

Appendix 2: Evidence Rating Methodology

The evidence table, modeled on the American College of Radiology Appropriateness Criteriaⁱ, summarizes each literature citation in a topic narrative by scoring the number of quality criteria achieved by each citation. We utilize the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system^{ii,iii}, which is particularly well-suited for cross sectional or cohort studies in patients with diagnostic uncertainty, i.e. patients to whom clinicians would apply the diagnostic test in the course of regular clinical practice.

Citation ¹	Objectives ²	# of patients ³	Results ⁴	Study type ⁵	Design ⁶	Quality Score ⁷
				Diagnostic	observational	1: Has all eight study quality criteria.
					experimental	2: Has six to seven quality criteria.
					review	3: Has three to five quality criteria.
					meta-analysis	4: Source is not useful as primary evidence.
				Therapeutic	observational	n/a
					experimental	1: Has five to six study quality criteria.
					review	2: Has three to four quality criteria.
					meta-analysis	3: Has one to two quality criteria.
						4: Source is not useful as primary evidence.
						n/a

1. Citation –Includes title, abbreviated authors list, year of publication, journal/book name.
2. Objective – Lists the main aim(s) of the study.
3. Number of patients – the higher the number, the better the case can be made for a true relationship to disease or therapeutic intervention and to may identify important differences among patient cohorts.
4. Results – List of the main findings.
5. Study type – Citation can either describe diagnostic tools or assess the use of treatments and interventions and have different study quality criteria to assess the amount of bias.
6. Design – Citations can be in one of more of 4 categories. Observational studies can detect relationship but experimental design may better determine causality but both can achieve a *quality score* between 1-3 based on the *criteria* below.
7. Quality score – Can take the values of 1 to 4 based on *criteria* listed below with a lower integer better in quality than a higher number. Review articles do not typically perform statistical measurements and hence can only achieve a maximum score of 4. Meta-analysis studies are not rated since this method is designed only to evaluate individual studies. The *criteria* are:
 - a. Statistical measures such as sensitivity, PPV, ROC analysis, etc. are present and facilitate comparisons across citations.
 - b. Measurements of uncertainty such as p-values, confidence intervals, etc. are present to provide a range for the statistical measurement.
 - c. Timing of the study, e.g. prospective studies designed prior to the data collection tend to reduce bias.
 - d. For diagnostic studies only – comparison with standard method was made, a reference standard has been applied to all subjects in the same way, recruitment was performed systematically, two or more independent readers were employed, and test results were interpreted blind to reference standard results.
 - e. For therapeutic studies only – presence of control and intervention groups, random allocation into these groups and length of follow-up or drop-out factors listed.

ⁱ American College of Radiology. ACR Appropriateness Criteria® Evidence Table Development. Available at: <https://www.acr.org/-/media/ACR/Files/Appropriateness-Criteria/EvidenceTableDevelopment.pdf>. Updated November 2015. Accessed December 26, 2017.

ⁱⁱ The GRADE working group. 2018. Last accessed at www.gradeworkinggroup.org on April 24, 2018.

ⁱⁱⁱ Grading quality of evidence and strength of recommendations for diagnostic tests and strategies BMJ 2008;336:1106.