Table E-1 Therapy Article Scoring Worksheet[43]

1. Was the study randomized?
2. Were all patients analyzed in the groups to which they were randomized?
3. Were patients in the treatment and control groups similar with respect to known prognostic factors?
4. Other than the intervention, were all patients treated equally?
5. Was there blinding on any level (investigator, patient, outcomes assessor)?
6. Was follow-up greater than 80% in all groups?[a]
7. Does the study report the size of the treatment effect?[b]
8. Does the study report the precision of the treatment effect (confidence intervals)?

SCORING: Yes = 1 No/Uncertain = 0

[a] Follow-up of 80 percent is derived from criteria described by Sackett et al.[43].

[b] Size of treatment effect can be reported as numbers-needed-to-treat (NNT), likelihood ratio (LR), relative risk, odds ratio, absolute risk difference, or survival rates.

<u>Questions excluded from scoring (explanation of why excluded):</u>

Did the experimental and control groups retain a similar prognosis after the study started? (Included in question #3 for the purposes of this investigation.)

Individual questions for blinding of patient, investigator, outcomes assessor. (As double blinding is nearly impossible for most surgical trials, for this investigation, we chose to accept any level of blinding.)

Were all clinically significant outcomes considered? and *Are the likely benefits worth the potential harm and costs?* (Determined to be too subjective for this analysis.)

Table E-2 Exposure/Risk/Harm Article Scoring Worksheet[43]

1. Were there clearly defined groups of patients, similar in all important characteristics except for exposure to the treatment or proposed etiologic agent?

2. Were treatments/exposures measured/administered in the same way in all groups?

3. Were outcomes assessed the same way in all groups?

4. Were outcomes assessed in a blinded fashion in all groups?

5. Was there follow-up sufficiently long to permit the outcome in question to occur in all groups?

6. Was follow-up greater than 80% in all groups?[a]

7. Is it clear that the exposure preceded the onset of the outcome in question?

SCORING: Yes = 1 No/Uncertain = 0

[a] Follow-up of 80 percent is derived from criteria described by Sackett et al.[43].

<u>Questions excluded from scoring (explanation of why excluded):</u>

Is there a dose-response effect? Is there positive evidence from a dechallenge-rechallenge study? Is the temporal relationship correct? Is the association specific (limited to a single putative cause and single effect? (Because these questions are so rarely possible to answer in contemporary orthopaedic research, for the purpose of the present investigation, they were not counted.)

Table E-3 Prognosis/Natural History Article Scoring Worksheet[43]

1. Was a defined, representative sample of patients assembled at a common point in the disease (inception cohort)?
2. Was the referral pattern described (to enable the readers to judge applicability to their practice)?
3. Was greater than 80% of the inception cohort analyzed or otherwise accounted for at the final time point[a]?
4. Were objective outcome criteria used or was a validated outcomes instrument applied to all patients at final follow-up?
5. Were outcome criteria applied in a blinded manner?
6. Are estimates of likelihood/survivorship of the important end point presented (e.g., Kaplan-Meier curves)?
7. Are confidence intervals or other appropriate measurements of uncertainty provided?
8. Was there follow-up sufficiently long to permit the outcome in question to occur in all groups?

SCORING: Yes = 1 No/Uncertain = 0

[a] Follow-up of 80 percent is derived from criteria described by Sackett et al.[43].

<u>Questions excluded from scoring (explanation of why excluded):</u>

Were patients sufficiently homogeneous with respect to prognostic risk? (This question relates to study generalizability and is not a direct measurement of study quality.)

Was there validation in an independent group ("test set") of patients? (This question is included in #4 and relates to the validity of the outcomes assessment.)

Table E-4 Diagnostic/Screening Test Article Scoring Worksheet[43]

1. Was there an independent, blinded comparison with a reference standard ("gold standard")?
2. Was the reference standard applied regardless of the diagnostic/screening test result?
3. Were the methods of performing the test described in sufficient detail to enable replication?
4. Are sufficient data reported to enable calculation of likelihood ratios or are they stated explicitly in the text?

<u>Questions excluded from scoring (explanation of why excluded):</u>

*Did the clinicians face diagnostic uncertainty? (*Frequently not possible to discern; also, this question relates to pretest probability in the context of generalizability more than study quality or internal validity.)

Was the diagnostic/screening test applied to an appropriate spectrum of patients? (Frequently not possible to discern; also, this question relates to study generalizability more than study quality or internal validity.)