods
• outcomes reported incompletely so they cannot be entered in a meta-analysis

See Section 8.14 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)
Other sources of bias

Low risk

• study appears to be free of other sources of risk

High risk

• issues specific to the study design
  • carry-over in cross-over trials
  • recruitment bias in cluster-randomised trials
  • non-randomised studies
  • baseline imbalance
  • blocked randomisation in unblinded trials
  • differential diagnostic activity
  • other bias

See Section 8.15 of the Handbook
Source: Cochrane Training (http://training.cochrane.org/)

Presentation of assessments of risk of bias

Risk of bias summary presents all of the judgments in a cross-tabulation of study by entry

Figure 8.6.c: Example of a ‘Risk of bias’ summary figure

Source: Cochrane Training (http://training.cochrane.org/)
Presentation of assessments of risk of bias

Risk of bias graph illustrates the proportion of studies with each of judgement across for each entry in the tool.

What about non-randomised studies?

- risk of bias must still be carefully assessed
- you may wish to add domains to your assessment
- you may wish to use an alternative, appropriate tool

A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI)
Edited by Jonathan AC Sterne, Julian PT Higgins and Barney C Reeves
on behalf of the development group for ACROMBAT-NRSI
Version 1.0.0, 24 September 2014

The Risk Of Bias In Non-randomized Studies – of Interventions (ROBINS-I) assessment tool
(version for cohort-type studies)
Developed by: Jonathan AC Sterne, Miguel A Hernán, Baranyi, Mohammed T Anani, Isabelle Boutron, James Carpenter, Avi Ramay, Deborah Regidor, Hannah Rothstein, Lakshmi Sandhu, P Valentine, Hugh Waddington, Elisabeth Waters, Penny Witting
Version 1.0, 1 August 2016

The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses
Nonrandomised studies, including case-control and cohort studies, can be challenging to implement and conduct. Assessment of the quality of such studies is essential for a proper understanding of nonrandomised studies. The Newcastle-Ottawa Scale (NOS) is an ongoing collaboration between the Universities of Newcastle, Australia and Ottawa, Canada. It was developed to assess the quality of nonrandomised studies with its design, content and ease of use directed to the lack of incorporating the quality assessments in the decision of study groups, the comparability of the groups, and the ascertainment of external exposure or outcomes of interest for case control or cohort studies respectively. The goal of this project is to develop an instrument providing an easy and consistent tool for quality assessment of nonrandomised studies to be used in a comparative tool.
Effectiveness of tranexamic acid in reducing blood loss during cytoreductive surgery for advanced ovarian cancer

Protocol

Chumnan Kietpeerakool C, Arnon Supoken A, Malinee Laopaiboon, Pisake Lumbiganon P.
First published: 3 June 2015
Editorial Group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group
DOI: 10.1002/14651858.CD011732

Assessment of risk of bias in included studies

We will assess and report on the methodological risk of bias of the included studies according to chapter 8 of the Cochrane Handbook (Higgins 2011), which recommends the explicit reporting of the following individual elements for RCTs:

1. Selection bias: random sequence generation and allocation concealment;
2. Performance bias: blinding of participants and personnel (participants and treatment providers);
3. Detection bias: blinding of outcome assessment;
4. Attrition bias: incomplete outcome data;
5. Reporting bias: selective reporting of outcomes.

Example

What to include in your protocol

• check with your CRG for standard text
• brief description of risk of bias assessment tool
  • list domains
  • refer to Handbook Chapter 8
• more than one author will assess risk of bias
• how will disagreements will be resolved?
• are there specific domains you consider to be important for the review?
• how will you incorporate findings into your analysis?

Source: Cochrane Training (http://training.cochrane.org/)
Take home message

• biased studies may lead to misleading reviews
• seven domains of bias to be assessed
• describe what happened in detail and give your judgement
• consider the possible effects and use appropriate caution in interpreting your results

Source: Cochrane Training (http://training.cochrane.org/)

Acknowledgements

• Materials for this presentation are based in part on material adapted from Cochrane Training (http://training.cochrane.org)
References


Outline II

Protocol development II:
- principles of meta-analysis
- steps in a meta-analysis
- presenting your results
- what is heterogeneity?
- assumptions about heterogeneity
- identifying heterogeneity

*With an example of a Cochrane protocol
Steps of a systematic review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Source: Cochrane Training (http://training.cochrane.org/)

Components of a protocol

- Title
- Main text
  - Background
  - Objectives
  - Methods
    - Criteria for considering studies for this review
    - Searching methods for identification of studies
    - Data collection and analysis
      - Selection of studies
      - Data extraction and management
      - Assessment of treatment effect
      - Data synthesis
      - Subgroup analysis and investigation of heterogeneity
      - Sensitivity analysis
Effect measure – Binary data

Comparing two groups

<table>
<thead>
<tr>
<th></th>
<th>Headache</th>
<th>No headache</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caffeine</td>
<td>17</td>
<td>51</td>
<td>68</td>
</tr>
<tr>
<td>Decaf</td>
<td>9</td>
<td>55</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>106</td>
<td>132</td>
</tr>
</tbody>
</table>

**effect measures**
- risk ratio/relative risk (RR)
- odds ratio (OR)
- risk difference (RD) (absolute risk reduction)

**all estimates are uncertain, and should be presented with a confidence interval**

Source: Cochrane Training (http://training.cochrane.org/)
Effect measure – Continuous data

Comparing two groups

<table>
<thead>
<tr>
<th>Irritability score</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caffeine</td>
<td>20</td>
<td>9.1</td>
<td>65</td>
</tr>
<tr>
<td>Decaf</td>
<td>33</td>
<td>8.6</td>
<td>67</td>
</tr>
</tbody>
</table>

- effect measures
  - mean difference (MD) (difference of means)
  - standardised mean difference (SMD) (effect size)
- all estimates are uncertain, and should be presented with a confidence interval

Source: Cochrane Training (http://training.cochrane.org/)

Study level

Study A: Outcome data -> Scale 1
Study B: Outcome data -> Scale 2
Study C: Outcome data -> Scale 1
Study D: Outcome data -> Scale 3

Review level

Effect measure

Source: Cochrane Training (http://training.cochrane.org/)
Standardised mean difference

- when different scales used to measure the same outcome
- SMD standardises the results
  - units of standard deviation
  - does not correct for direction – may need to multiply by -1

\[
SMD = \frac{\text{mean of intervention group} - \text{mean of control group}}{\text{pooled standard deviation of both groups}}
\]

When mean difference = 0, there is no difference between the groups

What is a meta-analysis?

- combines the results from two or more studies
- estimates an ‘average’ or ‘common’ effect
- optional part of a systematic review
Why perform a meta-analysis?

- quantify treatment effects and their uncertainty
- increase power
- increase precision
- explore differences between studies
- settle controversies from conflicting studies
- generate new hypotheses

Source: Cochrane Training (http://training.cochrane.org/)

When can you do a meta-analysis?

- more than one study has measured an effect
- the studies are sufficiently similar to produce a meaningful and useful result
- the outcome has been measured in similar ways
- data are available in a format we can use

Source: Cochrane Training (http://training.cochrane.org/)
Steps in a meta-analysis

- identify comparisons to be made
- identify outcomes to be reported and statistics to be used
- collect data from each relevant study
- combine the results to obtain the summary of effect
- explore differences between the studies
- interpret the results

Source: Cochrane Training [http://training.cochrane.org/]

Selecting comparisons

Hypothetical review: Caffeine for daytime drowsiness

caffeinated coffee vs decaffeinated coffee

Source: Cochrane Training [http://training.cochrane.org/]
Selecting outcomes & effect measures

Hypothetical review: Caffeine for daytime drowsiness

caffeinated coffee vs decaffeinated coffee

- asleep at end of trial (RR)
- irritability (MD)
- headaches (RR)

- for each comparison, select outcomes
- for each outcome, select an effect measure
  - may depend on the available data from included studies

Source: Cochrane Training (http://training.cochrane.org/)

Calculating the summary result

“Weighting studies”

- more weight to the studies which give more information
  - more participants, more events, narrower confidence interval
  - calculated using the effect estimate and its variance

- inverse-variance method:

\[
\text{weight} = \frac{1}{\text{variance of estimate}} = \frac{1}{SE^2}
\]

\[
\text{pooled estimate} = \frac{\text{sum of (estimate} \times \text{weight)}}{\text{sum of weights}}
\]

Source: Cochrane Training (http://training.cochrane.org/)
For example

<table>
<thead>
<tr>
<th></th>
<th>Caffeine</th>
<th>Decaf</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amore-Coffea 2000</td>
<td>2/31</td>
<td>10/34</td>
<td>6.6%</td>
</tr>
<tr>
<td>Deliciotta 2004</td>
<td>10/40</td>
<td>9/40</td>
<td>21.9%</td>
</tr>
<tr>
<td>Mama-Kaffa 1999</td>
<td>12/53</td>
<td>9/61</td>
<td>22.2%</td>
</tr>
<tr>
<td>Morrocona 1998</td>
<td>3/15</td>
<td>1/17</td>
<td>2.9%</td>
</tr>
<tr>
<td>Norscafe 1998</td>
<td>19/68</td>
<td>9/64</td>
<td>26.4%</td>
</tr>
<tr>
<td>Oohlahlazza 1998</td>
<td>4/35</td>
<td>2/37</td>
<td>5.1%</td>
</tr>
<tr>
<td>Piazza-Allerta 2003</td>
<td>8/35</td>
<td>6/37</td>
<td>14.9%</td>
</tr>
</tbody>
</table>

Source: Cochrane Training (http://training.cochrane.org/)

Presenting results

Outcome: Headache at 24 hours

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Caffeinated coffee</th>
<th>Decaffeinated coffee</th>
<th>Risk Ratio</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Events</td>
<td>Events</td>
<td>Total</td>
<td>Weight</td>
</tr>
<tr>
<td>Amore-Coffea 2000</td>
<td>31</td>
<td>34</td>
<td>2.22 (0.65, 0.83)</td>
<td></td>
</tr>
<tr>
<td>Deliciotta 2004</td>
<td>40</td>
<td>40</td>
<td>1.11 (0.51, 2.44)</td>
<td></td>
</tr>
<tr>
<td>Mama-Kaffa 1999</td>
<td>53</td>
<td>53</td>
<td>1.53 (0.70, 3.39)</td>
<td></td>
</tr>
<tr>
<td>Morrocona 1998</td>
<td>3</td>
<td>17</td>
<td>1.30 (0.39, 3.81)</td>
<td></td>
</tr>
<tr>
<td>Norscafe 1998</td>
<td>98</td>
<td>64</td>
<td>1.09 (0.47, 2.57)</td>
<td></td>
</tr>
<tr>
<td>Oohlahlazza 1998</td>
<td>36</td>
<td>37</td>
<td>2.11 (0.41, 10.83)</td>
<td></td>
</tr>
<tr>
<td>Piazza-Allerta 2003</td>
<td>35</td>
<td>37</td>
<td>1.41 (0.54, 3.85)</td>
<td></td>
</tr>
<tr>
<td>Total events for all studies</td>
<td>277</td>
<td>290</td>
<td>1.38 (0.96, 2.00)</td>
<td></td>
</tr>
</tbody>
</table>

Total events: 46
Heterogeneity: Chi² = 0.58, df = 6 (P = 0.98), I² = 30%
Test for overall effect: Z = 1.73 (P = 0.08)

Forest plots

Effect measure shown numerically and graphically

Source: Cochrane Training (http://training.cochrane.org/)
Effectiveness of tranexamic acid in reducing blood loss during cytoreductive surgery for advanced ovarian cancer


Example

What to include in the protocol

• how will you decide whether a meta-analysis is appropriate?
• meta-analysis model to be used

Source: Cochrane Training (http://training.cochrane.org/)
Take home message

- there are several advantages to performing a meta-analysis but it is not always possible (or appropriate)
- plan your analysis carefully, including comparisons, outcomes and meta-analysis methods
- forest plots display the results of meta-analyses graphically
- interpret your results with caution

Source: Cochrane Training (http://training.cochrane.org/)

Steps of a systematic review

1. define the question
2. plan eligibility criteria
3. plan methods
4. search for studies
5. apply eligibility criteria
6. collect data
7. assess studies for risk of bias
8. analyse and present results
9. interpret results and draw conclusions
10. improve and update review

Source: Cochrane Training (http://training.cochrane.org/)

See Chapter 9 of the Handbook
What is heterogeneity?

- Variation or differences
  - three broad types:
    - **clinical**
      (Participant – age, Intervention – dose, Outcome – follow-up duration, ways of measuring)
    - **methodological**
      (Design – RCT/non-RCT, Conduct – blind/not)
    - **statistical**
      there will always be some random (sampling) variation between the results of different studies

Source: Cochrane Training (http://training.cochrane.org/)

---

Fixed-effect vs random-effects

- Two models for meta-analysis available in RevMan
- Make different assumptions about heterogeneity
- Pre-specify your planned approach in your protocol

Source: Cochrane Training (http://training.cochrane.org/)
Fixed-effect model

- Assumes all studies are measuring the same treatment effect.
- Estimates that one effect.
- If not for random (sampling) error, all results would be identical.

Random-effects model

- Assumes the treatment effect varies between studies.
- Estimates the mean of the distribution of effects.
- Weighted for both within-study and between-study variation (tau², τ²).

Source: Julian Higgins
Source: Cochrane Training (http://training.cochrane.org/)
Random-effects model

- Almost identical to fixed-effect when there is no heterogeneity
- Similar to fixed-effect but with wider confidence intervals when there is heterogeneity of the sort assumed by RE model
- Different from fixed-effect meta-analyses when results are related to study size
- RE model gives relatively more weight to smaller studies

\[
\text{weight} = \frac{1}{\text{variance within} + \text{variance between}} = \frac{1}{\text{SE}^2 + \text{tau}^2}
\]

Source: Cochrane Training (http://training.cochrane.org/)

No heterogeneity

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Treatment Events</th>
<th>Control Events</th>
<th>Fixed Weight</th>
<th>Risk Ratio M-H, Fixed, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Random Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faschke 2006</td>
<td>1</td>
<td>7</td>
<td>0.5%</td>
<td>0.57 (0.38, 0.86)</td>
<td></td>
<td>3.2%</td>
<td></td>
</tr>
<tr>
<td>Hakim 2005</td>
<td>1</td>
<td>21</td>
<td>1.7%</td>
<td>1.92 (0.81, 4.49)</td>
<td></td>
<td>1.6%</td>
<td></td>
</tr>
<tr>
<td>Maser 1994</td>
<td>1</td>
<td>120</td>
<td>2.7%</td>
<td>1.01 (0.81, 1.25)</td>
<td></td>
<td>2.0%</td>
<td></td>
</tr>
<tr>
<td>Maser 2002</td>
<td>12</td>
<td>67</td>
<td>14.3%</td>
<td>2.22 (0.23, 20.50)</td>
<td></td>
<td>15.0%</td>
<td></td>
</tr>
<tr>
<td>Ohlin 2001A</td>
<td>17</td>
<td>87</td>
<td>39.0%</td>
<td>1.18 (0.80, 1.73)</td>
<td></td>
<td>37.2%</td>
<td></td>
</tr>
<tr>
<td>Ohlin 2001B</td>
<td>7</td>
<td>59</td>
<td>11.1%</td>
<td>1.75 (0.64, 5.00)</td>
<td></td>
<td>11.1%</td>
<td></td>
</tr>
<tr>
<td>Rønnegard 2000</td>
<td>20</td>
<td>115</td>
<td>24.8%</td>
<td>2.22 (0.23, 20.50)</td>
<td></td>
<td>27.7%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>484</td>
<td>100.0%</td>
<td>1.65 [1.12, 2.43]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: \(I^2 = 0.00\), \(Q = 3.14\), \(df = 7\) (\(P = 0.87\)), \(I^2 = 0.00\)

Test for overall effect: \(Z = 2.41\) (\(P = 0.02\))

Adapted from Ohlsson A, Aher SM. Early erythropoietin for preventing red blood cell transfusion in preterm and/or low birth weight infants. Cochrane Database of Systematic Reviews 2006, Issue 3.
Some heterogeneity

<table>
<thead>
<tr>
<th>Study/Subgroup</th>
<th>Chlorpromazine</th>
<th>Placebo</th>
<th>Fixed Weight</th>
<th>Risk Ratio M-H, Fixed, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chawawat 1999</td>
<td>14</td>
<td>21</td>
<td>18</td>
<td>21</td>
<td>0.5%</td>
<td>0.00 [0.49, 1.29]</td>
</tr>
<tr>
<td>Clark 1974a</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.27 [0.09, 0.62]</td>
<td>6.7%</td>
</tr>
<tr>
<td>Clark 1974b</td>
<td>10</td>
<td>53</td>
<td>6</td>
<td>18</td>
<td>2.7%</td>
<td>0.57 [0.24, 1.34]</td>
</tr>
<tr>
<td>Fleming 1980</td>
<td>5</td>
<td>21</td>
<td>13</td>
<td>21</td>
<td>5.0%</td>
<td>0.58 [0.17, 0.89]</td>
</tr>
<tr>
<td>Hart 1956</td>
<td>65</td>
<td>67</td>
<td>78</td>
<td>98</td>
<td>20.4%</td>
<td>0.84 [0.69, 1.10]</td>
</tr>
<tr>
<td>Penn 1948</td>
<td>37</td>
<td>416</td>
<td>70</td>
<td>212</td>
<td>37.6%</td>
<td>0.27 [0.19, 0.35]</td>
</tr>
<tr>
<td>Semple 1951</td>
<td>0</td>
<td>20</td>
<td>12</td>
<td>23</td>
<td>6.1%</td>
<td>0.04 [0.00, 0.63]</td>
</tr>
<tr>
<td>Serafinides 1972</td>
<td>6</td>
<td>14</td>
<td>3</td>
<td>13</td>
<td>1.0%</td>
<td>1.05 [0.59, 2.10]</td>
</tr>
<tr>
<td>Smith 1981</td>
<td>4</td>
<td>17</td>
<td>10</td>
<td>15</td>
<td>2.0%</td>
<td>0.40 [0.19, 1.12]</td>
</tr>
<tr>
<td>Semmeklie 1968</td>
<td>5</td>
<td>15</td>
<td>23</td>
<td>30</td>
<td>6.0%</td>
<td>0.45 [0.22, 0.96]</td>
</tr>
</tbody>
</table>

Total (95% CI) 675 452 100.0% 0.55 [0.47, 0.63]

Test of Heterogeneity: X² = 59.2, df = 9, P < 0.00001; I² = 69%
Test for overall effect: Z = 2.37, P = 0.02


Which to choose?

- plan your approach at the protocol stage
- do you expect your results to be very diverse?
- consider the underlying assumptions of the model
  - fixed-effect
    - may be unrealistic – ignores heterogeneity
  - random-effects
    - allows for heterogeneity
    - estimate of distribution of studies may not be accurate if biases are present, few studies or few events

Source: Cochrane Training (http://training.cochrane.org/)
Identifying heterogeneity

- Visual inspection of the forest plots
- Chi-squared ($\chi^2$) test (Cochrane Q test)
- $I^2$ statistic to quantify heterogeneity

Visual inspection

Forest plot A

Forest plot B

Source: Cochrane Training [http://training.cochrane.org/]
Chi-squared ($\chi^2$) test

- Tests the null hypothesis of homogeneity
  - low power with few studies

\[ Q = \sum_{i=1}^{k} W_i (Y_i - M)^2 \]
\[ Q = \sum_{i=1}^{k} \left( \frac{Y_i - M}{S_i} \right)^2 \]

\( Y_i \) = effect estimate
\( M \) = summary effect
\( k \) = no. of studies
\( w_i \) = weight = \( \frac{1}{\text{variance}} = \frac{1}{S_i^2} \)

\( Q \) = weighted sum of squares (WSS)

\( \text{df.} = k-1 \)

\( I^2 \) statistic

- \( I^2 \) statistic;
  \[ I^2 = \left( \frac{Q - df}{Q} \right) \times 100\% \]

\( I^2 \) describes the percentage of variability due to heterogeneity rather than chance (0% to 100%)

- Low values indicate no, or little, heterogeneity
- High values indicate a lot of heterogeneity
- Roughly,

- \( I^2 \) levels of heterogeneity
  - 25% low
  - 50% moderate
  - 75% high
### What to do about heterogeneity

- **Check that the data are correct**
  - especially if the direction of effect varies
- **If heterogeneity is very high**
  - may choose not to meta-analyse
    - average result may be meaningless in practice
    - consider clinical & methodological comparability of studies
- **Explore heterogeneity**

---

### Outcomes Table

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Treatment</th>
<th>Control</th>
<th>Odds Ratio</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome: bleeding (Y/N)</strong></td>
<td><strong>Treatment</strong></td>
<td><strong>Control</strong></td>
<td><strong>Odds Ratio</strong></td>
<td><strong>Odds Ratio</strong></td>
</tr>
<tr>
<td><strong>Study or Subgroup</strong></td>
<td><strong>Events</strong></td>
<td><strong>Total</strong></td>
<td><strong>Events</strong></td>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>1.3.1 Placebo control</td>
<td>20</td>
<td>89</td>
<td>22</td>
<td>89</td>
</tr>
<tr>
<td>Nelson</td>
<td>20</td>
<td>89</td>
<td>22</td>
<td>89</td>
</tr>
<tr>
<td>Crowther</td>
<td>12</td>
<td>56</td>
<td>15</td>
<td>59</td>
</tr>
<tr>
<td>Dwyer</td>
<td>40</td>
<td>412</td>
<td>56</td>
<td>421</td>
</tr>
<tr>
<td>Hamilton</td>
<td>20</td>
<td>97</td>
<td>31</td>
<td>96</td>
</tr>
<tr>
<td>Hotway</td>
<td>34</td>
<td>143</td>
<td>22</td>
<td>145</td>
</tr>
<tr>
<td>Henderson</td>
<td>3</td>
<td>63</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>Houghton</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Ouyse</td>
<td>57</td>
<td>612</td>
<td>53</td>
<td>617</td>
</tr>
<tr>
<td>Winsterbottom</td>
<td>10</td>
<td>102</td>
<td>26</td>
<td>103</td>
</tr>
<tr>
<td>McKnight</td>
<td>25</td>
<td>76</td>
<td>15</td>
<td>73</td>
</tr>
<tr>
<td>Mustard</td>
<td>34</td>
<td>764</td>
<td>65</td>
<td>854</td>
</tr>
<tr>
<td>Gates</td>
<td>0</td>
<td>32</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>Horsley</td>
<td>52</td>
<td>342</td>
<td>102</td>
<td>341</td>
</tr>
<tr>
<td>Sakata</td>
<td>12</td>
<td>44</td>
<td>4</td>
<td>44</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>2831</td>
<td>2762</td>
<td>100.0%</td>
<td>0.93 [0.80, 1.08]</td>
</tr>
</tbody>
</table>

---

### Diagram

Visual inspection of the forest plot

Q test: Low power - number of studies is small

---

**Source:** Julian Higgins

59

---

Source: Cochrane Training ([http://training.cochrane.org/](http://training.cochrane.org/))
Exploring your results

• What factors appear to modify the effect?
  • clinical diversity (population, interventions, outcomes)
  • methodological diversity (study design, risk of bias)

• Plan your strategy in your protocol
  • identify a limited number of important factors to investigate
  • have a scientific rationale for each factor chosen
  • declare any post-hoc investigations

Source: Cochrane Training (http://training.cochrane.org/)

Two methods available

• subgroup analysis
  • group studies by pre-specified factors
  • look for differences in results and heterogeneity

• meta-regression
  • examine interaction with categorical and continuous variables
  • not available in RevMan

Source: Cochrane Training (http://training.cochrane.org/)
Proceed with caution

• results are observational, not randomised
• be wary of multiple and post hoc comparisons
• may not be useful with few studies
• look for confounding factors
• follow the plan specified in the protocol without over-emphasising particular findings

Interpreting subgroup analyses

• look at results and heterogeneity within subgroups
• are the subgroups genuinely different?
  • if only 2 subgroups – do the confidence intervals overlap?
  • statistical tests for subgroup difference
• can be more confident about:
  • pre-specified analyses

Source: Cochrane Training (http://training.cochrane.org/)
### Participant subgroups

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>4 mg dose Events</th>
<th>Total</th>
<th>2 mg dose Events</th>
<th>Total</th>
<th>Weight</th>
<th>Risk Ratio IV, Fixed, 95% CI</th>
<th>Risk Ratio IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>5.1.1 High dependency smokers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiney 2010</td>
<td>24</td>
<td>116</td>
<td>18</td>
<td>116</td>
<td>20.0%</td>
<td>1.32 [0.76, 2.30]</td>
<td></td>
</tr>
<tr>
<td>Hemera 1995</td>
<td>30</td>
<td>87</td>
<td>13</td>
<td>87</td>
<td>18.3%</td>
<td>2.16 [1.21, 3.82]</td>
<td></td>
</tr>
<tr>
<td>Kondakgo 1987</td>
<td>24</td>
<td>73</td>
<td>16</td>
<td>73</td>
<td>21.2%</td>
<td>1.77 [0.82, 3.30]</td>
<td></td>
</tr>
<tr>
<td>Tonnessen 1993</td>
<td>12</td>
<td>27</td>
<td>4</td>
<td>27</td>
<td>8.5%</td>
<td>1.67 [1.33, 10.00]</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td>383</td>
<td>1315</td>
<td>4</td>
<td>383</td>
<td>67.0%</td>
<td>1.83 [1.34, 2.49]</td>
<td></td>
</tr>
<tr>
<td><strong>Total events</strong></td>
<td>90</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td><strong>Ch² = 3.44, df = 3 (P = 0.39), I² = 13%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test for overall effect</strong></td>
<td>Z = 3.84 (P = 0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5.1.2 Low Dependency Smokers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garsley 2000</td>
<td>16</td>
<td>67</td>
<td>17</td>
<td>67</td>
<td>17.0%</td>
<td>0.94 [0.51, 1.74]</td>
<td></td>
</tr>
<tr>
<td>Hughes 1990</td>
<td>5</td>
<td>13</td>
<td>8</td>
<td>13</td>
<td>7.5%</td>
<td>0.66 [0.26, 1.66]</td>
<td></td>
</tr>
<tr>
<td>Kondakgo 1987</td>
<td>5</td>
<td>17</td>
<td>5</td>
<td>17</td>
<td>7.7%</td>
<td>0.47 [0.19, 1.17]</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td>123</td>
<td>415</td>
<td>115</td>
<td>415</td>
<td>32.2%</td>
<td>0.73 [0.47, 1.15]</td>
<td></td>
</tr>
<tr>
<td><strong>Total events</strong></td>
<td>26</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td><strong>Ch² = 1.60, df = 2 (P = 0.45), I² = 0%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test for overall effect</strong></td>
<td>Z = 1.39 (P = 0.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>426</td>
<td>430</td>
<td>100.0%</td>
<td>1.36 [1.06, 1.75]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total events</strong></td>
<td>116</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td><strong>Ch² = 15.86, df = 6 (P = 0.01), I² = 62%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test for overall effect</strong></td>
<td>Z = 2.39 (P = 0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Source: Cochrane Training (http://training.cochrane.org/)

### Effectiveness of tranexamic acid in reducing blood loss during cytoreductive surgery for advanced ovarian cancer

#### Protocol

Chunnam Kietpeerakool ET, Amornrat Supoken, Malinee Laopaiboon, Pitsake Lumbiganon

First published: 3 June 2013

Editorial Group: Cochrane Gynaecological, Neuro-oncology and Orphan Cancer Group

DOI: 10.1002/14651858.CD011732 View summary

### Assessment of heterogeneity

Clinical heterogeneity will be assessed by considering the characteristics of the participants included in the studies.

We will also assess statistical heterogeneity in each meta-analysis using the $T^2$, $I^2$ and $Chi^2$ statistic (Higgins 2003). We will regard heterogeneity as substantial if $I^2$ is greater than 50% and either $T^2$ is greater than zero, or there is a low $P$ value (less than 0.10) in the $Chi^2$ test for heterogeneity (Deeks 2001).
Data synthesis

We will carry out statistical analysis using Review Manager software (RevMan 2014). The results of the included studies will be pooled in meta-analyses.

- For dichotomous outcomes, the risk ratio will be calculated for each study and these will be pooled.
- For continuous outcomes, the mean difference between the treatment arms will be pooled, if all trials measured the outcome on the same scale; otherwise standardised mean differences will be pooled.

We will use the random-effect model with inverse variance weighting for all meta-analyses (DeSimonian 1986).

We will prepare a ‘Summary of findings’ table to present the results of the meta-analysis, based on the methods described in chapter 11 of the Cochrane Handbook (Schünemann 2011). We will present the results of the meta-analysis for the primary outcome, and harms, as outlined in the Types of outcome measures section.

If we are unable to pool the data using meta-analysis methods, we will conduct a narrative synthesis of the results.

Subgroup analysis and investigation of heterogeneity

If we identify substantial heterogeneity, we will investigate it using subgroup analyses for the following factors:

1. Route of administration of tranexamic acid (intravenous, oral or topical administration).
2. Dose of tranexamic acid (single dose versus multiple doses).
3. Largest preoperative tumour size, excluding ovarian mass (≤5 versus >5 cm).

We will assess subgroup differences by interaction tests available within RevMan (RevMan 2014). We will report the results of subgroup analyses using the χ² statistic and P-value, and the interaction test P-value.

Sensitivity analysis

“Are the findings robust to the decisions made in the process of obtaining them?”

- not the same as subgroup analysis
- testing the impact of decisions made during the review
  - inclusion of studies in the review
  - definition of low risk of bias
  - choice of effect measure
  - assumptions about missing data
  - cut-off points for dichotomised ordinal scales
  - correlation coefficients
- repeat analysis using an alternative method or assumption
  - don’t present multiple forest plots – just report the results
  - if difference is minimal, can be more confident of conclusions
  - if difference is large, interpret results with caution

Source: Cochrane Training (http://training.cochrane.org/)
What to include in your protocol

• Assessment of heterogeneity
  • assessment of comparability of studies before meta-analysis
  • visual inspection and use of statistics such as $I^2$

• Data synthesis
  • fixed-effect or random-effects model (or both)

• Subgroup analyses and investigation of heterogeneity
  • planned subgroup analyses
  • any other strategies for investigating heterogeneity

Take home message

• statistical heterogeneity is the presence of differences between estimated intervention effects greater than expected because of random (sampling) variation alone

• it can be caused by clinical and methodological diversity

• fixed and random-effects models make different assumptions about heterogeneity

• explore any heterogeneity you find
Acknowledgements

• Materials for this presentation are based in part on material adapted from Cochrane Training (http://training.cochrane.org)

References