Definition of the Different Levels of Evidence (LoE)

### Articles on treatment

| Level  | Risk of bias               | Studies of therapy                                                                 |
|--------|---------------------------|-----------------------------------------------------------------------------------|
| 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 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 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                                                   |
|        |                           | • 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.

### Articles on prognosis or risk

| Level  | Risk of bias               | Studies of prognosis                                                                |
|--------|---------------------------|-------------------------------------------------------------------------------------|
| 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 the rest of the criteria in level I             |
| 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                                |
|        |                           | • Prospectively designed 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                 | 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.
*· Applies to cohort studies only.
*· Authors must consider other factors that might influence patient outcomes and should control for them if appropriate.

**Strength of Evidence for Existing Systematic Reviews**

Level of evidence ratings for Cochrane and other systematic reviews are assigned a baseline score of HIGH (I/II) or LOW (III/IV) after 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 or very low**: Evidence either is unavailable or does not permit a conclusion.

Additional 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.

**Definition of the Different Levels of Evidence for Reliability Studies**

| Level | Study type       | Criteria                                                                 |
|-------|------------------|--------------------------------------------------------------------------|
| 1     | Good quality     | • Broad spectrum of persons with the expected condition                  |
|       |                   | • Adequate description of methods for replication                        |
|       |                   | • Blind 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     | • Violation of any two of the criteria                                   |
| 4     | Very poor quality| • Violation of all three of the criteria                                  |

*a 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.
*b Additional domains: dose-response, strength of association, publication bias.
\[\text{Quality cohort}^2\]