Delfini Pearls: Evidence Grading

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.

Exaggeration of Differences Between Outcomes in Intervention vs Control Groups (High Quality vs Low Quality RCTs)

To see study citation and abstract, enter the PubMed Identification (PMID) number in the PubMed search window.

Study Area of Concern	Relative Exaggeration	Reference (PMID)
Inadequate Generation of Randomization Sequence	17% to 75%	Juni (11440947), Kjaergard (11730399), Van Tulder (19770609)
Concealment of Allocation	14% to 73%	Schulz (7823387), Kjaergard (11730399), Moher (9746022), Juni (11440947)
Inadequate Double Blinding	4%% to 72%%	Schulz (7823387), Poolman (17332104), Kjaergard (11730399), Moher (9746022), Juni (11440947)
Loss of Data (Up to 38%)	2% to 35%	Vvan Tulder (19770609), Tierney (15561753), Nüesch (19736281), Canadian Orthopaedic Trauma Society (17200303)
Assessing outcomes through models	50% or greater	Lachin (11018568)

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.

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.)

- For therapy, screening, prevention and diagnostic studies: RCTs. In some cases a single, large well-designed and conducted 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 study might receive a Grade A, but the Strength of the Evidence should include a cautionary note.
- For natural history and prognosis: Cohort studies

Grade B: Possibly Useful

The evidence appears potentially strong and is probably sufficient to use in making health care decisions - some threats to validity were identified

- For therapy, screening, prevention and diagnostic studies: RCTs. In some cases a single, large well-designed and conducted 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 study might receive a Grade A, but the Strength of the Evidence should include a cautionary note.
- Also for diagnosis, valid studies assessing test accuracy for detecting a condition when there is evidence of effectiveness from valid, applicable RCTs.
- For natural history and prognosis: Cohort studies

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.

Study quality is such that it appears likely that the evidence is sufficient to use in making health care decisions; however, there are some study issues that raise continued uncertainty. Health care decision-makers should be fully informed of the evidence quality.

Grade U: Uncertain Validity and/or Usefulness

There is sufficient uncertainty that caution is urged regarding its use in making health care decisions.

- Uncertain Validity: This may be due to uncertain validity due to methodology (enough threats to validity to raise concern – our suggestion would be to **not** use such a study in most circumstances) or may be due to conflicting results.
- Uncertain Usefulness: Or this may be due to uncertain applicability due to results (good methodology, but questions due to effect size, applicability of results when relating to biologic markers, or other issues). These latter studies may be useful and should be viewed in the context of the weight of the evidence.
- Uncertainty of Author: If the author has reached a conclusion that the findings are uncertain, doing a critical appraisal is unlikely to result in a different conclusion. The evidence leaves us uncertain regardless of whether the study is valid or not. Critical appraisal is at the discretion of the reviewer.

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 bia, 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