CDI QUALITY INSTITUTE

Provider Led Entity

Appropriate Use Criteria:
Quality Evaluation for a Body of Evidence

January 26, 2017

The CDI Quality Institute follows the recommended framework defined by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group in evaluating the quality of a body of evidence. Table 1 outlines the quality of evidence grades; table 2, the ADAPTE process for evaluation the quality of guidelines; and table 3, the GRADE criteria used to determine the quality of systemic reviews, meta-analyses and/or a body of literature.

.

Table 1: Quality of Evidence Grades

Strength of Evidence	*Definition*
High	Very confident in the accuracy of the estimate. We believe the findings are stable, i.e. future research is unlikely to change this estimate significantly.
Moderate	Moderate confidence in the accuracy of the estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that future research will change that estimate.
Low	Limited confidence in the accuracy of the estimate. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect. The true effect may be substantially different from the estimate.
Insufficient	We have no evidence or the body of evidence has unacceptable deficiencies. The evidence does not support an estimate of accuracy or outcome.

Table 2: ADAPTE process for evaluating the quality and applicability of specialty society and organizational guidelines

The ADAPTE process uses standardized and transparent tools to assess the quality, validity and transparency of guidelines developed within subspecialty societies and government organizations and to determine applicability for the development of appropriateness utilization criteria. The ADAPE evaluation centers on 6 domains:

Domain	*Definition and Elements*
Scope	1. Overall objectives of the guideline are specifically described. 2. Clinical questions covered by the guideline are specifically described. 3. Patients to whom the guideline is meant to apply are specifically described.
Stakeholder Involvement	1. Guideline development group includes individuals from all the relevant disciplines and/or stakeholders. 2. Patients' views and preferences have been sought. 3. Target users of the guideline are clearly defined. 4. The guideline has been piloted among target users.
Methodology	1. Systematic methods were used to search for evidence. 2. The criteria for selecting the evidence are clearly described. 3. The methods for formulating the recommendations are clearly described. 4. The health benefits, side effects and risks have been considered in formulating the recommendations. 5. There is an explicit link between the recommendations and the supporting evidence. 6. The guideline has been externally reviewed by experts prior to publication. 7. A procedure for updating the guideline is provided.

Presentation and Clarity	1. The recommendations are specific and unambiguous. 2. The different options for management of the condition are clearly presented. 3. Key recommendations are easily identifiable. 4. The guideline is supported with tools for application.
Applicability	1. The potential organizational barriers in applying the guideline have been discussed. 2. The potential cost implications of applying the recommendations have been considered. 3. The guideline presents key review criteria for monitoring and/or auditing purposes.
Conflict of Interests	1. The guideline is editorially independent from the funding body or subspecialty society. 2. Conflicts of interest of guideline development members have been recorded.
Overall Assessment	Overall assessment and recommendation for use in Appropriateness Use Criteria development is based on an overall score of > 90%, and a methodology score of > 50%.

Table 3: GRADE criteria for determining the quality of systemic reviews and a body of published articles

Study design is used to establish the initial quality of evidence. For questions about diagnostic testing and prognosis, well designed cross sectional and cohort studies are considered high quality evidence and can be downgraded according to the criteria below. For studies around therapy or management strategies, randomized control trials start as the highest quality of evidence and can be downgraded using relevant GRADE criteria.

Domain	*Definition and Elements*	*Application to the Evaluation of Diagnostic Test Performance*
Study Design	Cross-sectional or cohort studies in patients with diagnostic uncertainty and direct comparison of test results with an appropriate reference standard are considered high quality studies.	
Risk of Bias	Limitations in study design and execution may bias the estimates of test performance. Internal validity is assessed with attention to the following areas; • Patient selection – Consecutive patients as a single cohort and not classified according to disease state. • Index test – Interpreted and reported on each patient with appropriate blinding to the reference standard. Results of index test reported with estimates of diagnostic uncertainty and reliability. • Reference standard – Appropriate independent reference standard applied with blinding as to the results of the index test.	Using the QUADAS-2 tool, the risk of bias is classified as: • Low, • High or • Unclear. With a high risk of bias consider downgrading the evidence if significant or critical to the estimates of accuracy or outcome.

	• Flow and Timing – Index and reference standard applied at appropriate intervals to every patient and withdrawals from the study explained.	
Inconsistency	Inconsistency refers to unexplained heterogeneity in the test performance. Criteria to determine whether to downgrade for variability can be applied when one or more of the following criteria are met: • Wide variance of estimates across studies, • Minimal or no overlap of confidence intervals, or • Statistical criteria for tests of heterogeneity	Use one of three levels of consistency: • Consistent • Inconsistent • Unknown or not applicable (e.g. single study) Investigators should explore explanations for heterogeneity (study design, patient population or test variability), If it remains unexplained and is significant or critical to the estimates of accuracy or outcome consider downgrading the level of evidence.
Indirectness	Indirectness occurs when there are significant differences between the research and intended guideline population, test or intervention; when the measured outcomes are not directly relevant to patients; or when competing test strategies are not directly compared. Indirectness criteria should be applied to each of the following areas: • The comparability of the patient populations, interventions, equipment, diagnostic expertise and imaging protocols between the research and target settings, • The link between measures of test performance and patient outcomes,	Score dichotomously as one of two levels of directness: • Indirectness present • Not present If indirectness is present and significant or critical to the estimates of accuracy or outcome consider downgrading the level of evidence. If a decision is made to grade the strength of evidence of an intermediate outcome such as diagnostic accuracy, then the reviewer does not need to automatically "downgrade" this outcome for being indirect.

	safety, anxiety and/or management, or • Direct comparison versus indirect comparisons of the index test and next best testing strategy.	
Imprecision	Imprecision is the extent to which our confidence in the estimate of an effect is adequate to support a particular decision. Imprecision exists if there are few events and confidence intervals (CIs) show substantial overlap, particularly if the CI crosses the decision threshold for recommending and not recommending a test.	Score dichotomously as one of two levels of precision: • Precise • Imprecise Consider downgrading for imprecision if significant or critical to the estimates of accuracy or outcome.
Publication Bias	• Publication bias indicates that studies may have been published selectively and results in a systematic under-estimation or over-estimation of the underlying effect. With selective reporting, the estimate of test performance based on published studies may not reflect the true effect. Evidence from small studies of new tests or asymmetry in funnel plots should raise suspicion for publication bias.	Score dichotomously as one of two levels of reporting bias: • Suspected • Undetected Consider downgrading for publication bias if significant or critical to the estimates of accuracy or outcome.

Dose-response gradient	This association, either across or within studies, refers to a pattern of a larger effect with greater exposure (including dose, duration, and adherence). This indicates a putative cause-effect relationship and may increase our confidence in the findings and the quality of evidence. The presence of a dose-response gradient may support an underlying mechanism for detection and relevance of some tests that have continuous outcomes and/or multiple cutoffs [e.g., serum PSA (prostate-specific antigen) levels and ventilation/perfusion scanning].	Score dichotomously as one of two levels of publication bias: • Present • Undetected Consider upgrading the level of evidence when a dose-response gradient is present and significant.
Plausible unmeasured and confounding bias	Occasionally, plausible confounding factors would work in the direction opposite to that of the observed effect. Had these confounders not been present, the observed effect would have been larger. The impact of plausible unmeasured confounders may be relevant to testing strategies that predict outcomes. A study may be biased to find low diagnostic accuracy via spectrum bias yet still show high diagnostic accuracy.	Score dichotomously as one of two levels of publication bias: • Present • Absent If plausible unmeasured confounding factors exist and are significant or critical to the estimates of accuracy or outcome, consider upgrading the level of evidence.
Large Magnitude of Effect	When an observational study shows a large effect, we can be more confident about the results. One is more likely to rate up the quality of evidence when the effect is rapid, the effect is consistent across subjects, previous trajectory of disease is reversed, or the effect is supported by indirect evidence.	Use one of three levels for magnitude of effect: • Strong • Weak • Absent If a large magnitude of effect exists, consider upgrading the level of evidence.

Sources

AHRQ. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=1752

Methods Guide for Effectiveness and Comparative Effectiveness Reviews. (Berkman 2013, Singh 2012).

Schünemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Available at http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.f7lc8w9c3nh8

Schünemann H, Oxman AD, Brozek J, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ 2008;336:1106-1110.