### Articles on treatment

#### Risk of bias description

| Risk of bias description | Studies of therapy | Criteria |
|-------------------------|--------------------|----------|
| 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 author’s primary important outcomes |
|                         | • Co-interventions applied equally | • F/U rate of 80% or > |
|                        | Moderate or poor quality RCT | • Difference in follow-up between groups |
|                        | Moderate or poor quality RCT | • Controlling for possible confounding |
|                        | Poor quality RCT | Moderate or poor quality cohort |
|                        | Moderate or poor quality cohort | • Violation of one or two criteria for good quality RCT |
| High risk               | Study has significant potential for bias, lack of comparison group precludes direct assessment of important outcomes | Case series |
|                        | Case series | Any case series design |

#### Determination of Overall Strength (Quality) of Evidence (SOE)

After individual article evaluation, the overall quality of the body of evidence with respect to each outcome is likely to include steps 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). Quality assessment is performed considering the AHRQ required and additional domains. The table below provides an outline of the methods used to determine the final SOE.

The following four possible levels and their definition will be reported:

- **High**: High confidence that the evidence reflects the true effect. Further research is 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 is unavailable or does not permit a conclusion.

All AHRQ “required” and “additional” domains are assessed. Author’s final determination of overall quality of evidence is made based on all domains.

**Outcome assessment** is independent of health care personnel, investigator, or patient judgment.

**Authors** must provide a description of robust baseline characteristics, and control for those that are unequally distributed between treatment groups. RCTs get credit if there is a similar distribution of baseline characteristics between groups but must also control for confounding if distribution is not similar.

**Reliable data** are data such as mortality or reoperation.

---

### Articles on prognosis or risk

#### Risk of bias description

| Risk of bias description | Studies of prognosis | Criteria |
|-------------------------|---------------------|----------|
| 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 |
|                        | Moderate quality     | • F/U rate of ≥ 80% |
|                        | Poor quality study   | • Patients followed long enough for outcomes to occur |
| High risk               | Study has potential for some bias; does not meet all criteria for good quality cohort, or does not include design features geared toward minimizing bias and/or does not have a comparison group | Poor quality cohort |
|                        | Moderate quality case-control or cross-sectional study | Good quality case-control or cross-sectional study |
|                        | Moderate quality study | • Prospective design with violation of two or more criteria for good quality cohort, or |
|                        | A good case-control study | • Retrospective design with violation of one or more criteria for good quality cohort |
|                        | A good cross-sectional study | • A good cross-sectional study |

**Criteria** are assessed. Only those that influence the baseline grade are listed in table. Baseline strength. Risk of bias (including control of confounding) is accounted for in the individual article evaluations. High quality = majority of articles RCTs; Low quality = majority of articles cohort (observational) studies.

**Outcome measures** are assessed. Only those that influence the baseline grade are listed in table. Baseline strength. Risk of bias (including control of confounding) is accounted for in the individual article evaluations. High quality = majority of articles RCTs; Low quality = majority of articles cohort (observational) studies.

**Dependent variables** are assessed. Only those that influence the baseline grade are listed in table. Baseline strength. Risk of bias (including control of confounding) is accounted for in the individual article evaluations. High quality = majority of articles RCTs; Low quality = majority of articles cohort (observational) studies.

**Additional domains** are assessed. Only those that influence the baseline grade are listed in table. Baseline strength. Risk of bias (including control of confounding) is accounted for in the individual article evaluations. High quality = majority of articles RCTs; Low quality = majority of articles cohort (observational) studies.

### Definitions of the Different Levels of Evidence for Reliability Studies

| Level | Study type | Criteria |
|-------|------------|----------|
| 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 |

---

### 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 if RCTs were used. Low if observational studies were used. The rating can be upgraded or downgraded based on adherence to the core criteria for methods, quality, and quantitative analyses for systematic reviews (there is a reference/evaluation table for this).

---

### Additional Information

- **Outcome assessment** is independent of health care personnel, investigator, or patient judgment.
- **Authors** must provide a description of robust baseline characteristics, and control for those that are unequally distributed between treatment groups. RCTs get credit if there is a similar distribution of baseline characteristics between groups but must also control for confounding if distribution is not similar.
- **Reliable data** are data such as mortality or reoperation.

---

### Required domains

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

---

### Single study

- **Consistency unknown**.