### EBSJ Evidence Assessment: Definition of Risk of Bias (RoB)

#### Articles on treatment

| Risk of bias description | Studies of therapy | Studies of prognosis | Criteria |
|--------------------------|--------------------|----------------------|----------|
| Low risk                 |                    |                      |          |
| Study adheres to commonly held tenets of high-quality design, execution, and avoidance of bias | Good quality RCT | Good quality cohort |          |
|                         |                    |                      |          |
| Moderately low risk      |                    |                      |          |
| Study has potential for some bias, study does not meet all criteria for a good quality RCT, but deficiencies not likely to invalidate results or introduce significant bias | Moderate or poor quality RCT | Moderate quality cohort |          |
|                         |                    |                      |          |
| Moderately high risk     |                    |                      |          |
| Study has significant flaws in design and/or execution that increase potential for bias that may invalidate study results | Poor quality RCT | Poor quality cohort |          |
|                         |                    |                      |          |
| High risk                |                    |                      |          |
| Study has significant potential for bias, lack of comparison group precludes direct assessment of important outcomes | Case series | Poor quality case-control or cross-sectional study |          |

#### Articles on prognosis or risk

| Risk of bias description | Studies of therapy | Studies of prognosis | Criteria |
|--------------------------|--------------------|----------------------|----------|
| Low risk                 |                    |                      |          |
| Study adheres to commonly held tenets of high-quality design, execution, and avoidance of bias | Good quality cohort | Good quality cohort |          |
|                         |                    |                      |          |
| Moderately low risk      |                    |                      |          |
| Study has potential for some bias; does not meet all criteria for good quality cohort, but deficiencies not likely to invalidate results or introduce significant bias | Moderate quality cohort | Moderate quality cohort |          |
|                         |                    |                      |          |
| Moderately 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 cohort | Poor quality case-control or cross-sectional study |          |
|                         |                    |                      |          |
| 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 study | Poor quality case-control or cross-sectional study |          |
|                         |                    |                      |          |

#### 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 determined based on tenets 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 methods used to determine the final SoE.

| Baseline strength: Risk of bias (including control of confounding) is accounted for in the individual article | Outcome | Strength of evidence | Conclusions and comments | Baseline | Downgrade | Upgrade | Criteria |
|----------------------------------------------------------------------------------------------------------------|--------|----------------------|--------------------------|----------|-----------|---------|----------|
| High: High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect | High    | Summary of findings  | High RCTs | No Consistent, direct, and precise estimates | No     |          |          |
| 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 | Moderate| Summary of findings  | Low Cohort studies | No Consistent, direct, and precise estimates | Yes    | Large effect |          |
| 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 | Low     | Summary of findings  | High RCTs | Yes (2) Inconsistent findings | No     |          |          |

#### 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 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 |

---

*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.

*bReliable data are data such as mortality or reoperation.

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

---

*aModerate or poor quality RCT are listed in table.

*bBaseline strength: Risk of bias (including control of confounding) is accounted for in the individual article including tenets 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.

*bModerate or poor quality RCT are listed in table.

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

*bCochrane 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, qualitative, and quantitative analyses for systematic reviews (there is a reference/evaluation table for this).