

Which Risk of Bias Tool Should You Use?

The standard instrument is decided by the study design you are appraising, not by habit. A selector that matches the right tool to each design, with the traps that get reviews rejected. From PhD methodologists.



Written by PhD methodologists researchgold.org

The right tool follows from the design you are appraising

Assessing the quality of included studies is a required step in a systematic review, and reviewers lose credibility fast by using the wrong instrument for the design in front of them. There is no single universal tool. The design of each study, randomised, non-randomised, observational, or diagnostic, points to the instrument built for it.

PRINCIPLE

Match the instrument to the design, then apply it the same way to every study. A randomised trial and an observational cohort are exposed to different biases, so they are appraised with different tools. Using one generic checklist for everything is the most common appraisal mistake, and a reviewer will spot it immediately.

The instruments and what they are for

1 For randomised trials

The Cochrane risk of bias tool, in its second version known as RoB 2, is the standard. It assesses bias across domains such as randomisation, deviations from intended interventions, missing data, measurement, and selective reporting.

2 For non-randomised studies of interventions

ROBINS-I is built for studies where allocation was not randomised. Its central concern is confounding, the bias that randomisation would otherwise have handled.

3 For observational cohort and case-control studies

The Newcastle-Ottawa Scale is widely used to appraise selection, comparability, and outcome or exposure ascertainment in these designs.

4 For reviews of reviews and diagnostic accuracy

AMSTAR 2 appraises the quality of systematic reviews themselves, used in an umbrella review. QUADAS-2 is the standard for studies of diagnostic test accuracy.

Find your design, read across to your tool

Identify the design of the studies you included, then take the instrument on that row. Where designs are mixed, you will need more than one.

STUDY DESIGN TO APPRAISAL TOOL

Study design you included	Standard instrument	What it focuses on
Randomised controlled trial	RoB 2	Bias across five trial-specific domains
Non-randomised intervention study	ROBINS-I	Confounding and selection above all
Cohort or case-control study	Newcastle-Ottawa Scale	Selection, comparability, outcome ascertainment
Systematic review (for an umbrella review)	AMSTAR 2	Methodological quality of the review itself
Diagnostic test accuracy study	QUADAS-2	Patient selection, index test, reference standard, flow
Prevalence or other observational design	A JBI critical appraisal checklist	Design-specific quality criteria

WATCH OUT FOR

Risk of bias is about how a study was conducted, not how it was reported. A beautifully written paper can still be at high risk of bias, and a poorly written one at low risk. Appraise the methods and the conduct, and resist being swayed by the polish of the writing.

PRINCIPLE

Three points that separate a current appraisal from a dated one. First, the Newcastle-Ottawa Scale, though common, is increasingly discouraged: its star count has weak inter-rater reliability and no validated threshold, and Cochrane now points reviewers toward ROBINS-I (or ROBINS-E for exposures) for non-randomised studies. Second, RoB 2 and the ROBINS tools are *not* scored: you answer signalling questions per domain, an algorithm maps the answers to a domain judgement, and the domain judgements roll up to an overall rating. Never sum them into a numeric quality score or use that score as an inclusion cut-off, which is exactly the misuse the older scales invited. Third, RoB 2 asks you to specify the estimand first, the effect of assignment (intention-to-treat) or of adherence (per-protocol), because the bias from deviations is judged against whichever you chose.

How to use the tool so the appraisal is defensible

Choosing the right instrument is half the task. Applying it consistently and transparently is the other half.

- 1 Assess in duplicate**
Have two reviewers appraise each study independently, then reconcile. Bias judgements involve interpretation, and a second reviewer catches what one alone would miss.
- 2 Record the reason, not just the rating**
For each domain, note the evidence behind your judgement. A rating with no supporting reason cannot be checked and will not convince a careful reader.
- 3 Carry the judgement into the synthesis**
Risk of bias should shape interpretation. Consider a sensitivity analysis that excludes high-risk studies, and weigh the certainty of evidence accordingly.
- 4 Present it visually**
A traffic-light figure or summary plot lets readers see the pattern of bias across studies at a glance, and is now expected in a thorough review.

The appraisal readiness checklist

If every box is ticked, your quality appraisal will match the evidence and stand up to peer review.

- ☒ You identified the design of every included study.
- ☒ You chose the instrument built for that design, not a generic checklist.
- ☒ You will appraise each study in duplicate and reconcile disagreements.
- ☒ You will record the evidence behind every domain judgement.
- ☒ You used more than one tool where the included designs are mixed.
- ☒ You reached judgements through the tool's signalling-question algorithm, not a summed numeric quality score.
- ☒ You planned how risk of bias will feed into the synthesis and interpretation.
- ☒ You will present the results in a clear summary figure.

WHEN THE APPRAISAL HAS TO HOLD UP

Have risk of bias assessed correctly across your studies

If you want the right instrument chosen for each design, applied in duplicate, and carried properly into your synthesis and certainty rating, our PhD methodologists run the appraisal with you to a standard that survives peer review. You stay the author and make every judgement call.

Get a quote for your review

researchgold.org . Written by PhD methodologists . We work in writing, on your schedule.