Assessment of methodological quality and QUADAS-2

Yemisi Takwoingi
October 2015
Based on slides developed by Mariska Leeflang, Penny Whiting, Hans Reitsma and Sue Mallett
Learning objectives

• To be familiar with QUADAS-2

• To understand the need to adapt QUADAS-2 for the review

• To appreciate presenting and incorporating results of quality assessment
Outline

• Introduction to quality assessment of test accuracy studies

• QUADAS-2 tool

• Presenting and incorporating quality assessment results
Why assess quality?

- Problem 1: Bias in primary studies can lead to misleading summary estimates of accuracy
- Problem 2: The studies may not be applicable to the review question
- Problem 3: Results of primary studies may vary

Quality assessment to identify potential risk of bias and applicability to the review question and to guide interpretation of results
What do we mean by quality?

“both the risk of bias and applicability of a study;
(1) the degree to which estimates of diagnostic accuracy avoid risk of bias, and
(2) the extent to which primary studies are applicable to the review’s research question”
Sources of bias and variation

• **Bias**: flaws in the design of the study may result in invalid estimates of accuracy

• **Applicability**: variation across studies mean that the results may not be applicable to your review question
QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies

Penny F. Whiting, PhD; Anne W.S. Rutjes, PhD; Marle E. Westwood, PhD; Susan Mallett, PhD; Jonathan J. Deeks, PhD; Johannes B. Reitsma, MD, PhD; Mariska M.G. Leeﬂang, PhD; Jonathan A.C. Sterne, PhD; Patrick M.M. Bossuyt, PhD; and the QUADAS-2 Group*

In 2003, the QUADAS tool for systematic reviews of diagnostic accuracy studies was developed. Experience, anecdotal reports, and feedback suggested areas for improvement; therefore, QUADAS-2 was developed. This tool comprises 4 domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias, and the first 3 domains are also assessed in terms of concerns regarding applicability. Signalling questions are included to help judge risk of bias.

The QUADAS-2 tool is applied in 4 phases: summarize the review question, tailor the tool and produce review-speciﬁc guidance, construct a ﬂow diagram for the primary study, and judge bias and applicability. This tool will allow for more transparent rating of bias and applicability of primary diagnostic accuracy studies.

For author afﬁliations, see end of text.
* For members of the QUADAS-2 Group, see the Appendix (available at www.annals.org).
QUADAS-2

Three phased tool:

- Phase 1: Define the review question
- Phase 2: Draw a flow diagram
- Phase 3: Assessment of risk of bias and applicability

  ▪ Four domains:
    » Patient selection
    » Index test
    » Reference standard
    » Flow and timing
Phase 1: Define the question

Define the question:

<table>
<thead>
<tr>
<th>Patients:</th>
<th>Patients with joint symptoms &lt;12 months duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index test:</td>
<td>Second generation anti-CCP test analysed by ELISA</td>
</tr>
<tr>
<td>Comparator test (if applicable):</td>
<td>Rheumatoid factor detected by latex agglutination</td>
</tr>
<tr>
<td>Target condition:</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Reference Standard:</td>
<td>American College of Rheumatology (ACR) criteria</td>
</tr>
</tbody>
</table>
Phase 2: Flow diagram

467 Consecutive patients

Anti-CCP2: 467

Anti-CCP2+: 95
- RA: 82
- Other: 6
- Unclear: 7

Anti-CCP2−: 372
- Unclear: 100
- Other: 200
- RA: 71
Phase 3: Assessment of domains

• Description:
  – Record information used to make judgement

• Signalling questions:
  – Used to inform rating of bias
  – Items phrased so that yes indicates absence of bias

• Domains:
  – Judgement of risk of bias/applicability made
  – A “No” on one or more signalling questions means that risk of bias should be considered, it does not necessarily imply that the risk should be judged as “High”

• Background document:
  – Provides general guidance on scoring
  – Scoring guidance, specific to your review should be produced to help ensure objective and consistent rating
## QUADAS-2 domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Source of Bias</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient selection</td>
<td>Study design</td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td>Prospective/ retrospective</td>
<td>Previous tests</td>
</tr>
<tr>
<td></td>
<td>Sample selection</td>
<td>Presentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intended use of test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Setting</td>
</tr>
<tr>
<td>Index test</td>
<td>Blinding</td>
<td>Test technology</td>
</tr>
<tr>
<td></td>
<td>Threshold specification</td>
<td>Test execution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpretation setting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpreter expertise</td>
</tr>
<tr>
<td>Reference standard</td>
<td>Verification procedure</td>
<td>Reference standard</td>
</tr>
<tr>
<td></td>
<td>Independence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blinding</td>
<td></td>
</tr>
<tr>
<td>Patient Flow and Timing</td>
<td>Appropriate time interval</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Intervention between tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Verification bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriate exclusions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appropriate inclusions</td>
<td></td>
</tr>
</tbody>
</table>
Signalling questions

**DOMAIN 1: PATIENT SELECTION**

**A. Risk of Bias**

Describe methods of patient selection:

- Was a consecutive or random sample of patients enrolled?  
  [Yes/No/Unclear]
- Was a case-control design avoided?  
  [Yes/No/Unclear]
- Did the study avoid inappropriate exclusions?  
  [Yes/No/Unclear]
- Could the selection of patients have introduced bias?  
  RISK: LOW/HIGH/UNCLEAR

**B. Concerns regarding applicability**

Describe included patients (prior testing, presentation, intended use of index test and setting):

- Is there concern that the included patients do not match the review question?  
  CONCERN: LOW/HIGH/UNCLEAR

---

**DOMAIN 3: REFERENCE STANDARD**

**A. Risk of Bias**

Describe the reference standard and how it was conducted and interpreted:

- Is the reference standard likely to correctly classify the target condition?  
  [Yes/No/Unclear]
- Were the reference standard results interpreted without knowledge of the results of the index test?  
  [Yes/No/Unclear]
- Could the reference standard, its conduct, or its interpretation have introduced bias?  
  RISK: LOW/HIGH/UNCLEAR

**B. Concerns regarding applicability**

Is there concern that the target condition as defined by the reference standard does not match the review question?  
CONCERN: LOW/HIGH/UNCLEAR

---

**DOMAIN 2: INDEX TEST(S)**

If more than one index test was used, please complete for each test.

**A. Risk of Bias**

Describe the index test and how it was conducted and interpreted:

- Were the index test results interpreted without knowledge of the results of the reference standard?  
  [Yes/No/Unclear]
- If a threshold was used, was it pre-specified?  
  [Yes/No/Unclear]
- Could the conduct or interpretation of the index test have introduced bias?  
  RISK: LOW/HIGH/UNCLEAR

**B. Concerns regarding applicability**

Is there concern that the index test, its conduct, or interpretation differ from the review question?  
CONCERN: LOW/HIGH/UNCLEAR

---

**DOMAIN 4: FLOW AND TIMING**

**A. Risk of Bias**

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- Was there an appropriate interval between index test(s) and reference standard?  
  [Yes/No/Unclear]
- Did all patients receive a reference standard?  
  [Yes/No/Unclear]
- Did patients receive the same reference standard?  
  [Yes/No/Unclear]
- Were all patients included in the analysis?  
  [Yes/No/Unclear]
- Could the patient flow have introduced bias?  
  RISK: LOW/HIGH/UNCLEAR
Application of QUADAS-2

Off the shelf

Bespoke
Tailoring QUADAS-2 to your review

1. TAILOR TOOL CONTENT
   Consider adding/omitting signalling questions

2. DEVELOP SCORING GUIDELINES
   Produce clear guidelines for your review

3. PILOT TOOL AND GUIDELINES
   Apply QUADAS-2 in small number of studies

   GOOD AGREEMENT

4. APPLY TO ALL INCLUDED STUDIES
   Complete the QUADAS-2 assessment for all studies

   POOR AGREEMENT
   Refine tool content and/or guidelines

Focus on your review question
DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:

- Was a consecutive or random sample of patients enrolled?  Yes/No/Unclear
- Was a case-control design avoided?  Yes/No/Unclear
- Did the study avoid inappropriate exclusions?  Yes/No/Unclear

Could the selection of patients have introduced bias?  RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting):

Is there concern that the included patients do not match the review question?  CONCERN: LOW/HIGH/UNCLEAR
Patient selection - applicability

- Measures of accuracy may vary across patient groups:
  - Advanced versus early disease
  - Symptoms
  - Setting
  - Prior testing
  - Presence of alternative conditions
  - Demographic features
  - Intended use of the test
DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

- Were the index test results interpreted without knowledge of the results of the reference standard?  
  Yes/No/Unclear
- If a threshold was used, was it pre-specified?  
  Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias?  
RISK: LOW / HIGH / UNCLEAR

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question?  
CONCERN: LOW / HIGH / UNCLEAR
Index test

Bias: Blinding

- Knowledge of reference standard results when interpreting index test may lead to over-optimistic estimates of accuracy
- Less important for objective tests or if index test is interpreted prior to reference standard
Bias: Threshold selection

– Selecting threshold to maximise sensitivity and/or specificity may lead to overoptimistic measures of test performance.

– Performance of this cut-off in an independent set of patients will be likely to be lower, even if the study consists of patients from the same population.

– Important that the threshold is pre-specified rather than derived from the results of the study.
Index test

Applicability:

– If test conduct, technology, setting or interpretation differ from your review question the results may not be applicable
DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

- Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW/HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question? CONCERN: LOW/HIGH/UNCLEAR
Reference standard

Bias: Reference standard

– How do we know the truth?
  ▪ Post-mortem, histology, radiology, microbiology, chemical pathology

– Calculating accuracy assumes reference standard is 100% accurate – any disagreements assume index test is incorrect
Reference standard

Bias: Blinding

- Reference standard should be interpreted blind to index test results
- Related to degree of subjectiveness and order of tests.
- Index test should not form part of the reference standard

• Example:
  - ACR criteria for RA (reference standard) are applied some time after the anti-CCP test (Index test) and could therefore be influenced by knowledge of the test results
Reference standard

Applicability:

– Outcome of reference standard is decisive: if the reference standard does not detect the target condition defined in the review question results may not be applicable
– Critical to choose valid/optimal reference standard

Example:

– When defining UTI the reference standard is generally based on specimen culture but the threshold above which a result is considered positive may vary
### A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- Was there an appropriate interval between index test(s) and reference standard? [Yes/No/Unclear]
- Did all patients receive a reference standard? [Yes/No/Unclear]
- Did patients receive the same reference standard? [Yes/No/Unclear]
- Were all patients included in the analysis? [Yes/No/Unclear]

**Could the patient flow have introduced bias?**  RISK: LOW / HIGH / UNCLEAR
Flow and timing

Bias: Timing

– Delay between tests can cause misclassification due to recovery or progression to more advanced disease

– Length of time which may cause such bias will vary between conditions

Example

– For the evaluation of MRI for the early diagnosis of MS, a minimum follow-up period of around 10 years is required, for infectious disease e.g. UTI a delay of a few days may be important
Flow and timing

**Bias: Work-up/verification Bias**

- The reference standard may be expensive, risky or unpleasant – clinically unwilling to perform on “normals”
- If only cases who are test positive undergo reference standard may result in misclassifying false negatives as true negatives and will overestimate sensitivity and specificity.
- Alternative methods such as extended follow-up or random sample of test negatives may be appropriate

**Example**

- D-dimer test for diagnosis of PE, where ventilation perfusion scans are used (ref standard 1) in those testing positive and clinical follow-up in those testing negative (ref standard 2). Follow-up may miss some cases of PE and overestimate accuracy.
Incorporating study quality

- What do existing reviews do?
Incorporating study quality

• Present the results of the quality assessment:
  – In a table
Figure 5. 'Risk of bias' and applicability concerns summary: review authors’ judgements about each domain for each included study

<table>
<thead>
<tr>
<th></th>
<th>Patient Selection</th>
<th>Index Test</th>
<th>Reference Standard</th>
<th>Flow and Timing</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angiolli 2005</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Brun 2008</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Brun 2009</td>
<td>?</td>
<td>?</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Deffieux 2006</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Fagotti 2004</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Fagotti 2008</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Vergote 1998</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Incorporating study quality

• Present the results of the quality assessment:
  – In a table
  – Graphically
Figure 3. Risk of bias and applicability concerns graph: review authors’ judgements about each domain presented as percentages across included studies.

Citation: Taylor T, Dineen RA, Gardiner DC, Buss CH, Howatson A, Pace NL. Computed tomography (CT) angiography for confirmation of the clinical diagnosis of brain death. Cochrane Database of Systematic Reviews 2014, Issue 3. Art. No.: CD009694.
What do existing reviews do?

<table>
<thead>
<tr>
<th>Approach</th>
<th>Overall quality of included studies</th>
<th>Number N = 65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality mentioned in the main text</td>
<td></td>
<td>60 (92%)³</td>
</tr>
<tr>
<td>Results of quality assessment reported, no mention in discussion or conclusion</td>
<td></td>
<td>13 (20%)</td>
</tr>
<tr>
<td>Results of quality assessment reported and discussed, but quality not linked to conclusion</td>
<td></td>
<td>41 (63%)</td>
</tr>
<tr>
<td>Results of quality assessment reported and discussed, and conclusions regarding test accuracy linked to results of quality assessment</td>
<td></td>
<td>6 (9%)</td>
</tr>
<tr>
<td>Results of quality assessment reported and discussed, and recommendations based on general unspecified quality items</td>
<td></td>
<td>12 (18%)</td>
</tr>
</tbody>
</table>

³Quality was mentioned in one or more sections in the main text.

Ochodo et al. BMC Medical Research Methodology 2014, 14:33
Incorporating study quality

- Present the results of the quality assessment:
  - In a table
  - Graphically
- Investigate individual quality items as potential sources of heterogeneity
- Basis for recommendations for future research
Investigation of heterogeneity

• Sensitivity analysis to see whether the overall estimate changes when excluding specific studies

• Stratified analysis according to presence/absence of specific quality criteria

• Analysis using meta-regression

• Define methodological criteria a priori
Problems with quality assessment

• Not as straightforward as it might sound!
• Hampered by poor reporting
• Quality assessment is subjective
• Statistical incorporation of quality problematic with limited studies
Take home message

• Quality assessment is essential

• The QUADAS-2 tool is recommended by Cochrane

• The specific items and scoring guidelines should be tailored to your review question

• The results of the quality assessment should be presented

• No quality scores and cut-offs for ‘good’ quality

• Study quality should be incorporated into all reviews
Website

- [www.quadas.org](http://www.quadas.org)
ACKNOWLEDGEMENTS

Materials for this presentation are based in part on material adapted from members of the Cochrane Screening and Diagnostic Test Methods Group

See [http://dta.cochrane.org/dta-author-training-online-learning](http://dta.cochrane.org/dta-author-training-online-learning) for additional training materials