
Appendix Table 5. U.S. Preventive Services Task Force Study and Quality Rating Criteria*

Systematic reviews

Quality rating criteria

Comprehensiveness of sources considered or search strategy used

Standard appraisal of included studies

Validity of conclusions

Timeliness and relevance are especially important

Definition of ratings from above criteria

Good: Recent, relevant review with comprehensive sources and search strategies; explicit and relevant selection criteria; standard appraisal of included studies; and valid conclusions

Fair: Recent, relevant review that is not clearly biased but lacks comprehensive sources and search strategies

Poor: Outdated, irrelevant, or biased review without systematic search for studies, explicit selection criteria, or standard appraisal of studies

RCTs and cohort studies

Quality rating criteria

Initial assembly of comparable groups

RCTs: Adequate randomization, including first concealment and whether potential confounders 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, crossover, 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 the interventions

All important outcomes considered

Analysis: Adjustment for potential confounders for cohort studies or intention-to-treat analysis for RCTs

Definition of ratings from above criteria

Good: Meets all criteria: comparable groups are assembled initially and maintained throughout the study (follow-up $\geq 80\%$); reliable and valid measurement instruments are used and are applied equally to the groups; interventions are defined clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention-to-treat analysis is used.

Fair: Any or all of the following problems have occurred, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially, but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and are generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is done for RCTs.

Poor: Any of the following fatal flaws are present: Groups assembled initially are not close to being comparable or maintained throughout the study, unreliable or invalid measurement instruments are used or are not applied at all equally among groups (including not masking outcome assessment), and key confounders are given little or no attention. For RCTs, intention-to-treat analysis is lacking.

* Data are from references 25 and 26. RCT = randomized, controlled trial.