

Healthcare Information & Decision Equation: Information → Decision → Action → Outcome
Is it true → Is it useful → Is it usable?

Rationale for Evidence Grading

Effective critical appraisal requires assessing both validity and usefulness of studies or study results. An evidence grade rates a study or outcome. Higher grades of evidence reflect higher quality which is more likely to report more accurate estimates of effect.

Evidence Grading Systems

It is important to examine the criteria used in the various grading systems because some systems assign misleading quality grades by inflating lower quality or invalid studies.

Delfini Evidence Grading Scale & Strength of Evidence Considerations

Grades can be applied to individual studies, to conclusions within studies, a body of evidence or to secondary sources such as guidelines or clinical recommendations. General advice is provided below. (Due to complexities with studies of diagnostic tests, no recommendations for them are provided here.) All-or-none studies (observational) may be an exception and occur rarely.

Grade A: Useful

The evidence is strong and appears sufficient to use in making health care decisions—it is both valid and useful (e.g., meets standards for clinical significance, sufficient magnitude of effect size, physician and patient acceptability, etc.). Studies achieving this grade should be outstanding in design, methodology, execution and reporting and have successful study performance outcomes, providing useful information to aid clinical decision-making, enabling reasonable certitude in drawing conclusions.

- For a body of evidence: Several well-designed and conducted studies that consistently show similar results.
- For therapy, screening and prevention: RCTs. In some cases a single, large Grade A RCT may be sufficient; however, without confirmation from other studies, results could be due to chance, undetected significant biases, fraud, etc. In such instance, the SOE should include a cautionary note.
- For natural history and prognosis: Cohort studies
- Grade A should be rarely assigned to any study. (“Extra points” are not given for challenge or difficulty in answering the research question. Authors should not be given extra points by second-guessing them. Transparency is required.)

Grade B: Possibly Useful

Grade B studies should be very well designed and executed and meet most of the requirements that it takes to achieve a Grade A. Grade B evidence appears potentially strong and is probably sufficient to use in making health care decisions—some threats to validity have been identified. Studies achieving this grade should be of high quality and contain only non-lethal threats to validity and with sufficiently useful information to aid clinical decision-making, enabling reasonable certitude in drawing conclusions.

- For a body of evidence: The evidence is strong enough to conclude that the results are probably valid and useful (see above); however, study results from multiple studies are inconsistent or the studies may have some (but not lethal) threats to validity.
- For therapy, screening and prevention: RCTs. In some cases a single, large Grade B RCT may be sufficient; however, without confirmation from other studies results could be due to chance, undetected significant biases, fraud, etc. In such instance, the SOE should include a cautionary note.
- For natural history and prognosis: Cohort studies
- Grade B is more frequent than Grade A, but is still a difficult grade to achieve.

Grade B-U: Possible to uncertain usefulness

The evidence might be sufficient to use in making health care decisions; however, there remains sufficient uncertainty that the evidence cannot fully reach a Grade B, and the uncertainty is not great enough to fully warrant a Grade U.

Grade U: Uncertain Validity and/or Usefulness

There is sufficient uncertainty that caution is urged regarding its use in making health care decisions. Grade U should be assigned when there is sufficient uncertainty about the accuracy of the estimates of effect resulting in an inability to comfortably draw conclusions from the research and in comfortably applying results.

- We end up assigning most studies a Grade U. As stated, we generally never use Grade U studies to inform efficacy decisions, but we will use Grade U evidence for safety, being very careful to describe that the evidence is of low quality.

The Agency for Healthcare Research and Quality (AHRQ) has a simple, useful system for grading evidence of individual studies and the overall strength of evidence (SOE) considering all included studies:

- Individual study risk of bias ratings: high risk of bias, medium risk of bias, low risk of bias
- Overall SOE ratings: High, Moderate, Low, Insufficient

Delfini Modifications: Overall level of evidence (LOE) ratings: High, Moderate, Borderline, Inconclusive