### 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 > 90%  
• 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  
Good quality cohort | • 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 > 90%  
• Adequate sample size  
• Controlling for possible confounding |
| III   | Moderately high risk | Study has significant flaws in design and/or execution 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 design |
| 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 |

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

*Authors must provide a description of robust baseline characteristics, and control for those that are unequally distributed between treatment groups.*

### 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 domains. The table below provides an outline of the method used to determine the final SoE.

#### Strength of Evidence for Existing Systematic Reviews

<table>
<thead>
<tr>
<th>Level</th>
<th>Study type</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 cohort | • Prospective design  
• Patients at similar point in the course of their disease or treatment  
• F/U rate of ≥ 80%  
• Patients followed long enough for outcomes to occur  
• Accounting for other prognostic factors |
| II    | Moderately low risk | Study has potential for some bias; does not meet all criteria for level I but deficiencies not likely to invalidate results or introduce significant bias | 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  
• A good case-control study  
• A good cross-sectional study |
| III   | Moderately high risk | Study has flaws in design and/or execution that increase potential for bias that may invalidate study results | Poor quality cohort  
Good quality case-control or cross-sectional study | • 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  
• Any case series design |
| IV    | High risk  | Study has significant potential for bias; does not include design features geared toward minimizing bias and/or does not have a comparison group | 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 |

*Cohort studies follow individuals with the exposure of interest over time and monitor for occurrence of the outcome of interest.*

*Authors must consider other factors that might influence patient outcomes and should control for them if appropriate.*

*A good case-control study must have all of the following: all incident cases from the defined population over a specified time period, controls that represent the population from which the cases come, exposure that precedes an outcome of interest, and accounting for other prognostic factors.*

*A good cross-sectional study must have all of the following: a representative sample of the population of interest, an exposure that precedes an outcome of interest (e.g., sex, genetic factor), an accounting for other prognostic factors, and for surveys, at least an 80% return rate.*

*A case-series design for prognosis is one where all the patients in the study have the exposure of interest. Since all the patients have the exposure, risks of an outcome can be calculated only for those with the exposure, but cannot be compared with those who do not have the exposure. For example, a case-series evaluating the effect of smoking on spine fusion that only recruits patients who smoke can simply provide the risk of patients who smoke that result in pseudarthrosis but cannot compare this risk to those that do not smoke.*

### 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 |