Appendix 2. Study Quality Rating Criteria[1]

DIAGNOSTIC ACCURACY STUDIES

Criteria

Screening test relevant, available for primary care, adequately described?

Study uses a credible reference standard, performed regardless of test results?

Reference standard interpreted independently of screening test?

Handles indeterminate results in a reasonable manner?

Spectrum of patients included in study?

Sample size?

Administration of reliable screening test?

Definition of Ratings Based on Criteria

Good: Evaluates relevant available screening test; uses a credible reference standard; interprets reference standard independently of screening test; reliability of test assessed; has few or handles indeterminate results in a reasonable manner; includes large number (more than 100) broad-spectrum patients with and without disease.

Fair: Evaluates relevant available screening test; uses reasonable, although not best, standard; interprets reference standard independent of screening test; moderate sample size (50 to 100 subjects) and a "medium" spectrum of patients.

Poor: Has important limitation, such as uses inappropriate reference standard; screening test improperly administered; biased ascertainment of reference standard; very small sample size of very narrow selected spectrum of patients.

RANDOMIZED CONTROLLED TRIALS AND COHORT STUDIES

Criteria

Initial assembly of comparable groups: randomized controlled trials (RCTs)—adequate randomization, including concealment and whether potential con-founders were distributed equally among groups; cohort studies—consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts?

Maintenance of comparable groups (includes attrition, crossovers, adherence, contamination)?

Important differential loss to follow-up or overall high loss to follow-up?

Measurements: equal, reliable, and valid (includes masking of outcome assessment)?

Clear definition of interventions?

Important outcomes considered?

Analysis: adjustment for potential confounders for cohort studies, or intention-to-treat analysis for RCTs?

Definition of Ratings Based on Criteria

Good: Meets all criteria: comparable groups are assembled initially and maintained throughout the study (follow-up at least 80%); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; important outcomes are considered; and appropriate attention to confounders in analysis.

Fair: Studies will be graded "fair" if any or all of the following problems occur, without the important limitations noted in the "poor" category below: generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred in follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some, but not all, important outcomes are considered; and some, but not all, potential confounders are accounted for.

Poor: Studies will be graded "poor" if any of the following major limitations exists: groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention.

CASE CONTROL STUDIES

Criteria

Accurate ascertainment of cases?

Nonbiased selection of cases/controls with exclusion criteria applied equally to both?

Response rate?

Diagnostic testing procedures applied equally to each group?

Measurement of exposure accurate and applied equally to each group?

Appropriate attention to potential confounding variable?

Definition of Ratings Based on Criteria

Good: Appropriate ascertainment of cases and nonbiased selection of case and control participants; exclusion criteria applied equally to cases and controls; response rate equal to or greater than 80%; diagnostic procedures and measurements accurate and applied equally to cases and controls; and appropriate attention to confounding variables.

Fair: Recent, relevant, without major apparent selection or diagnostic workup bias but with response rate less than 80% or attention to some but not all important confounding variables.

Poor: Major selection or diagnostic workup biases, response rates less than 50%, or inattention to confounding variables.

Reference

1. Harris RP, Helfund M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* 2001;20:21-35.