Outline

9.10 – MECIR Standards and the CEU Screening Tool
9.30 – Screening a Protocol
10.10 – GRADE and SoF tables
10.45 – Coffee
11.15 – GRADE and SoF tables (cont’d)
12.25 – Checking consistency throughout the review
12.45 – Lunch
13.45 – Screening a Review
14.35 – Plain Language Summary
Review of MECIR Standards:

CEU Screening Programme

MECIR Standards

Methodological Expectations of Cochrane Intervention Reviews (MECIR)

• Methodological standards for the **conduct** of Cochrane Intervention Reviews (Version 2.3, 02 December 2013)

• Methodological standards for the **reporting** of Cochrane Intervention Reviews (version 1.1, 17 December 2012)
Methodological Expectations of Cochrane Intervention Reviews (MECIR)

Methodological standards for the conduct of new Cochrane Intervention Reviews

Version 2.3, 02 December 2013

Jackie Chandler, Rachel Churchill, Julian Higgins, Toby Lasserson and David Tovey

Preface

Cochrane Reviews are seen as exemplifying best practice in the quality of both their conduct and reporting. To maintain this position we need to improve and maintain the quality of our output as standards and expectations for systematic reviews increase generally; we also need to ensure consistency across all Cochrane Review Groups (CRGs) and all reviews. To this end we have undertaken within The Cochrane Collaboration to define Methodological Expectations for Cochrane Intervention Reviews (MECIR).

The documents associated with the MECIR project form a major step forward aimed at ensuring that both researchers and editorial teams have a shared understanding of the expectations of conduct and reporting for reviews in the Cochrane Database of Systematic Reviews (CDSR).

The standards below summarize attributes of the conduct of reviews of interventions described in the Cochrane Handbook that we have established should be either mandatory or highly desirable for new Cochrane Reviews. The judgments are accompanied by a rationale and reference to the appropriate section of the Cochrane Handbook.

We have described the process for determining the expectations for conducting Cochrane Reviews of interventions, including the methods used to develop the initial list and the management of all feedback received during the consultation process (see: www.editorial-unit.cochrane.org/mecir).

Finally, I want to pay tribute to my colleagues who have contributed to this work so far. Julian Higgins and Rachel Churchill have led this initiative with great expertise, perseverance and energy. An important feature of this project, at all levels, has been to reflect the importance of CRG teams and methodologists working alongside one another. Rachel and Julian have been supported by Jackie Chandler and Toby Lasserson, both of whom have made major contributions. In addition, scores of people from within the Collaboration either contributed to the working groups, without which we would have had no ‘long-list’ of proposed expectations to build on, or the consultation that succeeded the working groups. I hope that the Collaboration recognises the efforts of all the individuals involved and the true sense of collaboration that the work has engendered.

David Tovey, Editor in Chief of The Cochrane Library
Methodological Expectations of Cochrane Intervention Reviews (MECIR)

Standards for the reporting of new Cochrane Intervention Reviews

Version 1.1 17 December 2012

Preface

The standards below summarize proposed attributes of reporting that we consider should be either mandatory or highly desirable for Cochrane Intervention Reviews, with the rationale for this judgment. These standards are not intended to apply to protocols or updated reviews at this point, and these will be addressed in further work. There is also a separate project ongoing aimed at clarifying expectations for plain language summaries.

In order to provide the user with a succinct and relevant document, the methodology of a review should be reported in such a way that links the methods directly to the results of the present version of the review. Thus, details of methods that were planned in the protocol but were not implemented should generally be reported in the dedicated section for differences between the protocol and the review, or in an appendix.

The Cochrane Collaboration has adopted recommendations provided in the PRISMA statement [http://www.prisma-statement.org]. We believe the reporting standards below will ensure compliance with these recommendations. Some items have been included specifically to enable this (e.g. the standard relating to mentioning that the review has a published protocol). Extensions to the PRISMA statement may also be relevant to particular reviews, such as reviews addressing equity issues [http://equity.cochrane.org/equity-extension-prisma].

The ordering of the standards reflects the position in which each issue might be expected to be addressed in the main text of the review. In some items we have specified where things should be reported (e.g. for contents of the table of Characteristics of included studies). For other items, review authors should consider whether information should be reported in the main text, in tables, figures or appendices.

Further details of the MECIR project can be found at our website: www.editorial-unit.cochrane.org/mecir

David Tovey, Editor in Chief of The Cochrane Library
CEU Screening

- All of the MECIR standards are important at different stages of the review process.

- CEU has been screening new reviews against key MECIR standards since September 2013.

- Evolved to consider 3 core components of reviews as major determinants of overall review quality:
  1. implementation of protocol methods,
  2. interpretation of findings
  3. consistency of reporting
### Implementation of protocol methods

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>C27</td>
<td>Searching trials registers</td>
<td>Research trials registers and repositories of results, where relevant to the topic through ClinicalTrials.gov, the WHO International Clinical Trials Registry Platform (ICTRP) portal and other sources as appropriate.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>C37</td>
<td>Reporting</td>
<td>C76 Assessing the quality of the body of evidence</td>
<td>Use the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of the body of evidence for each outcome, and to draw conclusions about the quality of evidence within the text of the review.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>C40</td>
<td>Declaring</td>
<td>R97 ‘Summary of findings’ table</td>
<td>Prepare a summary of findings table according to recommendations described in Chapter 11 of the Cochrane Handbook (version 6 or later). (Specifically):  - Include results for one clearly defined population group (with few exceptions); - Indicate the intervention and the comparison intervention; - Include seven or fewer patient-important outcomes; - Describe the outcomes in a single, concise, follow-up.</td>
<td></td>
</tr>
</tbody>
</table>

### Completeness of reporting in the abstract & internal consistency

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R11</td>
<td>Abstract, Main results: bias assessment</td>
<td>Provide a comment on the findings of the bias assessment.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R12</td>
<td>Abstract, Main results: findings</td>
<td>Report findings for all primary outcomes, irrespective of the strength and direction of the result, and of the availability of data.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R13</td>
<td>Abstract, Main results: adverse effects</td>
<td>Ensure that any findings related to adverse effects are reported. If adverse effects data were sought, but availability of data was limited, this should be reported.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R18</td>
<td>Consistency of summary versions of the review</td>
<td>Ensure that reporting of objectives, important outcomes, results, caveats and conclusions is consistent across the text, the abstract, the plain language summary and the ‘Summary of findings’ table (if included).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>R86</td>
<td>Consistency of results</td>
<td>Ensure that all statistical results presented in the main review text are consistent between the text and the ‘Data and analyses’ tables.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stroke CRG Common Errors
Implementation of protocol methods
<table>
<thead>
<tr>
<th>Item No.</th>
<th>Item name</th>
<th>Standard</th>
<th>Met?</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Implementation of protocol methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C27</td>
<td>Searching trials registers</td>
<td>Search trials registers and repositories of results, where relevant to the topic through ClinicalTrials.gov, the WHO International Clinical Trials Registry Platform (ICTRP) portal and other sources as appropriate.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C37</td>
<td>Repeating searches</td>
<td>Rerun or update searches for all relevant databases within 12 months before publication of the review or review update, and screen the results for potentially eligible studies.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C40</td>
<td>Excluding studies without usable data</td>
<td>Include studies in the review irrespective of whether measured outcome data are reported in a ‘usable’ way.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C68</td>
<td>Comparing subgroups</td>
<td>If subgroup analyses are to be compared, and there are judged to be sufficient studies to do this meaningfully, use a formal statistical test to compare them.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R106</td>
<td>Changes from the protocol</td>
<td>Explain and justify any changes from the protocol (including any post-hoc decisions about eligibility criteria or the addition of subgroup analyses).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C27: Searching trials registers
C37: Rerunning searches

- Trial Registers
  - ClinicalTrials.gov
  - WHO International Clinical Trials Registry Platform (ICTRP)

- Search within 12 months of publication
C40: Excluding studies without useable data

Include studies in the review irrespective of whether measured outcome data are reported in a ‘usable’ way.
C68: Comparing Subgroups

- use a formal statistical test to compare subgroups

**Abstract**: Our Review suggests that (INTERVENTION) may have more beneficial effects in (SUBGROUP)

**PLS**: In the further analyses, there is evidence indicated that the effects of (INTERVENTION) in reducing (OUTCOME) rate may be different between (SUBGROUP 1) and (SUBGROUP 2), with more benefits observed in (SUBGROUP 1)
FLR (Funny Looking Results)

- Be on the look out for FLRs (Funny Looking Results)

1. Study weight at odd with sample size
2. Outliers
3. Study ID appearing more than once in a forest plot
4. Reporting at odds with forest plot
5. Relative effect X control group risk (Abstract/SoF Table)
FLR #1 – Study weight at odd with sample size
FLR #2 – Outliers

- SEMs used instead of SDs
- Minus sign left off mean
FLR #3 – Study ID appearing >1 in a forest plot

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Treatment Mean</th>
<th>SD</th>
<th>Total (SE)</th>
<th>Control Mean</th>
<th>SD</th>
<th>Total (SE)</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.2.1 Timepoint 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1</td>
<td>20.53</td>
<td>17</td>
<td>20.53</td>
<td>30.20</td>
<td>17</td>
<td>30.20</td>
<td>-9.64 [-12.28, -7.00]</td>
</tr>
<tr>
<td>Study 2</td>
<td>20.48</td>
<td>17</td>
<td>20.48</td>
<td>30.20</td>
<td>17</td>
<td>30.20</td>
<td>-9.64 [-12.28, -7.00]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-9.64 [-12.28, -7.00]</td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.2 Timepoint 2** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.3 Timepoint 3** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.4 Timepoint 4** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.5 Timepoint 5** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.6 Timepoint 6** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.7 Timepoint 7** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.8 Timepoint 8** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)

| **1.2.9 Timepoint 9** |                |    |            |              |    |            |                                        |
| Study 1             | 20.53          | 17 | 20.53      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Study 2             | 20.48          | 17 | 20.48      | 30.20        | 17 | 30.20      | -9.64 [-12.28, -7.00]                  |
| Total (95% CI)      |                |    |            |              |    |            | -9.64 [-12.28, -7.00]                  |

Heterogeneity: Tau² = 0.00; Chi² = 1.00, df = 2 (P = 0.57), I² = 0%
Test for overall effect: Z = 0.00 (P = 0.998)
Test for subgroup differences: Chi² = 13.81, df = 8 (P = 0.08), I² = 41.5%

Favours experimental Favours control
<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Treatment</th>
<th>Control</th>
<th>Std. Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
</tr>
<tr>
<td>Study 1</td>
<td>30</td>
<td>15.5</td>
<td>17</td>
</tr>
<tr>
<td>Study 1</td>
<td>20</td>
<td>15.45</td>
<td>17</td>
</tr>
<tr>
<td>Study 2</td>
<td>4.5</td>
<td>2.48</td>
<td>13</td>
</tr>
<tr>
<td>Study 2</td>
<td>4</td>
<td>2.97</td>
<td>13</td>
</tr>
<tr>
<td>Study 3</td>
<td>2.89</td>
<td>1.81</td>
<td>57</td>
</tr>
<tr>
<td>Study 3</td>
<td>2.66</td>
<td>1.47</td>
<td>57</td>
</tr>
<tr>
<td>Study 4</td>
<td>2.89</td>
<td>1.59</td>
<td>88</td>
</tr>
<tr>
<td>Study 4</td>
<td>2.88</td>
<td>1.72</td>
<td>88</td>
</tr>
<tr>
<td>Study 5</td>
<td>43.3</td>
<td>19.7</td>
<td>17</td>
</tr>
<tr>
<td>Study 6</td>
<td>3.7</td>
<td>2.4</td>
<td>64</td>
</tr>
<tr>
<td>Study 7</td>
<td>1.7</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Study 8</td>
<td>11</td>
<td>9.3</td>
<td>8</td>
</tr>
<tr>
<td>Study 8</td>
<td>51.6</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>Study 8</td>
<td>53.6</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

Total (95% CI) 472 471 100.0% -0.32 [-0.48, -0.17]

Heterogeneity: Tau² = 0.02; Chi² = 16.17, df = 13 (P = 0.24); I² = 20%
Test for overall effect: Z = 4.10 (P < 0.0001)
FLR #4 – Reporting at odds with forest plot

‘The confidence intervals for the estimated HR include large benefit and moderate harm of intervention (0.88; 95% CI 0.64 to 1.12), $P = 0.43$’
**FLR #5 – Relative effect X control group risk (Abstract/SoF Table)**

<table>
<thead>
<tr>
<th>Study population</th>
<th>See comment</th>
<th>101 (4 studies)</th>
<th>Risks were calculated from pooled risk differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(RD): RD 0.04, 95% CI -0.05 to 0.12; $\chi^2 = 0%$</td>
</tr>
<tr>
<td><strong>Safety/Acceptability</strong></td>
<td><strong>Drop-outs and adverse events</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 per 1000</td>
<td>-2147483648 per 1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>(-2147483648 to -2147483648)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 per 1000</td>
<td>-2147483648 per 1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-2147483648 to -2147483648)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
R106: Changes from the protocol

- Explain and justify any changes from the protocol
Original Protocol

Key considerations here are:

a) Have the protocol methods been implemented appropriately.

b) Were the original methods appropriate to begin with?
Full Group Exercise

As a group, we will now screen together a recently submitted Protocol:

Dual task training for improving balance and gait in patients with stroke
Interpretation of findings
<table>
<thead>
<tr>
<th>Page 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interpretation</strong></td>
</tr>
<tr>
<td><strong>Assessing the quality of the body of evidence</strong></td>
</tr>
<tr>
<td>Use the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of the body of evidence for each outcome, and to draw conclusions about the quality of evidence within the text of the review.</td>
</tr>
<tr>
<td><strong>Summary of findings table</strong></td>
</tr>
<tr>
<td>Present a 'Summary of Findings' table according to recommendations described in Chapter 11 of the Cochrane Handbook (version 5 or later). Specifically: include results for one clearly defined population group (with few exceptions); indicate the intervention and the comparison intervention;</td>
</tr>
<tr>
<td><strong>Interpreting results</strong></td>
</tr>
<tr>
<td>Interpret a statistically non-significant P value (e.g. larger than 0.05) as a finding of uncertainty unless confidence intervals are sufficiently narrow to rule out an important magnitude of effect.</td>
</tr>
<tr>
<td><strong>Assessing the quality of the body of evidence</strong></td>
</tr>
<tr>
<td>Use the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of the body of evidence for each outcome, and to draw conclusions about the quality of evidence within the text of the review.</td>
</tr>
<tr>
<td><strong>Formulating implications for practice</strong></td>
</tr>
<tr>
<td>Base conclusions only on findings from the synthesis (quantitative or narrative) of studies included in the review.</td>
</tr>
<tr>
<td><strong>Implications for practice</strong></td>
</tr>
<tr>
<td>Provide a general interpretation of the evidence so that it can inform healthcare or policy decisions. Avoid making recommendations for practice.</td>
</tr>
</tbody>
</table>

**Completeness of reporting in the abstract & Internal consistency**
C73 Interpreting Results

- Ensure statistically non-significant P value (e.g. larger than 0.05) are presented as a finding of uncertainty.
  - Authors commonly mistake a lack of evidence of effect as evidence of a lack of effect
on the border of significance (p=0.063)
on the borderline of significance (p=0.0699)
on the borderlines of significance (p=0.08)
on the boundaries of significance (p=0.056)
on the boundary of significance (p=0.055)
on the brink of significance (p=0.052)
on the cusp of conventional statistical significance (p=0.054)
on the cusp of significance (p=0.058)
on the edge of significance (p>0.08)
on the limit to significant (p=0.06)
on the margin of significance (p=0.051)
on the threshold of significance (p=0.059)
on the verge of significance (p=0.053)
on the very borderline of significance (0.05<p<0.06)
on the very fringes of significance (p=0.099)
on the very limits of significance (0.1>p>0.05)
only a little short of significance (p>0.05)
only just failed to meet statistical significance (p=0.051)
only just insignificant (p>0.10)
only just missed significance at the 5% level
only marginally fails to be significant at the 95% level (p=0.06)
only marginally nearly insignificant (p=0.059)
only marginally significant (p=0.9)
only slightly less than significant (p=0.08)
only slightly missed the conventional threshold of significance (p=0.062)
R97: ‘Summary of findings’ table

### Summary of Findings for the Main Comparison

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Intervention: Restrictive justice conference for reducing recidivism in young offenders</th>
<th>Relative effect (95% CI)</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suspected/confirmed recidivism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Controlled</td>
<td>0.9 (0.83 to 1.0)</td>
<td>627</td>
<td>very low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Restrictive justice conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assumed risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Corresponding risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number receiving administrative data: Follow-up 1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study population</td>
<td></td>
<td>0.9 (0.83 to 1.0)</td>
<td>627</td>
<td>very low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% CI: 0.83 to 1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% CI: 0.68 to 1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-intervention monthly offending rate (any offence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number receiving administrative data: Follow-up 1 year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study population</td>
<td></td>
<td>0.9 (0.83 to 1.0)</td>
<td>627</td>
<td>very low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95% CI: 0.83 to 1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Summary of Findings Table:**

- **Victim characterization:**
  - 630 per 1000
  - 630 per 1000
- **Study population:**
  - OR 4.03
  - (95% CI: 2.73 to 5.75)
- **I2:**
  - 43%
  - 2 studies
- **Number receiving administrative data:**
  - 630 per 1000
  - 630 per 1000
  - 630 per 1000
- **Post-intervention monthly offending rate (any offence):**
  - 321
  - OR 0.9
  - (95% CI: 0.68 to 1.00)

**Notes:**
- The basis for the assumed risk (e.g., the median control group risk across studies) is provided in footnotes. The contamination risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).
- CI: Confidence interval; OR: Odds ratio

**GRADE Working Group grades of evidence:**
- **High:** Further research is very unlikely to change our confidence in the estimate of effect.
- **Moderate:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- **Low:** 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:** We are very uncertain about the estimate.

- Randomization procedures were adequate. However, the risk of bias due to absence of blinding procedures and self-selection was high.
- Heterogeneity indicated by I² statistic (79%), although no significant detection was indicated (P = 0.17).
- Only two included studies make it difficult to ascertain the likelihood of publication bias.
- Review authors' choice of items to use as measures of imprecision is somewhat subjective.
- Review authors' choice of items to use as measure of confounding is somewhat subjective.
- There is moderate heterogeneity according to the I² statistic (59%), but this is not supported by the statistical significance (P = 0.12).
- Review authors' choice of items to use as measure of self-selection is somewhat subjective.
- There is moderate heterogeneity according to the I² statistic (57%), but this is not supported by the statistical significance (P = 0.44).
- There is moderate heterogeneity according to the I² statistic (57%), but this is not supported by the statistical significance (P = 0.18).
C76: Assessing the quality of the body of evidence

- **GRADE considerations**

- **Draw conclusions about the quality of evidence within the text of the review**

---

**Incorporating GRADE in Cochrane Reviews: Feedback from the CEU screening programme**

Faty Leclercq, Nancy Santens, Miranda Compton, Rachel Marshall & Oya M. Ogilvie

Assessing the quality of the evidence is an integral part of undertaking a Cochrane Review. GRADE is an established method that helps authors rate the quality of evidence and to communicate the key results of systematic reviews to users. It is a mandatory MEGIR requirement to state the interpretation of the evidence on the five GRADE considerations (imbalance, imprecision, inconsistency, indirectness, and publication bias).

Since the start of the CEU review screening programme in September 2011, we have been able to see how GRADE is used to assess and communicate the quality of a body of evidence. We also recognize some important challenges. Cochrane Review authors and editorial groups have encountered with implementing GRADE.

**Assessing the quality of the evidence**

As a result of this experience, we are sharing our thoughts on some of the reviews that highlight key aspects of incorporating GRADE methods and ratings into the text of Cochrane Reviews:

1. **Describing methods for assessing quality of the evidence**

   There is often only limited information presented in reviews about the implementation of GRADE. Given that GRADE is a method, it should be acknowledged as such under 'Data collection & analysis'. Methods for rating the quality of evidence should be considered as easily as possible in the review process, ideally at the protocol stage. However, even if GRADE has been adopted post-protocol, it is useful to know how this method has been applied to rate the quality of evidence. The following examples present information relevant to the implementation of GRADE methods and the selection of outcomes that are used in a Summary of Findings table (although note that the Cochrane Handbook recommends that the main outcomes for a Summary of Findings table should generally be included under Types of outcome measures).

   **Good practice examples**

   - For assessments of the overall quality of evidence for each outcome that included pooled data from RCTs only, we downgraded the evidence from 'High quality' by one level for serious or by two for very serious study limitations (risk of bias, indirectness of evidence, serious inconsistency, important effect estimates or potential publication bias). Data from observational studies started at lower quality.


   - The GRADE approach was employed in Interact, Ethiopia (Langendijk 2012) and the GRADE profiler (GRADEPRO) allowed us to import data from Review Manager 5.2 (RevMan). Managed to create a Summary of Findings table. These tables provide outcome-specific information concerning the overall quality of evidence from studies included in the comparison, the magnitude of effect of the interventions examined, and the sum of available data on the outcomes as considered.
C78: Formulating implications for practice
R101: Implications for practice

- Base conclusions only on findings from the synthesis.
- Avoid making recommendations for practice.
  - red flag words/phrases: ‘should’, or ‘safe’, etc)

AUTHORS’ CONCLUSIONS

Implications for practice
Due to the overall low quality of the included studies in this review, it is difficult to state with certainty the implications for the continued practice of restorative justice conferencing. While most included studies demonstrated low risk of selection bias and re-

Implications for research
This review has highlighted the need for further research in this area. With regards to study design, further attempts at randomised controlled trials are strongly encouraged. In particular, the pos-
Key consideration

• Can see here, a key aspect that we’re concerned with when we’re considering these ‘Interpretation’ domains is the GRADE and SoF table.

• So let’s stop now and consider these components of a review in more detail.
Checking ‘Summary of findings’ tables in Cochrane protocols and reviews
Key learning points

- SoF table(s) must be planned and thought out in detail right from the start (i.e. protocol stage)
- Creating the table is not an automatic process – manual amendments still need to be made
- The GRADE considerations are not always clear cut
  - Each involves a degree of subjective decision making
  - Each must be clearly justified in footnotes
- The findings from the table should be embedded throughout the review
Outline

• Planning the SoF table (drafting the protocol)

• Creating the SoF table (completing the review)
  ➢ Using the GDT Tool
  ➢ Using the GRADE considerations

• Presenting the SoF table (drafting the review)
Step 1: Planning the SoF table
Protocol common errors

• No plan in the protocol for including a SoF table

• Plan included as a brief sentence at the end of an existing section. No clear plan regarding:
  - Choice of comparisons and outcomes
  - How quality will be assessed using GRADE
  - Who will be involved in assessing quality
Avoiding protocol common errors

- Separate, headed protocol section on SoF tables
- One table per comparison (not per outcome)
- Seven clinically important outcomes
  - Consistent with review Objectives/PICO
  - Balanced overview – showing both ‘benefit’ and ‘harm’
- All GRADE considerations clearly described
- Quality assessed by two (unbiased) review authors
Example A

Data synthesis

Meta-analysis will be performed on the results assuming that at least two studies suitable for inclusion are found. Due to the expected heterogeneity among included studies, we will use a random-effects meta-analysis. When meta-analysis is inappropriate, we will provide a narrative description of the study results alone. We will perform both fixed and random-effects analyses as part of a sensitivity analysis. We will classify the quality of the evidence into one of four categories according to the GRADE approach. We will present the results of the GRADE assessment in a 'Summary of findings' table.
**Example B**

*‘Summary of findings’ table*

Based on the methods described in Chapter 11 of the Cochrane Handbook for Systematic Reviews of Interventions (Schünemann 2011), we will prepare a ‘Summary of findings’ table to present the meta-analysis results. Results of the meta-analysis will be presented for the main comparisons of the review, the primary outcome child pain and the following secondary outcomes: child satisfaction with virtual reality simulation, child pain-related distress and parent anxiety, as outlined in the section on Types of outcome measures. For each assumed risk cited in the table(s), we will provide a source and rationale, and the GRADE system will be used to rank the quality of the evidence using GRADEprofiler (GRADEpro) software (Schünemann 2011). If meta-analysis is not possible, we will present results in a narrative ‘Summary of findings’ table format (drawing on Chan 2011 as an example).

Step 2: Creating the SoF table
Creating the SoF table

• Should be created AFTER pooled analyses are completed, but BEFORE Results section is formally written
  ➢ It is possible to draft the table at protocol stage, and populate it after results are pooled

• Create manually (using RevMan) or online (using the Guideline Development Tool (GDT))
  ➢ Table created from GDT is a starting point, not a final product – manual amendments must still be made!
**Comparison**

**Patient or population:** CD in people with DS  
**Setting:**  
**Intervention:** Donepezil  
**Comparison:** Placebo

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Anticipated absolute effects* (95% CI)</th>
<th>Relative effect (95% CI)</th>
<th>N of participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive function</strong></td>
<td>The mean cognitive function was 0 The mean cognitive function in the intervention group was 0.52 standard deviations undefined more (0.27 fewer to 1.31 more)</td>
<td>-</td>
<td>105 (3 RCTs)</td>
<td>☐ ☐ ☐ ☐</td>
<td>VERY LOW 1,2,3</td>
</tr>
<tr>
<td><strong>Behavioural Problems</strong></td>
<td>The mean behavioural Problems was 0 The mean behavioural Problems in the intervention group was 0.42 standard deviations undefined more (0.06 fewer to 0.89 more)</td>
<td>-</td>
<td>157 (3 RCTs)</td>
<td>☐ ☐ ☐ ☐</td>
<td>VERY LOW 1,3,4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adverse Events</th>
<th>Study population</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>381 per 1000</td>
<td>667 per 1000</td>
<td>(502 to 799)</td>
<td>OR 3.26 (1.64 to 6.47)</td>
<td>171 (3 RCTs)</td>
</tr>
<tr>
<td><strong>Moderate</strong></td>
<td>214 per 1000</td>
<td>471 per 1000</td>
<td>(309 to 638)</td>
<td></td>
<td>LOW 1,3</td>
</tr>
</tbody>
</table>

*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; SMD: Standardised mean difference; OR: Odds ratio

**GRADE Working Group grades of evidence**  
**High quality:** We are very confident that the true effect lies close to that of the estimate of the effect  
**Moderate quality:** We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different  
**Low quality:** Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect  
**Very low quality:** We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect

1. Downgraded one level for high risk of bias (Kishnai 2009)  
2. Downgraded two levels for heterogeneity ($I^2 = 74\%$)  
3. Downgraded one level for inconsistency (wide confidence interval)  
4. Downgraded one level for heterogeneity ($I^2 = 36\%$)  
5. 
Common errors in SoF tables

- Details frequently missing from tables:
  - Comparison
  - PICO and setting
  - Length of follow-up
  - Outcome measurements

- Inconsistency between the data presented in SoF tables, and data presented in ‘Data and analysis’ section

- Planned outcomes omitted from final table, most often due to:
  - No data available
  - Narrative synthesis only
Assessing evidence with GRADE

- Detailed guidance available online (wwwGRADEpro.org > GRADE Handbook)
### GRADE considerations

<table>
<thead>
<tr>
<th>Downgrade for:</th>
<th>Upgrade for:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study limitations</td>
<td>Large magnitude of effect</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Dose response gradient</td>
</tr>
<tr>
<td>Indirectness</td>
<td>Effect of plausible residual confounding</td>
</tr>
<tr>
<td>Imprecision</td>
<td></td>
</tr>
<tr>
<td>Publication bias</td>
<td></td>
</tr>
</tbody>
</table>

- Judgements are not always clear cut – borderline decisions are acceptable, providing they are robust and transparent
Study Limitations

Common error:

Downgrading an outcome for ‘study limitations’ on the basis of:

• Irrelevant studies
• Irrelevant Risk of Bias domains
Example 1

‘Downgraded one level due to risk of bias. There are serious concerns in relation to blinding of participants and outcome assessors.’
Example 1

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Events</th>
<th>Total</th>
<th>Control Events</th>
<th>Total</th>
<th>Weight</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
<th>Total (95% CI)</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>112</td>
<td>176</td>
<td>16</td>
<td>168</td>
<td>41.8%</td>
<td>16.63 [8.13, 30.82]</td>
<td>319</td>
<td>9.91 [6.35, 15.47]</td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>21</td>
<td>111</td>
<td>0</td>
<td>109</td>
<td>51.8%</td>
<td>2.50 [1.13, 5.96]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 6</td>
<td>27</td>
<td>32</td>
<td>6</td>
<td>34</td>
<td>6.4%</td>
<td>25.20 [6.87, 92.40]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>160</td>
<td></td>
<td>311</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total events
Heterogeneity: $\chi^2 = 14.84$, df = 2 ($P = 0.0006$), $I^2 = 87%$
Test for overall effect: $Z = 10.10$ ($P < 0.00001$)

Risk of bias legend:
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias
Study Limitations

How to avoid:

• Consider how each Risk of Bias domain may impact the specific outcome.

• Consider the study limitations of the studies contributing to each outcome only.
Inconsistency

Common error:

Assuming that a high $I^2$ value automatically leads to downgrading for inconsistency.
Example 2

Heterogeneity was high ($I^2=65\%$), so the quality of evidence was downgraded one level due to inconsistency.
Inconsistency

How to avoid:

• Consider the consistency of the direction of effect.
• Consider how subgroups have been used to explain heterogeneity.
Indirectness

Common error:

Frequently ignored/omitted from the process due to poor understanding of how to assess ‘indirectness’.
Example 3

Objectives
The aim of the review was to explore the effectiveness of treatment Z in adult community-dwelling patients with syndrome Y.

Main results
Seven studies recruiting 1450 participants met the review inclusion criteria. The studies recruited community-dwelling people with severe syndrome Y who had a mean age of 70 years. In five of the included studies (Study 1, Study 2, Study 3, Study 5, Study 6), participants were 65 years or older.

Figure 8 (Analysis 1.5)

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Treatment Z Events</th>
<th>Events</th>
<th>Total</th>
<th>Weight</th>
<th>M-H, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 3</td>
<td>45</td>
<td>176</td>
<td>18</td>
<td>168</td>
<td>54.3%</td>
</tr>
<tr>
<td>Study 4</td>
<td>21</td>
<td>111</td>
<td>9</td>
<td>109</td>
<td>32.8%</td>
</tr>
<tr>
<td>Study 7</td>
<td>16</td>
<td>32</td>
<td>6</td>
<td>34</td>
<td>13.0%</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>319</td>
<td>311</td>
<td>100.0%</td>
<td>3.23</td>
<td>[2.05, 5.07]</td>
</tr>
</tbody>
</table>

Total events 82 31
Heterogeneity: Chi² = 6.66, df = 2 (P = 0.071); I² = 0%
Test for overall effect: Z = 5.97 (P = 0.0003)

Caption
Forest plot of comparison: 1 Treatment Z versus control, outcome: 1.5 Hospital admission.

‘Statistical analyses were all based on direct comparisons’
Indirectness

How to avoid:

• Consider how applicable were the Populations/ Interventions?
• Were ‘surrogate’ outcomes used?
• Were follow up lengths sufficient?
Imprecision

Common error:

Basing the decision solely on whether the confidence interval is too ‘wide’ or crosses the ‘line of no effect’.
Example 4

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Experimental Events</th>
<th>Total</th>
<th>Control Events</th>
<th>Total</th>
<th>Weight</th>
<th>Odds Ratio M-H, Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 3</td>
<td>36</td>
<td>50</td>
<td>8</td>
<td>60</td>
<td>50.5%</td>
<td>16.71 [6.35, 43.96]</td>
</tr>
<tr>
<td>Study 4</td>
<td>21</td>
<td>35</td>
<td>5</td>
<td>35</td>
<td>49.5%</td>
<td>9.00 [2.81, 28.81]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>57</td>
<td>85</td>
<td>95</td>
<td>95</td>
<td>100.0%</td>
<td>12.89 [6.13, 27.09]</td>
</tr>
</tbody>
</table>

Heterogeneity: \( \chi^2 = 0.64, \text{df} = 1 \) (\( P = 0.42 \)); \( P = 0% \)

Test for overall effect: \( Z = 6.75 \) (\( P < 0.00001 \))

‘95% Confidence Interval excluded no effect, so there was no need to downgrade for imprecision’
Imprecision

How to avoid:

• Consider whether
  ➢ ‘Optimal Information Size’ is met
  ➢ Confidence Interval fails to exclude ‘important’ benefit or ‘important’ harm.
  ➢ 95% Confidence Interval excludes no effect.
Publication Bias

Common error:

Judgement based only on Funnel Plot asymmetry, thus omitted from consideration if funnel plots were not constructed.
Example 5

There was no indication of publication bias, as funnel plot was not constructed, as less than 10 studies were included in the analysis.
Publication Bias

How to avoid:

• Consider beyond the funnel plot;
  ➢ Selective Outcome Reporting judgements?
  ➢ Search sufficiently comprehensive?
  ➢ Discrepancy between published and unpublished data?
Common errors when upgrading

**Common error:** Upgrading evidence from RCTs.

**How to avoid:**

- Decision to upgrade quality of evidence should only be made when serious limitations in any of the 5 areas reducing the quality of evidence are absent.

- Theoretically possible to rate up results from randomized control trials, but compelling example yet to be found.
Footnotes

Common Errors:

• Inconsistency between footnotes and final judgement.

• No details in footnotes of which GRADE consideration is being referred to, or how many levels each footnote is downgrading the outcome.
## Footnotes

<table>
<thead>
<tr>
<th>Victim/key stakeholder satisfaction with overall process participant interviews</th>
<th>Study population</th>
<th>OR 4.03 (0.59 to 27.75)</th>
<th>428 (2 studies)</th>
<th>very low[^3,^9]</th>
</tr>
</thead>
<tbody>
<tr>
<td>690 per 1000</td>
<td>900 per 1000 (568 to 984)</td>
<td>Moderate</td>
<td>640 per 1000</td>
<td>878 per 1000 (512 to 980)</td>
</tr>
</tbody>
</table>

*The basis for the assumed risk (e.g. the median control group risk across studies) is provided in footnotes. The corresponding risk (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). CI: Confidence interval; OR: Odds ratio

**GRADE Working Group grades of evidence**

- **High quality:** Further research is very unlikely to change our confidence in the estimate of effect
- **Moderate quality:** Further research is likely to have an important 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:** We are very uncertain about the estimate

[^3]: Random allocation procedures were adequate. However, the risk of bias due to absence of blinding procedures and self-selection was high
[^9]: Heterogeneity indicated by the I² statistic (70%), although no significant detection was indicated (P = 0.07)
[^3]: Only two included studies make it difficult to ascertain the likelihood of publication bias
[^3]: Review authors’ choice of item to use as measure of remorse is somewhat subjective
[^3]: Review authors’ choice of item to use as measure of recognition of wrongdoing is somewhat subjective
[^3]: There is moderate heterogeneity according to the I² statistic (59%), but this is not supported by the statistical significance (P = 0.12)
[^3]: Review authors’ choice of item to use as measure of self-perception is somewhat subjective
[^3]: There is moderate heterogeneity according to the I² statistic (87%), but this is not supported by the statistical significance (P = 0.44)
[^3]: There is moderate heterogeneity according to the I² statistic (87%), but this is not supported by the statistical significance (P = 0.16)
Avoiding Footnote common errors

• One footnote per outcome
• Standard template for footnotes:
  ➢ “We downgraded (X) levels for (serious/very serious) (GRADE consideration) due to (brief description of support for judgement)”
Small Group Exercise

- Read through the SoF table and the accompanying review sections

- Check for common errors in:
  - the SoF table
  - the use of the GRADE considerations
Step 3: Presenting the SoF table
Common errors

The outputs of the SoF table, and GRADE quality ratings should be drawn upon anywhere the findings of the review are being described, summarised or interpreted. Specifically:

- Abstract
- Plain language summary
- Results > Effects of interventions
- Discussion > Quality of the evidence
- Authors’ conclusions > Implications for practice
Abstract

Using a wider gauge needle (23 G 25 mm) may slightly reduce procedural pain (low quality evidence) and probably leads to a slight reduction in the duration of crying time immediately after vaccination (moderate quality evidence) compared with a narrower gauge (25 G 25 mm) needle (one trial, 320 participants). The effects are probably not large enough to be of any clinical relevance. The 25 G 25 mm needle may produce a small reduction in the incidence of local reactions after each dose of a DTwP vaccine compared with the 23 G 25 mm needle, but the effect estimates are imprecise (low quality evidence, two trials, numbers of participants in analyses range from 100 to 459).

Plain language summary

Quality of the evidence

We included five studies involving 1350 people. We rated the quality of the evidence from studies using four levels: very low, low, moderate, or high. Very low quality evidence means that we are very uncertain about the results. High quality evidence means that we are very confident in the results. There were problems with the design of some studies and there were not enough data to answer some parts of our review question. The quality of the evidence from two studies was too low to allow us to draw any conclusions about the effects of the needles that were compared in the studies. There was sufficient evidence from the remaining three studies to allow us to reach some conclusions.

Results: Effects of intervention

38 mm versus 25 mm needles - effects on vaccine immunogenicity

In Middleman 2010, seroprotection rates against hepatitis B were 93% (14/15) in the 38 mm group and 91% (10/11) in the 25 mm group (RD 2%, 95% CI -19% to 24%). Median antibody titres to hepatitis B surface antigen were higher in the 38 mm compared with the 25 mm group (345.4 mIU/mL (IQR 243 to 464.2) in 38 mm group versus 189.8 mIU/mL (IQR 143.6 to 324.7) in 25 mm group; P value = 0.03). The latter analysis did not include the two trial participants (one in each needle size group) who failed to reach antibody titre level thresholds of protection against hepatitis B.

We judged the quality of evidence for these immunogenicity outcomes to be very low. We downgraded the evidence by one level for indirectness due to use of a substitute endpoint in lieu of patient-important outcomes, by one level for imprecision due to the width of the CIs around effect estimates and by one level for risk of bias taking into account the absence of allocation concealment and the disparity between the numbers of participants randomised and analysed in the trial.

Discussion: Quality of evidence

Limitations in study design and implementation
Our risk of bias assessment showed that limitations in study design and implementation varied considerably among the included studies (Risk of bias in included studies). Only eleven studies clearly addressed the randomisation method and only four study authors stated how they concealed the allocation, which represents a potential risk of selection bias. More than half of the studies had a double-blind design. Of the 14 single-blind studies, only the investigator was blinded in 12 as the study medications varied in application frequency or were applied in different vehicles, however in 2 studies it was not clear who was blinded. Given that most of the ten non-blinded trials assessed interventions of minor clinical interest, the risk of performance and detection bias for comparisons of major interest may not be unduly affected. We rated one third of the trials to be at high or unclear risk of attrition bias.

Indirectness of the evidence
The studies included in this review assessed representative populations, though only a few studies included children (less than 18 years). However, psoriasis in children is a rather rare condition. In this review, we included vehicle- and active controlled trials, which allowed a clear judgment on comparative efficacy for most interventions.

Inconsistency of the results
In only three instances we downgraded the quality of evidence because of heterogeneity among the trial results. However, in one case heterogeneity was only moderate, and thus we did not seek to identify a plausible explanation. Study results from the 4 trials (Jemec 2008, Kragballe 2008, NCT01195831, van de Kerkhof 2009) that assessed 'IGA response' for the comparison 'steroid plus vitamin D versus vitamin D' were substantially heterogenous. The study populations differed in mean age and percentage of female participants. Two of the studies only masked the outcome assessor (Jemec 2008, Kragballe 2008) and only two had a double-blind design (Jemec 2008, van de Kerkhof 2009). However, we had serious doubts that these aspects alone were responsible for the variability of results.

Imprecision of the results
We lowered the quality of evidence in only two instances because of serious imprecision. In both cases the confidence interval crossed the Minimal Important Difference (MID) thresholds.

Publication bias
The assessment of publication bias was not feasible, as none of the comparisons included more than 10 studies. For this reason we did not create any funnel plot, since it would not give any meaningful information.

Authors’ conclusions

Implications for practice

Our review findings are most applicable to healthy infants between the ages of approximately two and six months undergoing intramuscular vaccination in the anterolateral thigh with combined vaccines containing diphtheria, tetanus, and whole-cell pertussis antigens (DTPwP vaccines) using an injection technique (WHO technique) where the skin is stretched flat and the needle inserted at a 90° angle through the skin and up to the needle hub:

1. using either a 25 G 25 mm or a 23 G 25 mm needle for the vaccination procedure probably leads to fewer severe and non-severe post-vaccination local reactions while achieving a comparable immune response to 25 G 16 mm needles (moderate quality evidence);

2. using a wider gauge 23 G 25 mm needle may slightly reduce the pain associated with the vaccination procedure (low quality evidence) and probably leads to a slight reduction in the duration of crying time immediately following vaccination (moderate quality evidence) compared with a narrower gauge 25 G 25 mm needle. The estimated effects are probably not large enough to be of any practical importance to patients, parents, and healthcare providers;

Summary

- SoF table must be planned and thought out in detail right from the start (i.e. protocol stage)
- Creating the table is not an automatic process – manual amendments still need to be made
- The GRADE considerations are not always clear cut
  - Each involves a degree of subjective decision making
  - Each must be clearly justified in footnotes
- The findings from the table should be embedded throughout the review
Consistency of reporting
| R101 | Implications for practice | Provide a general interpretation of the evidence so that it can inform healthcare or policy decisions. Avoid making recommendations for practice. | | MECR conduct standard 79 (Avoid | |
| R11  | Abstract, Main results: bias assessment | Provide a comment on the findings of the bias assessment. | | | |
| R12  | Abstract, Main results: findings | Report findings for all primary outcomes, irrespective of the strength and direction of the result, and of the availability of data. | | | |
| R13  | Abstract, Main results: adverse effects | Ensure that any findings related to adverse effects are reported. If adverse effects data were sought, but availability of data was limited, this should be reported. | | See Handbook 11.8 | |
| R18  | Consistency of Summary versions of the review | Ensure that reporting of objectives, important outcomes, results, caveats and conclusions is consistent across the text, the abstract, the plain language summary and the ‘Summary of findings’ table (if included). | | Summary versions of the review should read in isolation from the rest of the report | |
| R86  | Consistency of results | Ensure that all statistical results presented in the main review text are consistent between the text and the ‘Data and analysis’ tables. | | | |
• Provide a comment on the findings of the bias assessment.

• Ensure that any findings related to adverse effects are reported.

Main results

Four trials including a total of 1,447 young offenders were included in the review. Results failed to find a significant effect for restorative justice conferencing over normal court procedures for any of the main analyses, including number re-arrested (odds ratio (OR) 1.00, 95% confidence interval (CI) 0.59 to 1.71; \( P = 0.99 \)), monthly rate of reoffending (standardised mean difference (SMD) -0.06, 95% CI -0.28 to 0.16; \( P = 0.61 \)), young person’s remorse following conference (OR 1.73, 95% CI 0.97 to 3.10; \( P = 0.06 \)), young person’s recognition of wrongdoing following conference (OR 1.97, 95% CI 0.81 to 4.80; \( P = 0.14 \)), young person’s self-perception following conference (OR 0.95, 95% CI 0.55 to 1.63; \( P = 0.85 \)), young person’s satisfaction following conference (OR 0.42, 95% CI 0.04 to 4.07; \( P = 0.45 \)) and victim’s satisfaction following conference (OR 4.05, 95% CI 0.56 to 29.04; \( P = 0.16 \)). A small number of sensitivity analyses did indicate significant effects, although all are to be interpreted with caution.

Authors’ conclusions

There is currently a lack of high quality evidence regarding the effectiveness of restorative justice conferencing for young offenders. Caution is urged in interpreting the results of this review considering the small number of included studies, subsequent low power and high risk of bias. The effects may potentially be more evident for victims than offenders. The need for further research in this area is highlighted.

E.G. – no mention of adverse effects)
R12: Abstract, Main results: findings

- Consistently report findings for all primary outcomes across:
  - Abstract
  - PLS
  - SoF

## Summary of Findings for the Main Comparison (Supplementary)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of Trials</th>
<th>Number of Studies</th>
<th>Number of Participants</th>
<th>Risk of Bias</th>
<th>Heterogeneity</th>
<th>Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>10</td>
<td>8</td>
<td>900</td>
<td>Low</td>
<td>Low</td>
<td>Good</td>
</tr>
<tr>
<td>PLS</td>
<td>12</td>
<td>10</td>
<td>1200</td>
<td>Low</td>
<td>Low</td>
<td>Good</td>
</tr>
<tr>
<td>SoF</td>
<td>15</td>
<td>12</td>
<td>1500</td>
<td>Low</td>
<td>Low</td>
<td>Good</td>
</tr>
</tbody>
</table>

### Main results

Four trials including a total of 1,647 young offenders were included in the review. The studies failed to find a significant effect of innovative justice conferencing vs. normal court procedures for any of the main analyses, including number of contacts (odds ratio 0.80, 95% confidence interval 0.59 to 1.08, P = 0.19), number of contacts (odds ratio 0.79, 95% confidence interval 0.58 to 1.08, P = 0.14), young person's self-reported following conference (OR 0.89, 95% CI 0.72 to 1.11, P = 0.30), young person's self-esteem following conference (OR 0.89, 95% CI 0.72 to 1.11, P = 0.30), and victim satisfaction following conference (OR 0.89, 95% CI 0.72 to 1.11, P = 0.30). A small number of sensitivity analyses did indicate significant effects, although all were not interpreted with caution.

### Authors' conclusions

There is currently no high quality evidence regarding the effectiveness of innovative justice conferencing for young offenders. Caution is urged in interpreting the results of this review considering the small number of included studies, substantial low power and high risk of bias. The effects may potentially be more evident for victims than offenders. The need for further research in this area is highlighted.

### Plain Language Summary

Innovative justice conferencing for reducing recidivism in young offenders

The number of young people who commit offenses remains an area of concern in many countries, particularly considering the high rate of those who do go on to offend. An increasingly popular technique used with young offenders is an alternative to normal court proceedings, to conduct a Revolutionary Justice Conference. This conference involves bringing together the offender, the victim or victims, the representatives of both and a conference moderator. The conference gives all individuals involved a chance to share their experiences and to decide together how best to repair the harm caused by the offense. It is believed that providing an opportunity for the offender to make amends for what has been done, along with the victim's forgiveness, increases the satisfaction of all those involved and reduces the likelihood of reoffending. The purpose of this review was to look at whether young people who are part of a conference justice conference were less likely to reoffend than those who go through normal court proceedings. Four randomized controlled trials were included in the review. Findings indicate that there was no difference between those who were part of innovative justice conferencing and those in normal court proceedings in terms of the rate of reoffending after the intervention. There was also no difference between these two groups in terms of change in their attitudes toward their participation with the process. Results may indicate that victims who are part of a conference justice conference are more satisfied than those who are part of court proceedings. The quality of the included studies was low. More high-quality research using a design that participants are randomly allocated to an intervention or control group is needed.
R18: Consistency of summary versions of the review
R86: Consistency of results

- Objectives, outcomes, results, caveats and conclusions
- ‘Data and analysis’ tables.
Common Issues

Inconsistency within full text & between data/summary versions

Omission

- key information (RoB; key outcomes & GRADE ratings)
Abstract, Main results: findings

- Consider adding a statement about the risk of bias to the results of the abstract. The statement at the end of the abstract conveys slightly different information about the study follow-up and sample size. I have copied the following text from the discussion as I think it is a nice summary of the ROB judgements: ‘More than half the trials had no blinding of participants or outcome assessment. Most of the trials had uncertain selection and reporting bias as they did not report allocation concealment and there was evidence of selective reporting’.
Common Issues

Inconsistency within full text & between data/summary versions

Omission
- key information (RoB; key outcomes & GRADE ratings)

Differences
- methods reported/used
- analyses reported table/text
We assessed the following for each study.

1. Random sequence generation (checking for possible selection bias). We assessed the method used to generate the allocation sequence as: low risk of bias (i.e. any truly random process, for example random number table; computer random number generator); or unclear risk of bias (when the method used to generate the sequence is not clearly stated). We excluded studies at high risk of bias that used a non-random process (for example, odd or even date of birth, hospital or clinic record number).

2. Allocation concealment (checking for possible selection bias). The method used to conceal allocation to interventions prior to assignment determines whether intervention allocation could have been foreseen in advance of, or during, recruitment, or changed after assignment. We assessed the methods as: low risk of bias (for example, telephone or central randomisation; consecutively numbered, sealed, opaque envelopes); or unclear risk of bias (when the method is not clearly stated). We excluded studies that did not conceal allocation and were therefore at a high risk of bias (for example, open list).

3. Blinding of outcome assessment (checking for possible detection bias). For this review, it was unlikely that these studies would be blinded to the investigators or participants. We assessed any methods used to blind the outcome assessors from knowledge of which intervention a participant received. We assessed the methods as: low risk of bias (e.g. study stated that it was single-blinded and described the method used to achieve blinding of the outcome assessor); unclear risk of bias (study stated that outcome assessors were blinded but did not provide an adequate description of how it was achieved); or high risk of bias (outcome assessors were not blinded). We excluded studies at a high risk of bias that were not single-blind.

4. Incomplete outcome data (checking for possible attrition bias due to the amount, nature, and handling of incomplete outcome data). We assessed the methods used to deal with incomplete data as: low risk of bias (i.e. less than 10% of participants did not complete the study, or the study used 'baseline observation carried forward' (BOCF) analysis, or both); unclear risk of bias (used 'last observation carried forward' (LOCF) analysis); or high risk of bias (used 'complete' analysis).

5. Size of study (checking for possible biases confounded by small size). We assessed studies as being at low risk of bias (200 participants or more per treatment arm); unclear risk of bias (50 to 199 participants per treatment arm); or high risk of bias (fewer than 50 participants per treatment arm).

**Risk of bias in included studies**

See Figure 2 and Figure 3 for the summary graphs.

**Allocation (selection bias)**

All three included studies were randomised.

Two studies were at low risk of random sequence generation (select sequence was only described as 'randomly assigned').

Allocation concealment (selection bias) was low for Sutters 2010, b.

**Blinding (performance bias and detection bias)**

It was not possible to assess the included studies for blinding of the three included studies were at high risk of detection bias. Parents w the assessors were blinded to the intervention.

**Incomplete outcome data (attrition bias)**

All three included studies were at low risk of attrition bias. In Romal consent.

**Selective reporting (reporting bias)**

All included studies were at low risk of reporting bias.

**Size of study**

---

**Risk of bias table**

<table>
<thead>
<tr>
<th>Bias</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Random sequence generation (selection bias)</td>
<td></td>
</tr>
<tr>
<td>Allocation concealment (selection bias)</td>
<td></td>
</tr>
<tr>
<td>Incomplete outcome data (attrition bias)</td>
<td></td>
</tr>
<tr>
<td>Selective reporting (reporting bias)</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
</tr>
</tbody>
</table>
# Results

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative comparative risks (^{25% \text{ CI}})</th>
<th>Relative effect (^{95% \text{ CI}})</th>
<th>No of Participants (studies)</th>
<th>Quality of the evidence (GRADE)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>BMSG</td>
<td>RR 0.62 (0.31 to 1.28)</td>
<td>1,138 participants</td>
<td>low</td>
<td>A combination of low events and discordant results from one study leads to low confidence in the estimate of the effect. This is likely to change with further research.</td>
</tr>
<tr>
<td>Mortality - Short term follow-up (±12 months)</td>
<td>27 per 1000 (9 to 652)</td>
<td>RR 0.29 (0.15 to 0.58)</td>
<td>494 participants (8 studies)</td>
<td>low</td>
<td>As above.</td>
</tr>
<tr>
<td>Mortality - Long term follow-up (±12 months)</td>
<td>136 per 1000 (9 to 292)</td>
<td>RR 0.29 (0.15 to 0.58)</td>
<td>494 participants (8 studies)</td>
<td>low</td>
<td>As above.</td>
</tr>
</tbody>
</table>

**Primary angioplasty, surgery or administration of stem cell mobilising agents, were included where administered to treatment and control arms equally.**

## Data collection and analysis

Eligibility screening of all references, assessment of trial quality and data extraction was conducted by two reviewers independently. A quantitative evaluation of data was undertaken using fixed effect meta analyses. Heterogeneity was evaluated using the \( I^2 \) statistic; considerable heterogeneity (\( I^2 > 75\% \)) was explored using random effects models and subgroup analysis.

## Main results

Twenty-three RCTs involving a total of 1,255 participants were included in this study. Risk of bias was generally low with the majority of studies reporting appropriate methods of randomisation and blinding. Administration bone marrow stem cell treatment significantly reduced the incidence of long-term mortality (RR 0.52, 95% CI 0.45 to 0.62, \( p = 0.0001 \)) and returns hospitalisation due to heart failure (RR 0.25, 95% CI 0.17 to 0.38, \( p = 0.0001 \)) and stroke volume index (MD 5.62, 95% CI 1.51 to 11.74, \( p = 0.01 \)), and a significant improvement in left ventricular ejection fraction (LVEF) (MD 2.02%, 95% CI 0.20% to 4.73%, \( p = 0.02 \)) all in long term follow up. Overall, a significant reduction in functional class NYHA class in favour of BMSG treatment was observed during short term (MD -0.63, 95% CI -1.09 to -0.18, \( p = 0.04 \)) and long term (MD -0.33, 95% CI -0.66 to -0.00, \( p = 0.05 \)).
Common Issues

Inconsistency within full text & between data/summary versions

Omission
- key information (RoB; key outcomes & GRADE ratings)

Differences
- methods reported/used
- analyses reported table/text

Emphasis
- Certain results, strength of evidence or conclusions in abstract/PLS
Quality of the evidence

For many outcomes the quality of the evidence was low or moderate. For many of the studies the researchers collecting information about infections knew who received which treatment and this could have influenced the results. Overall the quality of evidence from the studies is rather poor. The studies did not use an agreed upon, standardised definition of what constituted infection of the surgical site, which would be useful to include in future research. Many of the studies were published 20 to 30 years ago, so the participants in them may not be representative of people today who require this type of surgery.

Not commented on in discussion, downgrading decisions or main conclusions of review
## PIC O

**Quality of the evidence**

**Overall completeness and applicability of evidence**

**Authors' conclusions**

**Abstract**

**Plain language summary**

---

### Quality of the evidence

**GRADE**

- Level of evidence:
  - A: Strong evidence, consistent results from multiple, well-conducted studies.
  - B: Moderate evidence, conflicting results or significant heterogeneity.
  - C: Low evidence, insufficient data or inconsistent results.

- Recommendation:
  - Strong recommendation, high certainty of effect.
  - Weak recommendation, low certainty of effect.

### Overall completeness and applicability of evidence

- PIC O framework:
  - **Population**: Who is the study about?
  - **Intervention**: What was done?
  - **Comparison**: What was compared?
  - **Outcome**: What was measured?

### Authors' conclusions

- Summary of findings:
  - Table of results:
    - Study or subgroup | L | SE | Total | Weight | M | N | E   |
    - Treatment 1     |   |    |       |        |   |   |     |
    - Placebo         |   |    |       |        |   |   |     |
  - Risk of bias:
    - Selection bias
    - Performance bias
    - Detection bias
    - Attrition bias
    - Other bias

### Abstract

- A brief summary of the study's purpose, methods, results, and conclusions.

### Plain language summary

- A simplified explanation of the study findings for non-specialist readers.
Key questions for assessing inconsistency

• Are the following methods consistent across the review?
  – Risk of bias domains & judgements
  – Measures of effect
  – Unit of analysis
  – Data synthesis
  – Subgroup/sensitivity analyses?

• Are same outcomes reported across summary versions?
  Do results match with data tables?

• Are conclusions in abstract/PLS similar to the implications for practice?
Small Group Exercise

In small groups, we will now screen together a recently submitted Review, using the CEU Screening Tool:

Self-management programs for quality of life in people with stroke
Plain Language Summary

• Aim to summarize the review in a straightforward style that can be understood by consumers of health care
• Plain language summaries are made freely available on the internet, so will often be read as stand-alone documents
• Consist of:
  – Plain Language Title
  – Summary Text
Methodological Expectations of Cochrane Intervention Reviews (MECIR)

Standards for the reporting of Plain Language Summaries in new Cochrane Intervention Reviews (PLEACS)

Version 1.0 19 December 2012

Status: Mandatory means that a new review should not be published if this is not reported. Highly desirable means that this should generally be done, but that there are justifiable exceptions.

Preface

Plain Language Expectations for Authors of Cochrane Summaries (PLEACS) is a special working group comprised of consumers, methodologists, and editors. During the past year, this group has been developing a set of standard requirements for plain language summaries (PLS) of Cochrane Reviews. This work complements the MECIR project which has so far delivered standards relating to conduct and reporting of Cochrane Intervention Reviews [see: http://www.editorial-unit.cochrane.org/mecir].

The standards below summarize proposed attributes of reporting that we consider either mandatory or highly desirable for PLS of Cochrane Intervention Reviews. For each standard we have given a reason for our judgment alongside some examples. These standards are not intended to apply to updated reviews at this point.

During July and August 2012, members of The Collaboration and the public were invited to comment on the draft standards through an open consultation process. Key comments revolved around the issues of 1) reading age for PLS, 2) the presentation of information about systematic reviews and Cochrane in PLS, 3) the use of headings to break-up the text, and 4) explanations about the quality of the evidence. The working group reviewed all the comments and amended the standards in response.

The ordering of the standards reflects the position in which each issue might be expected to be addressed in the PLS. Work on establishing the most suitable format for structuring the PLS is ongoing and as an interim measure we have associated each standard with provisional considerations to help orientate authors, editors and readers (see PLS3 below).

During the early part of 2013, the PLEACS working group will begin development of good-practice examples to aid authors and Cochrane Review Groups implement the standards.

Catherine McIlwain, Consumer Coordinator, The Cochrane Collaboration on behalf of the PLEACS committee†

† Catherine McIlwain, Nancy Santesso, Silvana Simi, Maryann Napoli, Toby Lasserson, Emma Welsh, Rachel Churchill, Tamara Rader, Jackie Chandler, David Tovey, Lorne Becker, Gill Gyte, Amellise Symon
Plain language title

- Restatement of the review’s title using plain language term.
- If review title is easily understood, this should simply be restated as the plain language title.
- Plain language title should not be declarative (it should not reflect the conclusions of the review).
Summary Text

- no more than 400 words
- Consisting of:
  - Review Question
  - Background (why the review is important)
  - Study characteristics
  - Key results (including any adverse effect)
  - Quality of the evidence (limitations of the review)
Review Question

• convey the question addressed by the review so that results and conclusions can be contextualised
• “We reviewed the evidence about the effect of $X$ on $Y$ in people with $Z$.

PLAIN LANGUAGE SUMMARY

Physical fitness training for stroke survivors

Review question

We reviewed the evidence that examines whether physical fitness training is beneficial for a range of health and function outcomes in people with stroke.
Background

- short description of the population, intervention, and outcomes.
- Highlight the uncertainties that the review intended to address.

**Background**

Physical fitness is important to allow people to carry out everyday activities such as walking and climbing stairs. Physical fitness varies among everyone. For example, fitness in men tends to be a little higher than in women and everyone’s fitness becomes reduced as we get older and if we become less physically active. Physical fitness is often particularly low in stroke survivors. It may limit their ability to perform everyday activities and also worsen any stroke-related disability. For this reason fitness training has been proposed as a beneficial approach for people with stroke. However, taking part in fitness training could have a range of other benefits important to people with stroke such as improving cognitive function (thinking skills), improving mood, and quality of life, and it could reduce the chance of having another stroke.
Study characteristics

- State the how current the evidence is: (“The evidence is current to MM YYYY”)

- Include information on the condition, the intervention(s), the population and the setting, total number of included studies, the duration of the trials, number of participants.

Study characteristics

By February 2015 we identified 58 trials for inclusion in the review. The trials involved a total of 2797 participants at all stages of care including being in hospital or back living at home. Most of the people who took part were able to walk on their own. The trials tested different forms of fitness training; these included 1) cardiorespiratory or ‘endurance’ training, 2) resistance or ‘strength’ training, or 3) mixed training, which is a combination of cardiorespiratory plus resistance training.
Key Results

• Present the results for all main outcomes (i.e. SoF table outcomes).
• Present the findings for harms (adverse events).
• Use consistent wording across outcomes
• Explain complicated or uncommon outcomes

Key results

Electromechanical and robot-assisted arm and hand training improved activities of daily living in people after stroke and function and muscle strength of the affected arm. As adverse events such as injuries and pain were seldom described, these devices can be applied as a rehabilitation tool, but we still do not know when or how often they should be used.
Quality of the evidence

- Describe the overall quality of the evidence for each of the main outcomes based on the five GRADE considerations.
- Provide key reasons for the quality of the evidence or limitations in lay terms.

Quality of the evidence

The quality of the evidence was low to very low.

Quality of the evidence

The majority of studies had a low to moderate risk of bias based on study limitations and inconsistency. Most studies reported the outcomes they stated they would.

Quality of the evidence: We classified the quality of the evidence as low for arm function. The quality of the evidence was very low for walking ability, global motor function and independence in performing daily activities. The quality of the evidence for each outcome was limited due to small numbers of study participants, inconsistent results across studies and poor reporting of study details.
General Points

• Limit sentences to no more than approximately 20 words
• Don't introduce more than one idea/statement/point in a sentence.
• Avoid potentially misunderstood words or phrases or words with dual or nuanced meanings (e.g. drugs; diet)
• Avoid more than 2 hard words (e.g., technical words, jargon) in a sentence unless it is a term that is explained
• Avoid words or terms that are regional (A&E versus ER); remember to write for an international audience.
PLAIN LANGUAGE SUMMARY

Stem cell transplantation for ischemic stroke

Once stroke-induced cell damage has occurred, little can be done to improve functional outcome. Stem cell transplantation has been shown to be safe and effective in ischemic stroke animal models but evidence in patients is still lacking. This review found three very small randomized controlled trials. Two of them could not be included in this meta-analysis because additional data are needed. The third one, which included 30 participants, found a statistically non-significant functional improvement in patients treated with autologous mesenchymal stem cells at longer follow-up. No adverse cell-related events were reported. Larger and better-designed trials are needed.
Plain language summary

Timing of Intervention for individuals with recently symptomatic carotid artery stenosis

Background
Ischemic stroke occurs when blood flow is cut off from part of the brain. It can be caused by a plaque on the carotid artery that can lead to an extremely narrowing artery, or disruption of the plaque and obstruction of a distal smaller vessel. Carotid revascularization can reestablish adequate blood flow by removal of the carotid stenosis, which can be accomplished by performing a surgery to remove the plaque, or by deploying a stent to cover the entire lesion.

There is uncertainty about whether to perform the operation immediately, or to wait a few days. Early revascularization can improve cerebral blood flow, and prevent new ischemic events. However, an early intervention may have a higher risk of dislodging the plaque or transformation of a preexisting infarct area into a hemorrhagic lesion.

Review question
We reviewed the effectiveness of performing very early intervention (within two days) compared with delayed treatment (after two days) for individuals with recently symptomatic carotid artery stenosis.

Study characteristics
The review found only one randomized trial which assessed the effect of the timing of surgery with a total of 40 participants, with ages ranging from 47 to 94 years. These studies had been published up to January, 2016.

Key results
From the limited evidence available, the timing of surgery was not a critical factor in determining the outcome from a recently symptomatic carotid artery stenosis.

Quality of the evidence
There is no evidence on the best time for surgical treatment of individuals with recently symptomatic carotid artery stenosis. The overall quality of the evidence was very low due to the small number of participants in only one trial and also because of the missing outcome data. Further studies with a larger number of patients are needed.
Any Final Questions?