GRADE for overviews:
In systematic reviews we trust?

Joanne McKenzie¹ and Sue Brennan²

Acknowledgements
Co-investigators: Elie Akl, Philippa Middleton, Alex Pollock
Researchers: Jane Reid and Carole Lunny
Funding: Monash University Strategic Project Grant (SGS17-0704)

¹ School of Public Health and Preventive Medicine
² Cochrane Australia, School of Public Health and Preventive Medicine
Joanne McKenzie is funded by a National Health and Medical Research Council (NHMRC) career development fellowship and holds NHMRC project grant funding. Receives funding to undertake commissioned systematic reviews and methodological review from the NHMRC Health Evidence Panel. Contributes to the Melbourne GRADE centre and is a member of the GRADE working group. Is co-convenor of the Cochrane Statistical Methods Group.

Sue Brennan is employed by Cochrane Australia which is funded by the NHMRC. Receives funding to undertake commissioned systematic reviews and methodological review from the NHMRC Health Evidence Panel. Leads the Melbourne GRADE centre and is a member of the GRADE working group.
Why so much interest in overviews?
Accessible: front end for decision makers to access findings of multiple, related reviews

68 Cochrane reviews on fertility treatments

**Abstract**

**Background**

As many as one in six couples will encounter problems with fertility, defined as failure to achieve a clinical pregnancy after regular intercourse for 12 months. Increasingly, couples are turning to assisted reproductive technology (ART) for help with conceiving and ultimately giving birth to a healthy live baby of their own. Fertility treatments are complex, and each ART cycle consists of several steps. If one of these steps is incorrectly applied, the stakes are high as conception may not occur. With this in mind, it is important that each step of the ART cycle is supported by good evidence from well-designed studies.

**Objectives**

To summarise the evidence from Cochrane systematic reviews on procedures and treatment options available to couples with RTI procedures.
Health advice

We provide Australians with the best available evidence-based advice about improving health and preventing disease.

We develop specific guidelines on various aspects of health, health care, health research and environmental health. We also review the latest research to provide information and advice about specific issues that might concern Australian communities.
The Review of the Australian Government Rebate on Private Health Insurance for Natural Therapies

To synthesis evidence about the effects of natural therapies, overview methods were used to aggregate findings from
• 348 systematic reviews (675 component trials, >75,000 participants)
• 68 conditions
Why overviews?

**Efficiency:** expedite the review of large and complex bodies of evidence

**Reduce waste:** capitalise on existing reviews rather than duplicating (41,563 SRs on PROSPERO)
Unique challenges arise in synthesising review evidence, for which guidance is yet to be agreed.

These challenges have ramifications for assessing certainty of the evidence in an overview using GRADE or other approaches.

For example, authors must decide:

• whether they will consider limitations of the included SRs in their assessment and, if so, how,

• how they will deal with overlapping data (e.g. where the same primary studies contribute to meta-analyses in multiple SRs), and

• how they will deal with missing or discordant data needed to assess certainty (e.g. conflicting risk of bias (RoB) assessments).
This project

**Objective:** to inform methods guidance by examining *current practice* and *expert perspectives* on assessing the certainty of the evidence in overviews.
Current practice: systematic review of methods used to assess certainty of the evidence

Medline search for overviews
(01 Jan 2015 – 11 Mar 2017)

Randomly sorted citations

Applied eligibility criteria to select up to 50 overviews that reported on certainty of the evidence
• GRADE (RoB, indirectness, inconsistency, imprecision, publication bias, +/-)
• Other tool or method (e.g. EPC strength of evidence)
• A mix (as used in the SR)
• Other explicit criteria
1707 citations retrieved and screened by 2 authors

381 full text articles screened
- Excluded: 250 – not an overview
- 90 of 131 overviews did not assess certainty (formally)

41 overviews met all eligibility criteria and were included

What tool was used?
- GRADE methods most common (19 overviews)
- 15 reported another tool or method, 1 a mix, 6 used own criteria

Who did the assessment?
- Most overview authors did their own assessment (35 overviews)
- 3 extracted assessments from included SRs
- 3 were unable to complete the assessment due to missing data for GRADE

Was anything done differently from an assessment of primary studies in a SR?
- No - GRADE was applied using the same approach as in a SR of primary studies
- Concerns about SR process were not integrated into the assessment
Expert perspectives: interviews and working meeting with methodologists (n=23)

Participants (at interview and meeting) were asked to:

• consider how they would GRADE the evidence presented in each of four scenarios commonly encountered by overview authors

• describe the process and reasoning behind their GRADE assessment

Working meeting: asked participants to critique key considerations and approaches identified from interviews
### Scenario 1

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Counselling Events</th>
<th>Counselling Total</th>
<th>Usual care Events</th>
<th>Usual care Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook 1995</td>
<td>8</td>
<td>23</td>
<td>2</td>
<td>20</td>
<td>2.8%</td>
<td>3.48 [0.83, 14.52]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Dornelas 2006</td>
<td>15</td>
<td>53</td>
<td>5</td>
<td>52</td>
<td>5.7%</td>
<td>2.94 [1.15, 7.51]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Gielen 1997</td>
<td>12</td>
<td>193</td>
<td>11</td>
<td>198</td>
<td>7.2%</td>
<td>1.12 [0.51, 2.48]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Hajek 2001</td>
<td>80</td>
<td>365</td>
<td>73</td>
<td>367</td>
<td>18.1%</td>
<td>1.10 [0.83, 1.46]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Kendrick 1995</td>
<td>48</td>
<td>822</td>
<td>65</td>
<td>1063</td>
<td>15.9%</td>
<td>0.95 [0.67, 1.37]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>McLeod 2004</td>
<td>37</td>
<td>163</td>
<td>14</td>
<td>109</td>
<td>10.9%</td>
<td>1.77 [1.00, 3.11]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Stotts 2002</td>
<td>27</td>
<td>134</td>
<td>28</td>
<td>135</td>
<td>13.0%</td>
<td>0.97 [0.61, 1.56]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Valbo 1996</td>
<td>5</td>
<td>52</td>
<td>8</td>
<td>78</td>
<td>4.7%</td>
<td>0.94 [0.32, 2.71]</td>
<td>? ? + + + + +</td>
</tr>
<tr>
<td>Windsor 1993</td>
<td>57</td>
<td>400</td>
<td>35</td>
<td>414</td>
<td>14.9%</td>
<td>1.69 [1.13, 2.51]</td>
<td>? ? + + + + +</td>
</tr>
</tbody>
</table>

Total (95% CI) 2532 2756 100.0% 1.39 [1.07, 1.79]

Total events 312 248

Heterogeneity: Tau² = 0.07; Chi² = 18.15, df = 9 (P = 0.03); I² = 50%

Test for overall effect: Z = 2.50 (P = 0.01)
Scenario 1

Overview: P. Any population
I. Smoking cessation counselling

SR1 2010
10 trials

ROBIS (high risk 2 domains)
Concerns about methods used to select studies, collect data and appraise studies (one author screened full text, extracted data, assessed RoB)
AMSTAR: 10/11

SR1
P. Women during pregnancy, any age
Overview: P. Any population
I. Smoking cessation counselling

SR1
P. Women during pregnancy, any age
4 trials
ROBIS (high risk 2 domains)
Concerns about methods used to select studies, collect data and appraise studies (one author screened full text, extracted data, assessed RoB)
AMSTAR: 10/11

SR2
P. Women during pregnancy, ≥18 yr
6 trials
ROBIS (high risk 2 domains)
Concerns about specification of eligibility criteria and synthesis (no a priori design; no assessment of likelihood of reporting bias)
AMSTAR: 9/11

SR1 2010
4 trials

SR2 2015
3 trials

Overview: P. Any population
I. Smoking cessation counselling

Scenario 2
## Scenario 2

### Table 1: Study or Subgroup Information

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Counselling Events</th>
<th>Total</th>
<th>Usual care Events</th>
<th>Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook 1995</td>
<td>8</td>
<td>23</td>
<td>2</td>
<td>20</td>
<td>2.8%</td>
<td>3.48 [0.83, 14.52]</td>
<td></td>
</tr>
<tr>
<td>Domelas 2006</td>
<td>15</td>
<td>53</td>
<td>5</td>
<td>52</td>
<td>5.7%</td>
<td>2.94 [1.15, 7.51]</td>
<td></td>
</tr>
<tr>
<td>Gielin 1997</td>
<td>12</td>
<td>193</td>
<td>11</td>
<td>198</td>
<td>7.2%</td>
<td>1.12 [0.51, 2.48]</td>
<td></td>
</tr>
<tr>
<td>Hajek 2001</td>
<td>80</td>
<td>365</td>
<td>73</td>
<td>367</td>
<td>18.1%</td>
<td>1.10 [0.83, 1.46]</td>
<td></td>
</tr>
<tr>
<td>Hegaard 2003</td>
<td>23</td>
<td>327</td>
<td>7</td>
<td>320</td>
<td>6.7%</td>
<td>3.22 [1.40, 7.39]</td>
<td></td>
</tr>
<tr>
<td>Kendrick 1995</td>
<td>48</td>
<td>822</td>
<td>65</td>
<td>1063</td>
<td>15.9%</td>
<td>0.95 [0.67, 1.37]</td>
<td></td>
</tr>
<tr>
<td>McLeod 2004</td>
<td>37</td>
<td>163</td>
<td>14</td>
<td>109</td>
<td>10.9%</td>
<td>1.77 [1.00, 3.11]</td>
<td></td>
</tr>
<tr>
<td>Stotts 2002</td>
<td>27</td>
<td>134</td>
<td>28</td>
<td>135</td>
<td>13.0%</td>
<td>0.97 [0.61, 1.56]</td>
<td></td>
</tr>
<tr>
<td>Valbo 1996</td>
<td>5</td>
<td>52</td>
<td>8</td>
<td>78</td>
<td>4.7%</td>
<td>0.94 [0.32, 2.71]</td>
<td></td>
</tr>
<tr>
<td>Windsor 1993</td>
<td>57</td>
<td>400</td>
<td>35</td>
<td>414</td>
<td>14.9%</td>
<td>1.69 [1.13, 2.51]</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>2532</td>
<td>2756</td>
<td>100.0%</td>
<td>1.39 [1.07, 1.79]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Study or Subgroup Information

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Counselling Events</th>
<th>Total</th>
<th>Usual care Events</th>
<th>Total</th>
<th>Weight</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domelas 2006</td>
<td>15</td>
<td>53</td>
<td>5</td>
<td>52</td>
<td>3.7%</td>
<td>2.94 [1.15, 7.51]</td>
<td></td>
</tr>
<tr>
<td>Gielin 1997</td>
<td>12</td>
<td>193</td>
<td>11</td>
<td>198</td>
<td>4.9%</td>
<td>1.12 [0.51, 2.49]</td>
<td></td>
</tr>
<tr>
<td>Hajek 2001</td>
<td>80</td>
<td>365</td>
<td>73</td>
<td>367</td>
<td>20.0%</td>
<td>1.10 [0.83, 1.48]</td>
<td></td>
</tr>
<tr>
<td>Kendrick 1995</td>
<td>48</td>
<td>822</td>
<td>65</td>
<td>1063</td>
<td>15.6%</td>
<td>0.95 [0.67, 1.37]</td>
<td></td>
</tr>
<tr>
<td>Lee 2015</td>
<td>21</td>
<td>140</td>
<td>16</td>
<td>137</td>
<td>7.7%</td>
<td>1.28 [0.70, 2.35]</td>
<td></td>
</tr>
<tr>
<td>McLeod 2004</td>
<td>37</td>
<td>163</td>
<td>14</td>
<td>109</td>
<td>8.6%</td>
<td>1.77 [1.00, 3.11]</td>
<td></td>
</tr>
<tr>
<td>Moore 2002</td>
<td>88</td>
<td>523</td>
<td>108</td>
<td>567</td>
<td>21.7%</td>
<td>0.88 [0.68, 1.14]</td>
<td></td>
</tr>
<tr>
<td>Stotts 2002</td>
<td>27</td>
<td>134</td>
<td>28</td>
<td>135</td>
<td>11.2%</td>
<td>0.97 [0.61, 1.56]</td>
<td></td>
</tr>
<tr>
<td>Wilkinson 2012</td>
<td>9</td>
<td>29</td>
<td>14</td>
<td>31</td>
<td>6.6%</td>
<td>0.89 [0.35, 1.34]</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>2422</td>
<td>2659</td>
<td>100.0%</td>
<td>1.07 [0.89, 1.30]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Diagram

- **Heterogeneity:** 
  - Tau²: 0.07
  - Chi²: 18.15
  - df: 9
  - P = 0.03
  - I² = 50%

- **Test for overall effect:** Z = 2.50 (P = 0.01)
Scenario 3

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Counselling Events</th>
<th>Counselling Total</th>
<th>Usual care Events</th>
<th>Usual care Total</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk Ratio M-H, Random, 95% CI</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cook 1995</td>
<td>8</td>
<td>23</td>
<td>2</td>
<td>20</td>
<td>3.48 [0.83, 14.52]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Domelas 2006</td>
<td>15</td>
<td>53</td>
<td>5</td>
<td>52</td>
<td>2.94 [1.15, 7.51]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Gielin 1997</td>
<td>12</td>
<td>193</td>
<td>11</td>
<td>198</td>
<td>1.12 [0.51, 2.48]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Hajek 2001</td>
<td>80</td>
<td>365</td>
<td>73</td>
<td>367</td>
<td>1.10 [0.83, 1.46]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Kendrick 1995</td>
<td>48</td>
<td>822</td>
<td>65</td>
<td>1063</td>
<td>0.95 [0.67, 1.37]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Lee 2015</td>
<td>21</td>
<td>140</td>
<td>16</td>
<td>137</td>
<td>1.28 [0.70, 2.35]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Moore 2002</td>
<td>88</td>
<td>523</td>
<td>108</td>
<td>567</td>
<td>0.88 [0.68, 1.14]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Stotts 2002</td>
<td>27</td>
<td>134</td>
<td>28</td>
<td>135</td>
<td>0.97 [0.61, 1.56]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Valbo 1996</td>
<td>5</td>
<td>52</td>
<td>8</td>
<td>78</td>
<td>0.94 [0.32, 2.71]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
<tr>
<td>Wilkinson 2012</td>
<td>9</td>
<td>29</td>
<td>14</td>
<td>31</td>
<td>0.69 [0.35, 1.34]</td>
<td></td>
<td>? ? ? + + +</td>
</tr>
</tbody>
</table>

Total (95% CI) 3224 3491 100.0% 1.24 [1.00, 1.54]

Total events 430 386

Heterogeneity: Tau² = 0.07; Chi² = 26.17, df = 12 (P = 0.01); I² = 54%
Test for overall effect: Z = 1.92 (P = 0.05)
Results

• Findings of key considerations, methods / approaches are presented
Aspects to assessing the certainty of the evidence: interview participants

Assessing the certainty of the evidence in the primary studies

Systematic review process

GRADEing process

Three aspects need to be considered in the overall assessment of the certainty of the evidence
Options proposed for integrating concerns about the SR process

**Option 1:** separate GRADE assessment of the primary studies and of the SR process

Assessing the certainty of the evidence in the primary studies:
- **Indirectness**
- **Risk of bias**
- **Imprecision**
- **Inconsistency**
- **Publication bias**

<table>
<thead>
<tr>
<th></th>
<th>HIGH</th>
<th>MODERATE</th>
<th>LOW</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirectness</td>
<td>★★★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Risk of bias</td>
<td>★★★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Imprecision</td>
<td>★★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>Publication bias</td>
<td>★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
</tbody>
</table>

Systematic review process:
- **? domain**

<table>
<thead>
<tr>
<th></th>
<th>HIGH</th>
<th>MODERATE</th>
<th>LOW</th>
<th>VERY LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>? domain</td>
<td>★★★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>? domain</td>
<td>★★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>? domain</td>
<td>★★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
<tr>
<td>? domain</td>
<td>★</td>
<td>★★★</td>
<td>★★</td>
<td>★</td>
</tr>
</tbody>
</table>
Options proposed for integrating concerns about the SR process

**Option 2:** add new domain(s) to GRADE to assess the SR process
Options proposed for integrating concerns about the SR process

**Option 3:** map concerns in the SR process to current GRADE domains

“In terms of review indirectness at SR level it is ok, but may be a problem at the overview level.”

- Indirectness
- Risk of bias
- Imprecision
- Inconsistency
- Publication bias

HIGH: ★★★★★
MODERATE: ★★★★★
LOW: ★★★★★
VERY LOW: ★★★★★
Options proposed for integrating concerns about the SR process

**Option 3:** map concerns in the SR process to current GRADE domains

“To me it seems kind of intuitively sensible to try to come up with some kind of combined assessment based on RoB across the two if possible. If high RoB in two ROBIS domains, plus potential moderate RoB within the included studies, might say overall concerned about high RoB, in terms of the overall assessment of RoB.”
Options proposed for integrating concerns about the SR process

Option 3: map concerns in the SR process to current GRADE domains

“If search is not comprehensive maybe fits onto indirectness. Or publication bias, which we always say look at [the] search strategy.”
Options proposed for integrating concerns about the SR process: meeting participants’ views

Option 1: separate GRADE assessment of the primary studies and of the SR process
   →? Participants’ felt that as a minimum an assessment of the SR process should be reported in the SoFs table

Option 2: add new domain(s) to GRADE to assess the SR process
   →? Some participants’ felt that uncertainty arising from the SR process was an artefact and should not be integrated with the assessment of primary studies

Option 3: map concerns in the SR process to current GRADE domains
   → Rejected. Participants’ felt that many concerns about the SR process do not map well to current GRADE domains
Concerns in the SR process: what were these?

From the interviews participants, variability in opinion about which elements of the SR process may introduce bias.
"One author extracted data"

Less concerned

"... it bothers me but it doesn't bother me that much, especially if the single person has expertise in an area ..."

"... not sure if the best way is for two people [to] do ... data extraction and so on, as I think it also heavily depends on the experience, so even though it is a gold standard, then I’m not sure ...

More concerned

"... Data extraction and risk of bias assessment are the most worrying factors described here..."
## SR process: elements of concern

- No protocol or *a priori* specification of inclusion criteria
- Inclusion criteria of SRs strangely specified
- Non-comprehensive search
- Omitted studies that should have been included
- Lack of expertise of the SR author team / do not know the team
- Conflict of interest of the SR author team
- Lack of detailed reporting of primary study characteristics within SR
- Date SR published (may be more lenient on older SRs)
- Use of outdated SR methods (e.g. Jadad quality scale)
- No RoB undertaken
- Inappropriate analysis (e.g. fixed effect model when there is statistical heterogeneity)
- High risk of bias on any of the ROBIS domains
Additional challenges with results from multiple SRs

How to deal with different RoB / quality scales across SRs? Should we treat trials where quality scales have been reported as ‘missing assessments’ and downgrade?

Does having knowledge that there were missing studies lead to upgrading (now have some of the missing studies), or downgrading (evidence that authors missed studies, so there may be others)?

Do differences in the data reported and RoB assessments for the same studies across the SRs lead to upgrading (if no differences) or downgrading (if differences)?
Conclusions: assessing the certainty of evidence

Assessing the certainty of evidence in overviews

• Limitations of the systematic review process need to be considered in assessing the certainty of evidence

• Different options for how this could be done
  • Separate GRADE assessment of the primary studies and of the SR process
  • Add new domain(s) to GRADE to assess the SR process
  • Map concerns in the SR process to current GRADE domains

• Diverse views on the elements of the SR process that are most important to consider when assessing certainty

• Next step is to develop principles for applying GRADE in overviews with project group members
Conclusions: broader issues

• Meta-epidemiological evidence has informed which elements of trial conduct may lead to biased intervention estimates
  → Cochrane RoB tool

• More limited empirical evidence to guide which elements of the SR process may lead to biased intervention estimates
Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records

Phil Edwards, Mike Clarke, Carolyn DiGuiseppi, Sarah Pratap, Ian Roberts and Reinhard Wentz

Testing the Risk of Bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs

Lisa Hanting, Michele P. Hamn, Andrea Milne, Ben Vandermeer, P. Lina Santaguida, Mohammed Ansari, Alexander Tsirtsvadze, Susanne Hempel, Paul Shekelle and Donna M. Dryden

Dual computer monitors to increase efficiency of conducting systematic reviews

Zhen Wang, Noor Asi, Tarig A. Elraiyah, Abd Mooin Abu Dahrh, Chaitanya Undavalli, Paul Glasziou, Victor Montori, Mohammad Hassan Murad

*Knowledge & Evaluation Research Unit, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA
*Centre for Research in Evidence-Based Practice, Faculty of Health Sciences and Medicine, Bond University, Queensland 4229, Australia
Accepted 1 June 2014; Published online 30 July 2014

Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews

Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study

Betta Tendal, PhD student; Julian P T Higgins, senior statistician; Peter Jürg, head of division; Asbjørn Hoblarbo, senior researcher; Sven Trelle, associate director; Eveline Niessen, PhD student; Simon Wandel, PhD student; Anders W Jørgensen, PhD student; Katarina Gesser, PhD student; Saren Iten-Kristensen, PhD student; Peter C Gøtzsche, director
Conclusions: broader issues (continued)

• Meta-epidemiological evidence has informed which elements of trial conduct may lead to biased intervention estimates
  → Cochrane RoB tool

• More limited empirical evidence to guide which elements of the SR process may lead to biased intervention estimates

• Having more empirical evidence of the key elements would provide a better foundation for knowing which SRs we should trust more

• Systematic collation of these empirical evaluations, with continued updating (particularly in the assessment of automation technologies), is required
Acknowledgements

We thank the following participants for their contributions to interviews and the working meeting: Elie Akl, Ed Aromataris, Madeleine Ballard, Stephanie Chang, Saskia Cheyne, Vittal Katikireddi, Reinar Liv Merete, Steve McDonald, Philippa Middleton, Rintaro Mori, Zachary Munn, Reem Mustafa, Liesl Nicol, Dawid Pieper, Karen Pilkington, Rebecca Ryan, Nancy Santesso, Ein-soon Shin, Airton Stein, Nandi Siegfried, Hilary Thomson, Robert Vander Stichele, Penny Whiting