**Evidence Assessment: Definition of Risk of Bias (RoB)**

**Articles on treatment**

| Risk of bias description | Studies of therapy | Criteria |
|--------------------------|--------------------|---------|
| Low risk                 | Good quality RCT   | • Random sequence generation  
|                          |                    | • Allocation concealment  
|                          |                    | • Intent-to-treat analysis  
|                          |                    | • Blind or independent assess- 
|                          |                    |   ment for author’s primary  
|                          |                    |   important outcomes*  
|                          |                    | • Co-interventions applied equally  
|                          |                    | • F/U rate of 80% - <10%  
|                          |                    | • Difference in follow-up between  
|                          |                    |   groups  
|                          |                    | • Controlling for possible  
|                          |                    |   confounding*b  
| Moderately low risk      | Moderate or poor quality RCT | • Violation of one or two criteria  
|                          |                    | for good quality RCT  
|                          |                    | • Blind or independent assess- 
|                          |                    |   ment in a prospective study, 
|                          |                    |   or use of reliable data in a 
|                          |                    |   retrospective study  
|                          |                    | • Co-interventions applied equally  
|                          |                    | • F/U rate of 80% - <10%  
|                          |                    | • Difference in follow-up between  
|                          |                    |   groups  
|                          |                    | • Controlling for possible  
|                          |                    |   confounding*b  
| Moderately high risk     | Poor quality RCT   | • Violation of three or more of 
|                          |                    |   the criteria for a good quality RCT  
|                          | Moderate or poor quality cohort | • No comparison group  
|                          |                    | • Random sequence generation  
|                          |                    | • Allocation concealment  
|                          |                    | • Intent-to-treat analysis  
|                          |                    | • Blind or independent assess- 
|                          |                    |   ment for author’s primary 
|                          |                    |   important outcomes*  
|                          |                    | • Co-interventions applied equally  
|                          |                    | • F/U rate of 80% - <10%  
|                          |                    | • Difference in follow-up between 
|                          |                    |   groups 
| High risk                | Case control       | • Any case-control design  
|                          |                    | • Random sequence generation  
|                          |                    | • Allocation concealment  
|                          |                    | • Intent-to-treat analysis  
|                          |                    | • Blind or independent assess- 
|                          |                    |   ment for author’s primary  
|                          |                    |   important outcomes*  
|                          |                    | • Co-interventions applied equally 
|                          |                    | • F/U rate of 80% - <10%  
|                          |                    | • Difference in follow-up between 
|                          |                    |   groups  
|                          |                    | • Controlling for possible  
|                          |                    |   confounding*b  

*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.

**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 results outlined by the 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.

The following four possible levels and their definitions will be 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 |
|---------|-----------------------|--------------------------|----------|-----------|---------|
| Outcome | High                  | Summary of findings      | High RCT | No        | Yes     |
|         |                       |                          | Consistent, direct, and precise estimates | No | Upgrade |
| Outcome | Moderate              | Summary of findings      | Low Cohort studies | No | Yes     |
|         |                       |                          | Consistent, direct, and precise estimates | Yes | Large effect |
| Outcome | Low                   | Summary of findings      | High RCT | Yes (2)   | No      |
|         |                       |                          | Inconsistent blinded | Inconsistent blinded | No |

*Additional domains: dose-response, strength of association, publication bias.

Single study - “consistency unknown.”

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

| 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  

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

**Articles on prognosis or risk**

| Risk of bias description | Studies of prognosis | Criteria |
|--------------------------|----------------------|---------|
| Low risk                 | Good quality cohort* | • Prospective design  
|                          |                      | • Patients at similar point in the course of their disease or treatment  
|                          |                      | • F/U rate of ≥ 80%i  
|                          |                      | • Patients followed long enough for outcomes to occur  
|                          |                      | • Accounting for other prognostic factorsb  
|                          |                      | • Objective and unbiased outcome measure usedc  
| 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 the rest of the criteria in good quality cohort  
| Moderately high risk     | Poor quality cohort  | • Prospective design with violation of two or more criteria for good quality cohort study  
|                          |                      | • Retrospective design with violation of one or more criteria for good quality cohort study  
|                          |                      | • A good case-control studyd  
|                          |                      | • A good cross-sectional studyd  
| High risk                | Poor quality cohort  | • 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.

*Examples of objective outcomes: weight loss, change in blood pressure, speed of walking, reoperation, death, etc.; examples of subjective outcomes are patient reported outcomes and self-reported pain.

*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, and 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**

| 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  

*aBroad spectrum of persons with the expected condition.

*bAdequate description of methods for replication.

*cBlinded performance of tests, measurements, or interpretation.

*dSecond test/interpretation performed independently of the first.