**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  
• Counterventions applied equally  
• FU 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  
• Counterventions applied equally  
• FU 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.

**Articles on prognosis or risk**

<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 cohort | • Prospective design  
• Patients at similar point in the course of their disease or treatment  
• FU rate of ≥ 80%  
• Patients followed long enough for outcomes to occur  
• Accounting for other prognostic factors* |
| II    | Moderately low risk | Moderate quality cohort | • Prospective design, with violation of one of the other criteria for good quality cohort study  
• Retrospective design, meeting all of the criteria in level I |
| III   | Moderately high risk | Poor quality cohort | • Prospective design with violation of 2 or more criteria for good quality cohort, or  
• Retrospective design with violation of 1 or more criteria for good quality cohort  
• A good case-control study*  
• A good cross-sectional study* |
| IV    | High risk    | Poor quality-case-control or cross-sectional Case series* | • Other than a good case-control study  
• Other than a good cross-sectional study  
• Any case series* design |

*Applies to cohort studies only.

**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 precepts 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 (BHS) when used. LOW if observational studies were used. 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 table.

**Outcome Strength of evidence Conclusions and comments Baseline Downgrade Upgrade**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Level of evidence</th>
<th>Summary of findings</th>
<th>Baseline</th>
<th>Downgrade</th>
<th>Upgrade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Summary of findings</td>
<td>No Consistent, direct, and precise estimates</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>Summary of findings</td>
<td>No Consistent, direct, and precise estimates</td>
<td>Yes</td>
<td>Large effect</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Summary of findings</td>
<td>Yes (2) Inconsistent indirect</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Required domains: risk of bias, consistency, directness, precision. Plausible confounding that would decrease observed effect is accounted for in our baseline risk of bias assessment through individual article evaluation. Additional domains: dose-response, strength of association, publication bias.

---

**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 |