Definition of the Different Levels of Evidence (LoE)

Articles on treatment

<table>
<thead>
<tr>
<th>Level</th>
<th>Risk of bias</th>
<th>Study design</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| I     | Low risk     | Study adheres to commonly held tenets of high quality design, execution and avoidance of bias | Good quality RCT | • Random sequence generation  
• Allocation concealment  
• Intent-to-treat analysis  
• Blind or independent assessment for important outcomes  
• Counterinterventions applied equally  
• F/U rate of above 80%  
• Adequate sample size |
| II    | Moderately low risk | Study has potential for some bias; study does not meet all criteria for level I, but deficiencies not likely to invalidate results or introduce significant bias | Moderate or poor quality RCT | • Violation of one of the criteria for good quality RCT  
• Blind or independent assessment in a prospective study, or use of reliable data in a retrospective study  
• Counterinterventions applied equally  
• F/U rate of above 80%  
• Adequate sample size  
• Controlling for possible confounding |
| III   | Moderately high risk | Study has significant flaws in design and/or execution that increase potential for bias that may invalidate study results | Moderate or poor quality cohort  
Case-control | • Violation of any of the criteria for good quality cohort  
• Any case-control study |
| IV    | High risk | Study has significant potential for bias; lack of comparison group precludes direct assessment of important outcomes | Case series  
Any case series design |

Outcomes assessment is independent of healthcare personnel judgment. Reliable data are data such as mortality or re-operation.

Additional domains: dose-response, strength of association, publication bias.

Determination of Overall Strength of Evidence (SoE)

After individual article evaluation, the overall body of evidence with respect to each outcome is determined based on guidelines outlined by the Grades of Recommendation Assessment, Development and Evaluation (GRADE) Working Group and recommendations made by the Agency for Healthcare Research and Quality (AHRQ). Qualitative analysis is performed considering the AHRQ required and additional domains. The table below provides an outline of the method used to determine the final SoE.

Strength of Evidence for Existing Systematic Reviews

Level of evidence ratings for Cochrane reviews and other systematic reviews are assigned a baseline score of High (I/II/C) or Low (III/IV/D). The rating can be upgraded or downgraded based on adherence to the core criteria for methods, qualitative, and quantitative analyses for systematic reviews (there is a reference/evaluation table for this).

The following four possible levels and their definitions are reported:

- **High**: High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
- **Moderate**: Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.
- **Low**: Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and likely to change the estimate.
- **Insufficient**: Evidence either is unavailable or does not permit a conclusion.

All AHRQ “required” and “additional” domains are assessed. Only those that influence the baseline grade are listed in the table.

- Baseline strength: Risk of bias (including control of confounding) is accounted for in the individual article evaluation. High = majority of articles level III; Low = majority of articles level IV/D.
- Downgrade: Inconsistency4 of results (1 or 2); Indirectness of evidence (1 or 2); Imprecision of effect estimates (1 or 2); Sub-group analyses not stated apriori and no test for interaction (2).
- Upgrade: Large magnitude of effect (1 or 2); Dose response gradient (1)

Definitions of the Different Levels of Evidence for Reliability Studies

<table>
<thead>
<tr>
<th>Level</th>
<th>Study type</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| 1     | Good quality study | • Broad spectrum of persons with the expected condition  
• Adequate description of methods for replication  
• Blinded performance of tests, measurements or interpretation  
• Second test/interpretation performed independently of the first |
| 2     | Moderate quality | • Violation of any one of the criteria for a good quality study |
| 3     | Poor quality study | • Violation of any two of the criteria |
| 4     | Very poor quality study | • Violation of all three of the criteria |