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     | Good quality RCT | • Random sequence generation  
|       |              |              | • Allocation concealment  
|       |              |              | • Intent-to-treat analysis  
|       |              |              | • Blind or independent assessment for important outcomes  
|       |              |              | • Co-interventions applied equally  
|       |              |              | • F/U rate of 80%  
|       |              |              | • Adequate sample size |
| II    | Moderately low risk | 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  
|       |              |              | • Co-interventions applied equally  
|       |              |              | • F/U rate of 80%  
|       |              |              | • Adequate sample size  
|       |              |              | • Controlling for possible confounding* |
| III   | Moderately high risk | Moderate or poor quality cohort | • Violation of any of the criteria for good quality cohort  
|       |              |              | • Any case-control design |
| IV    | High risk    | Case-control | • Any case series design |

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

**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 1 (high quality) or 2 (low quality). Levels of evidence ratings are used to determine the final SoE. 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 the effect.
- **Moderate**: Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of the 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 the effect and likely to change the estimate.
- **Insufficient**: Evidence either is unavailable or does not permit a conclusion.

**Criteria for level I, but deficiencies not likely to invalidate results or introduce significant bias**

- Study has signifi cant potential for bias; study does not meet all criteria for level I, but deficiencies not likely to invalidate results or introduce significant bias.
- Study has signifi cant potential for bias; lack of comparison group precludes direct assessment of important outcomes.
- Upgrading: Large magnitude of effect (1 or 2); Dose response gradient (1)
- Downgrading: Inconsistency 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)
- Other than a good case-control study
- A good cross-sectional study

**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 principles 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 criteria in level I. The table below provides an outline of the method used to determine the final SoE.

**Criteria for level I but deficiencies not likely to invalidate results or introduce significant bias**

- Study has signifi cant potential for bias; lack of comparison group precludes direct assessment of important outcomes.
- Upgrading: Large magnitude of effect (1 or 2); Dose response gradient (1)
- Downgrading: Inconsistency 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)
- Other than a good case-control study
- A good cross-sectional study

**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 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 |