Combining studies: from heterogeneity to similarity

Jos Verbeek, Jani Ruotsalainen
Cochrane OSH Review Group
Finnish Institute of Occupational Health
Kuopio Finland
A systematic review

1. Well-formulated question (PICO)
2. Thorough search
3. Objective selection of studies
4. Critical assessment of methodological quality
5. Objective data extraction
6. Synthesis of the data
   a) appropriate comparisons of interventions and controls
   b) meta-analysis per comparison
7. Conclusions for practice and research
Scope of a review

• Interventions for.....
  • Cochrane Library
    • intervention* NOT pharmacological in title
      • 411 reviews and 261 protocols

• Resulting in a variety of interventions
  • non-drug
  • complex interventions
    • multi-faceted
    • multi-component
    • behavioural
    • team based
    • community-based
    • rehabilitation
    • exercises for..
    • educational
When to combine studies?

• Usual text in Cochrane Protocol Method Section
  • We will assess clinical heterogeneity by examining types of participants, interventions, and outcomes in each study
  • We will pool data from studies judged to be clinically homogeneous with RevMan 5 software.

• When you are going to look for heterogeneity you will probably find it and then it will be difficult to pool studies

• Better to state: We will look for *similarity* between studies
Heterogeneity? Similarity?

In the review “Interventions for preventing noise-induced hearing loss in workers” you find the following 8 studies. How are they similar assuming that study design is the same in all?

1. Ear muffs vs plugs for preventing hearing loss (STS) in construction workers
2. Legislation versus no legislation for reducing noise exposure (dB) in mines
3. Worker training in ear plug use vs no training for reducing noise exposure (dB) in metal sheet workers
4. Subsidies for employers vs no subsidies for reducing noise exposure (dB) in offshore oil platforms
5. Stronger ear plugs vs lighter ear plugs for preventing hearing loss (STS) in farmers
6. Magnesium vs placebo for preventing hearing loss (STS) in noise-exposed workers
7. Inspections and penalties for preventing noise exposure (dB) in construction industry
8. Occupational health advice to decrease noise exposure (dB) in small companies
Heterogeneity? Similarity?

- In the review “Interventions for preventing noise-induced hearing loss in workers” you find the following 8 studies. How are they similar assuming that study designs are all similar?

1. **Hearing Protection for workers**
   - Ear muffs vs plugs for preventing hearing loss (STS) in construction workers
   - Stronger ear plugs vs lighter ear plugs for preventing hearing loss (STS) in farmers
   - Worker training in ear plug use vs no training for reducing noise exposure (dB) in metal sheet workers

2. **Incentives for employers/firms**
   - Legislation versus no legislation for reducing noise exposure (dB) in mines
   - Subsidies for employers vs no subsidies for reducing noise exposure (dB) in offshore oil platforms
   - Inspections and penalties for preventing noise exposure (dB) in construction industry
   - Occupational health advice to decrease noise exposure (dB) in small companies

3. **Drugs to prevent hearing loss in workers**
   - Magnesium vs placebo for preventing hearing loss (STS) in noise-exposed workers
Too heterogeneous for meta-analysis?

- Cochrane Systematic Review, Rehabilitation for older people in long-term care, CD004294

- Objective: to evaluate physical rehabilitation interventions directed at improving physical function among older people in long-term care.

- ...From these, 49 studies fulfilled the eligibility criteria and are included in this review.

- ...The included studies are heterogeneous. They examine different types of intervention, and evaluate them with a wide battery of outcome measures. Such variety made a meta-analysis unfeasible.

- Are the authors correct?
Solutions for heterogeneity 1: focus

1. Narrow down the scope of the review

2. When few studies expected, formulate on beforehand which comparisons will be judged sufficiently similar to be combined.
Need for intervention classification

• Preferably classification should...
  
    • be based on mechanism
    • have a practical meaning
    • have consequences in resource use
Criteria for intervention classification

- **Outcome**
  - that the intervention aims to reduce
  - exposure, worker behaviour, occupational disease, disability, injury
- **Mode of action**
  - environmental, behavioural, clinical
- **Level or point of action**
  - individual, group, societal level (legal)
- **Complexity**
  - simple, multi-component, multi-actor
- **Target Group**
  - workers, students, specific occupations
- **Place of delivery or setting**
  - hospital, primary care, workplace
- **Moment of application**
  - preventive (without request for help), treatment
- **Mode of delivery**
  - verbal, written, web-based, media
Use Excel Pivot Table
Excel Pivot Table
Pivot Table
Alternatives for dealing with heterogeneity?

• Because the studies were too heterogeneous to combine them in a meta-analysis we used a levels of evidence synthesis (best evidence synthesis) to combine them.

• **Data synthesis:** The selected studies were very heterogeneous in types of interventions, types of complaints, study population and outcomes measures, and therefore meta-analyses were not performed. Findings were reported narratively.

• **Levels of evidence:** For a more qualitative approach to synthesise the findings from included studies, so-called ‘levels of evidence’ were used (Ostelo 2002; Van Tulder 1997; Van Tulder 2001).

  - Levels of evidence:
    1. Good evidence - provided by generally consistent findings in two or more high-quality studies
    2. Moderate evidence - provided by generally consistent findings in one high-quality study and one or more low-quality studies, or by generally consistent findings in two or more low-quality studies
    3. Limited or conflicting evidence - only one study (either high or low quality), or inconsistent findings in two or more studies
    4. No evidence - no studies. (Henken 2007)

• Are the authors correct?
Too heterogeneous to combine thus..

• Worker training to prevent injuries

• Outcome: Reported Injuries
  • Peterson 2001 found after one year follow-up:
    • intervention: 15 injuries / 450 workers
    • control: 19 injuries / 370 workers
    • RR 0.61 (95% CI 0.3 to 1.2)
    • author's conclusions: non-significant outcome
  • Hansson 2004 found after one year follow-up:
    • intervention: 15 injuries / 402,000 working hours
    • control: 23 injuries / 386,000 working hours
    • RR 0.63 (95% CI 0.3 to 1.2)
    • author's conclusions: non-significant outcome

• No quantitative analysis possible we combined studies qualitatively:
  • conclusion: based on two studies with a non-significant outcome we found no evidence of effectiveness
Solution to heterogeneity 2: recalculate

- Recalculate all outcomes on similar scale
  - 2000 working hours = 1 working year (US)

- Combine in meta-analysis

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental</th>
<th>Control</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
<td>Events</td>
</tr>
<tr>
<td>Hansen 2004</td>
<td>14</td>
<td>450</td>
<td>19</td>
</tr>
<tr>
<td>Peterson 2001</td>
<td>16</td>
<td>200</td>
<td>23</td>
</tr>
</tbody>
</table>

Total (95% CI) 650 562 100.0% 0.62 [0.39, 0.97]

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 0.00$, df = 1 ($P = 0.94$); $I^2 = 0$
Test for overall effect: $Z = 2.07$ ($P = 0.04$)

- Review Conclusion:
  - the intervention reduces injuries with 38%
Solution to heterogeneity 3: narrative synthesis

• Studies used different outcomes and interventions and therefore we did not combine them but described them in a narrative way.

• How would you perform a narrative synthesis?
Narrative synthesis

• Rodgers et al 2009:
  • developing a theory of how the intervention works, why and for whom;
  • developing a preliminary synthesis;
  • exploring relationships within and between studies;
  • assessing the robustness of the synthesis product.
Narrative synthesis

Figure 1. Synthesis Process
Narrative synthesis

• Rodgers et al. Evaluation 2009 (15) 47-79

• Guidance-led narrative synthesis against a meta-analysis of the same study data.
  • The conclusions of the two syntheses were broadly similar.
  • However, conclusions about the
    • impact of moderators of effect appeared stronger when derived from the meta-analysis,
    • whereas implications for future research appeared more extensive when derived from the narrative synthesis.
Statistical heterogeneity
Statistical heterogeneity

A

$Q = 3.00$

$(Q - df) < 0$

$p = 0.70$

$T^2 = 0.00$

$T = 0.00$

$I^2 = 0.00\%$

---

$-1.0$  $0.0$  $1.0$

B

$Q = 12.00$

$(Q - df) = 7.00$

$p = 0.035$

$T^2 = 0.006$

$T = 0.077$

$I^2 = 58.34\%$

---

$-1.0$  $0.0$  $1.0$

C

$Q = 3.00$

$(Q - df) < 0$

$p = 0.70$

$T^2 = 0.00$

$T = 0.00$

$I^2 = 0.00\%$

---

$-1.0$  $0.0$  $1.0$

D

$Q = 12.00$

$(Q - df) = 7.00$

$p = 0.035$

$T^2 = 0.037$

$T = 0.193$

$I^2 = 58.34\%$

---

$-1.0$  $0.0$  $1.0$
Statistical heterogeneity

- Between-studies $\sigma^2$ is low because total $\sigma^2$ is low.
- Between-studies $\sigma^2$ is low because within-studies $\sigma^2$ is high.
- Between-studies $\sigma^2$ is high because total $\sigma^2$ is high. And within-studies $\sigma^2$ is low.
Zinc for Common Cold

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Intervention</th>
<th>Control</th>
<th>Mean Difference IV, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kurugol 2006a</td>
<td>4.7</td>
<td>0.8</td>
<td>97</td>
</tr>
<tr>
<td>Kurugol 2007</td>
<td>5.5</td>
<td>1.97</td>
<td>60</td>
</tr>
<tr>
<td>Macknin 1998</td>
<td>8.5</td>
<td>2.85</td>
<td>125</td>
</tr>
<tr>
<td>Petrus 1998</td>
<td>4.4</td>
<td>1.4</td>
<td>52</td>
</tr>
<tr>
<td>Prasad 2000</td>
<td>4.5</td>
<td>1.6</td>
<td>25</td>
</tr>
<tr>
<td>Prasad 2008</td>
<td>4</td>
<td>1.04</td>
<td>25</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>384</td>
<td>378</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Heterogeneity: $\chi^2 = 90.33$, df = 5 ($P < 0.00001$); $I^2 = 94$
Test for overall effect: $Z = 9.76$ ($P < 0.00001$)
## Statistical heterogeneity

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Blunted needles</th>
<th>Sharp needles</th>
<th>Risk Ratio</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
<td>Events</td>
<td>Total</td>
</tr>
<tr>
<td>Wright 1993</td>
<td>18</td>
<td>76</td>
<td>31</td>
<td>62</td>
</tr>
<tr>
<td>Thomas 1995</td>
<td>14</td>
<td>40</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Hartley 1996</td>
<td>3</td>
<td>46</td>
<td>14</td>
<td>39</td>
</tr>
<tr>
<td>Mingoli 1996</td>
<td>49</td>
<td>390</td>
<td>102</td>
<td>392</td>
</tr>
<tr>
<td>Rice 1996</td>
<td>0</td>
<td>36</td>
<td>10</td>
<td>64</td>
</tr>
<tr>
<td>Botet 1998</td>
<td>6</td>
<td>200</td>
<td>63</td>
<td>200</td>
</tr>
<tr>
<td>Ablett 1998</td>
<td>9</td>
<td>104</td>
<td>15</td>
<td>91</td>
</tr>
<tr>
<td>Nordkam 2005</td>
<td>12</td>
<td>100</td>
<td>28</td>
<td>100</td>
</tr>
<tr>
<td>Wilson 2008</td>
<td>4</td>
<td>217</td>
<td>5</td>
<td>221</td>
</tr>
<tr>
<td>Sullivan 2009</td>
<td>7</td>
<td>97</td>
<td>17</td>
<td>97</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>1306</td>
<td>1306</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>122</td>
<td>305</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.18; \chi^2 = 23.42, df = 9 (P = 0.005); I^2 = 62\%$

Test for overall effect: $Z = 5.00 (P < 0.00001)$
### Statistical heterogeneity

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Blunted needles</th>
<th>Sharp needles</th>
<th>Risk Ratio</th>
<th>M-H, Random, 95% CI</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
<td>Events</td>
<td>Total</td>
<td>Weight</td>
</tr>
<tr>
<td>Wright 1993</td>
<td>18</td>
<td>76</td>
<td>31</td>
<td>62</td>
<td>17.4%</td>
</tr>
<tr>
<td>Thomas 1995</td>
<td>14</td>
<td>40</td>
<td>20</td>
<td>40</td>
<td>14.2%</td>
</tr>
<tr>
<td>Mingoli 1996</td>
<td>49</td>
<td>390</td>
<td>102</td>
<td>392</td>
<td>40.5%</td>
</tr>
<tr>
<td>Hartley 1996</td>
<td>3</td>
<td>46</td>
<td>14</td>
<td>39</td>
<td>2.8%</td>
</tr>
<tr>
<td>Rice 1996</td>
<td>0</td>
<td>36</td>
<td>10</td>
<td>64</td>
<td>0.5%</td>
</tr>
<tr>
<td>Ablett 1998</td>
<td>9</td>
<td>104</td>
<td>15</td>
<td>91</td>
<td>6.5%</td>
</tr>
<tr>
<td>Nordkam 2005</td>
<td>12</td>
<td>100</td>
<td>28</td>
<td>100</td>
<td>10.3%</td>
</tr>
<tr>
<td>Wilson 2008</td>
<td>4</td>
<td>217</td>
<td>5</td>
<td>221</td>
<td>2.3%</td>
</tr>
<tr>
<td>Sullivan 2009</td>
<td>7</td>
<td>97</td>
<td>17</td>
<td>97</td>
<td>5.6%</td>
</tr>
<tr>
<td><strong>Total (95% CI)</strong></td>
<td>1106</td>
<td></td>
<td>1106</td>
<td></td>
<td>100.0%</td>
</tr>
<tr>
<td>Total events</td>
<td>116</td>
<td></td>
<td>242</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 7.24$, df = 8 ($P = 0.51$); $I^2 = 0$

Test for overall effect: $Z = 7.13$ ($P < 0.00001$)
Statistical heterogeneity
Meta-regression

- Linear regression model
  - Dependent variable:
    - effect size (SMD, ln OR)
  - Independent variables
    - any study characteristic (‘subgroup’)

- Tests for differences between ‘subgroups’

- Needs at least 10 studies

- Can be best performed in STATA
Flowchart

List of included studies

Check the conceptual similarity of the items 1 to 7 and in that order

1. Interventions / Exposure
2. Control condition
3. Participants
4. Study Design
5. Outcome
6. Follow-up time
7. Effect Size (RR, MD)

Transform ES if necessary

Perform Meta-Analysis

Check / Explain Remaining Statistical Heterogeneity

1. If a little dissimilar, consider making subgroups
2. If quite dissimilar consider narrative synthesis
3. If very dissimilar consider describing studies separately
4. Always report and pool different study designs separately