CHECKLIST FOR DIAGNOSTIC TEST ACCURACY STUDIES

Critical Appraisal tools for use in JBI Systematic Reviews

INTRODUCTION

JBI is an international research organisation based in the Faculty of Health and Medical Sciences at the University of Adelaide, South Australia. JBI develops and delivers unique evidence-based information, software, education and training designed to improve healthcare practice and health outcomes. With over 70 Collaborating Entities, servicing over 90 countries, JBI is a recognised global leader in evidence-based healthcare.

JBI Systematic Reviews

The core of evidence synthesis is the systematic review of literature of a particular intervention, condition or issue. The systematic review is essentially an analysis of the available literature (that is, evidence) and a judgment of the effectiveness or otherwise of a practice, involving a series of complex steps. JBI takes a particular view on what counts as evidence and the methods utilised to synthesise those different types of evidence. In line with this broader view of evidence, JBI has developed theories, methodologies and rigorous processes for the critical appraisal and synthesis of these diverse forms of evidence in order to aid in clinical decision-making in healthcare. There now exists JBI guidance for conducting reviews of effectiveness research, qualitative research, prevalence/incidence, etiology/risk, economic evaluations, text/opinion, diagnostic test accuracy, mixed-methods, umbrella reviews and scoping reviews. Further information regarding JBI systematic reviews can be found in the JBI Evidence Synthesis Manual.

JBI Critical Appraisal Tools

All systematic reviews incorporate a process of critique or appraisal of the research evidence. The purpose of this appraisal is to assess the methodological quality of a study and to determine the extent to which a study has addressed the possibility of bias in its design, conduct and analysis. All papers selected for inclusion in the systematic review (that is – those that meet the inclusion criteria described in the protocol) need to be subjected to rigorous appraisal by two critical appraisers. The results of this appraisal can then be used to inform synthesis and interpretation of the results of the study. JBI Critical appraisal tools have been developed by the JBI and collaborators and approved by the JBI Scientific Committee following extensive peer review. Although designed for use in systematic reviews, JBI critical appraisal tools can also be used when creating Critically Appraised Topics (CAT), in journal clubs and as an educational tool.

JBI CRITICAL APPRAISAL CHECKLIST FOR DIAGNOSTIC TEST ACCURACY STUDIES

Reviewer_____Date_____

Author_____Year_____ Record Number_____

	Yes	No	Unclear	Not applicable
1. Was a consecutive or random sample of patients enrolled?	☐	☐	☐	☐
2. Was a case control design avoided?	☐	☐	☐	☐
3. Did the study avoid inappropriate exclusions?	☐	☐	☐	☐
4. Were the index test results interpreted without knowledge of the results of the reference standard?	☐	☐	☐	☐
5. If a threshold was used, was it pre-specified?	☐	☐	☐	☐
6. Is the reference standard likely to correctly classify the target condition?	☐	☐	☐	☐
7. Were the reference standard results interpreted without knowledge of the results of the index test?	☐	☐	☐	☐
8. Was there an appropriate interval between index test and reference standard?	☐	☐	☐	☐
9. Did all patients receive the same reference standard?	☐	☐	☐	☐
10. Were all patients included in the analysis?	☐	☐	☐	☐

Overall appraisal: Include ☐   Exclude ☐   Seek further info ☐

Comments (Including reason for exclusion)

_____
_____

DIAGNOSTIC TEST ACCURACY STUDIES
CRITICAL APPRAISAL TOOL

How to cite: *Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.*

Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, White S, Munn Z. Diagnostic test accuracy: methods for systematic review and meta-analysis. Int J Evid Based Healthc. 2015;13(3):154-62.

Answers: Yes, No, Unclear or Not/Applicable

PATIENT SELECTION

1. **Was a consecutive or random sample of patients enrolled?**

   Studies should state or describe their method of enrolment. If it is claimed that a random sample was chosen the method of randomization should be stated (and appropriate). It is acceptable if studies do not say 'consecutive' but instead describe consecutive enrolment; i.e. 'all patients from …. till …. were included'.

2. **Was a case control design avoided?**

   Case control studies are described in detail in the reviewers manual. In essence, if a study design involves recruiting participants who are already known by other means to have the diagnosis of interest and investigating whether the test of interest correctly identifies them as such, the answer is 'No'.

3. **Did the study avoid inappropriate exclusions?**

   If patients are excluded for reasons that would likely influence the conduct, interpretation or results of the test, this may bias the results. Examples include: excluding patients on which the test is difficult to conduct, excluding patients with borderline results, excluding patients with clear clinical indicators of the diagnosis of interest.

INDEX TEST

4. **Were the index test results interpreted without knowledge of the results of the reference standard?**

   The results of the index test should be interpreted by someone who is blind to the results of the reference test. The reference test may not have been conducted at the point that the index test is carried out, if so the answer to this question will be 'Yes'. If the person who interprets the index test also interpreted the reference test then it is assumed that this question will be answered 'No' unless there are other factors in play (for instance, the interpretation of the results may be separate from their collection, in which case the interpreter may be blinded to patient identity and past reference test results).

5. **If a threshold was used, was it pre-specified?**

Diagnostic thresholds may be chosen based on what gives the optimum accuracy from the data, or they may be pre-specified. When no diagnostic threshold is applied (i.e. the results of a test is based on the observation of a specific characteristic which is either there or not) this question will be answered NA.

6. **Is the reference standard likely to correctly classify the target condition?**

The reference test should be the gold standard for the diagnosis of the condition of interest. Additionally, the reporting of the study should describe its conduct in sufficient detail that the reviewers can be confident that it has been correctly and competently implemented.

7. **Were the reference standard results interpreted without knowledge of the results of the index test?**

The points made for criteria 4 apply equally here. The results of the reference test should be interpreted by someone who is blind to the results of the index test. The index test may not have been conducted at the point that the reference test is carried out, if so the answer to this question will be 'Yes'. If the person who interprets the reference test also interpreted the index test then it is assumed that this question will be answered 'No' unless there are other factors in play (for instance, the interpretation of the results may be separate from their collection, in which case the interpreter may be blinded to patient identity and past index test results).

8. **Was there an appropriate interval between index test and reference standard?**

The index test and the reference test should be carried out close enough together that the status of the patient could not have meaningfully changed. The maximum acceptable time will vary based on characteristics of the population and condition of interest.

9. **Did all patients receive the same reference standard?**

The reference standard by which patients are classed as having or not having the condition of interest should be the same for all patients. If the results of the index test influence how or whether the reference test is used (i.e. where an apparent false negative may be detected the study design may call for a 'double check') this may result in biased estimates of sensitivity and specificity. Additionally, in some studies two parallel reference tests may be used (on different patients) and the results then pooled. In either case the results should be 'No'.

10. **Were all patients included in the analysis?**

Loses to follow up should be explained and there cause and frequency should be considered in whether they are likely to have had an effect on the results (Subjectivity may exist in this context, overall low tolerance should be applied in deciding to answer 'No' to this question, but a single withdrawal from a large cohort should not necessarily force a negative response). However, if a

patient's results being difficult to interpret causes their data to be excluded from the analysis this will exaggerate the estimate of DTA, and this question should definitely be answered 'No' .