ebhc CONFERENCE 2026

THE ECOSYSTEM OF EVIDENCE

Navigating the future of the ecosystem of evidence

Artificial intelligence and digital transformation in research and healthcare

11th EBHC International Conference
Taormina, 21st – 24th October 2026

#EBHC2026

QUADAS-3

Updated tool to evaluate risk of bias and applicability concerns in diagnostic test accuracy studies

Penny Whiting, on behalf of the QUADAS-3 Team

Eve Tomlinson, Anne Rutjes, Clare Davenport, Bada Yang, Jude Holmes, Mariska Leeflang, Sue Mallett and the QUADAS-3 Steering Group



KEY TOOL

LATITUDES-network.org

GIMBE EVIDENCE FOR HEALTH

University of BRISTOL · UNICAMILLUS International Medical University in Rome · UNIVERSITY OF BIRMINGHAM · Amsterdam UMC University Medical Centers · UMC Utrecht · UCL

Background

- QUADAS - 2 (2011) was designed to evaluate the **risk of bias** and **applicability** of test accuracy studies
- QUADAS-2 is widely used (cited >12, 000 times) and recommended by major HTA organizations and Cochrane
- Feedback on QUADAS-2 has generally been positive, but some signaling questions have been identified as problematic and the tool could be improved based on features included in more recently developed tools

Aims

- Introduce you to QUADAS-3
- Briefly outline methods to develop the tool



Development Steps



Step 1

Identification of evidence



Step 2

Face-to-face meeting of Steering Group

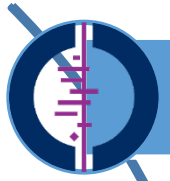


Step 3

Pilot and refine

Step 1: Evidence





A. Modifications to QUADAS-2 for the Cochrane Handbook



B. Developments from more recent risk-of-bias tools



C. Review of methodological studies that have evaluated QUADAS-2



D. Review of published systematic reviews that have used QUADAS-2



E. Web-based survey of review authors who have used QUADAS-2



F. Review of guidance from HTA organizations

Step 2: F2F Meeting



Meeting in London, September 2023 to discuss:

- Findings from step 1 (evidence review)
- Scope of QUADAS-3
- Applicability
- Feasibility of algorithms for arriving at domain-level ratings
- Small group work on draft tool

Step 2: Pilot and Refine



- V0.4 • Revised version of tool created and shared with Steering Group for feedback
- V0.5 • Consideration of incorporation of QUADAS-C
- V0.6 • Piloting of tool on five papers by the Steering Group
- V0.8 • Piloting of tool at Global Evidence Summit 2024
- V0.11 • Piloting in a series of think aloud interviews
- V0.13 • Piloting in “real” systematic reviews
- V0.14 • Final tool shared with steering group for comments
- V1.0 • Final version agreed by steering group

QUADAS-3

6 Phases:

REVIEW	1 Synthesis question	2 Ideal test accuracy trial
STUDY	3 Flow diagram	4 Select accuracy estimates
ESTIMATE	5 Domain judgments	6 Overall judgment

Phase 1: Synthesis question(s)



- A single review may address more than one synthesis question.
- The first phase of the QUADAS-3 assessment is to specify with sufficient detail the synthesis question(s) that you will answer with your review

Question 1	
Question 2 (if applicable)	

Phase 2: Define the ideal test accuracy trial

Component
Objective
Participants
Index Test(s)
Definition of the condition
Analysis

Table 3: Framework to define the ideal test accuracy trial. Use this information to help you complete Table 4

Component	General features
Overall	• Pre-specified protocol. • Sufficient total sample size to expect adequate numbers of participants with and without the target condition to generate sufficiently precise estimates of test accuracy (e.g. sensitivity and specificity).
Objective	• Assess the accuracy of the index test(s) in its proposed role and position in the clinical pathway, including the clinical decision that the test is informing.
Participants	• Prospective enrolment of a single group of consecutive or random group of eligible participants suspected of having the target condition. • Participants are a representative sample of the intended-use population, as defined in the synthesis question. • Details of symptoms and prior test results of included patients reported.
Index test	• Index test should be the appropriate technical version of the index test, conducted using recommended instructions and in the intended role in the clinical pathway. • Performed after recruitment of trial participants. • Interpretation of the index test result should not be affected by external information that would not usually be available to the person interpreting the test in its proposed role in practice e.g. information on the reference standard, disease state, or additional test results. • Test conducted by operators who would usually conduct and interpret the test in practice, in a setting representative of the review question. • Definition of threshold(s) to classify test results as positive or negative pre-specified.
Definition of the target condition	• Reference standard and reference standard threshold should correctly classify all participants as having or not having the target condition. • Protocol for undergoing the reference standard is the same for all study participants, and the reference standard is performed in all participants. • Reference standard performed after recruitment of trial participants. • Reference standard interpretation is independent of the index test; those interpreting the reference standard should not be aware of the index test results, and the index test should not be part of the reference standard where a composite reference standard is used. • The time between the index test and reference standard is sufficiently short to be confident that the disease status has not changed.
Analysis	• All participants included in the analysis. • Unit of analysis appropriate for the synthesis question. • Appropriate methods used to calculate estimates of accuracy (e.g. sensitivity and specificity).

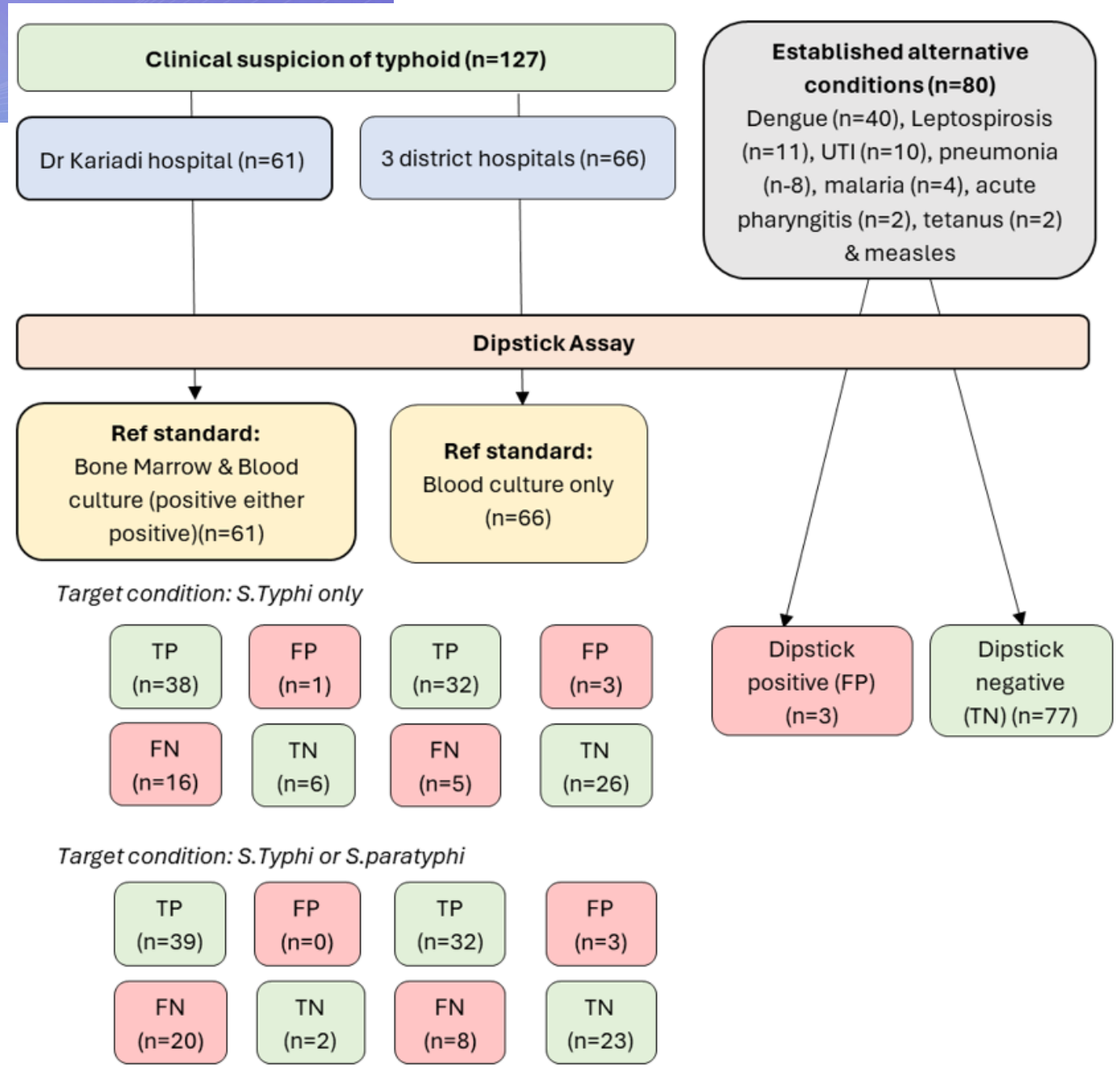
The “ideal” trial to address each synthesis question

Deviations from this design have potential to introduce bias or affect applicability

Forms formal part of applicability assessment



Phase 3: Flow Diagram



Phase 4: Identify accuracy estimate

	Estimate 1	**Estimate 2**	**Estimate 3**
Synthesis Question (e.g. Q1, Q2 from table X)			
Numerical test accuracy result e.g. sensitivity=80% and specificity=76%			
Participants			
Index Test			
Index test threshold (if applicable)			
Target condition			
Reference standard			
Unit of analysis (e.g. participant, tumour, lesion, sample)			
Analysis (e.g. analysis method, participants included in analysis)			
Which QUADAS-3 domains will be assessed (Participants, Index Test, Target condition, Analysis)*			

Phase 5: Domain level assessment



Participants



Index Test



Target Condition



Analysis



Phase 5: Domain Structure - Participants

Clarification on handling drop-outs

More structured support for judgement sections

Changes to signalling questions

Change to answers

Change to statement

Addition of rationale for judgement

Clear link to test accuracy trial

Rationale for judgement

A. Risk of bias	
Note: If participants dropped out or were excluded from a study because they did not receive the index test and/or reference standard, then this should be [illegible] main.	
*[illegible]re enrolled, e.g. **single-gate** (a group of part[illegible] not yet known. This could include multiple [illegible], e.g. different hospitals) or **multi-gate** ([illegible] known diagnosis):*	
Describe whether enrolment was prospective or retrospective:	
Describe how participants were selected for enrolment (consecutive or random):	
Describe any restrictions to enrolment:	
[illegible] used?	Y/PY/PN/N/NI
[illegible]pectively enrolled?	Y/PY/PN/N/NI
[illegible]ve or random sample of participants included?	
1.4 Is the study group a re[illegible]esentative sample of the intended-use population?	
Risk that the selection of participants has introduced bias	
Rationale for judgment	
B. Concerns regarding applicability to the systematic review synthesis question	
Concern that the included participants do not match those in the ideal test accuracy trial	
Rationale for judgment *Describe how differences between the included participants (e.g. presentation, prior testing, setting, intended use of index test) and the ideal [illegible] trial defined in Table 4 have led to this judgment*	

Phase 5: Index Test

A. Risk of Bias	
Describe the index test: Describe whether the index test was conducted and interpreted according to recommended instructions (e.g. following manufacturers instructions, following protocol, standard clinical methods): Describe what information was available to those interpreting the index test: If a threshold was used, describe how this was selected (e.g. standard pre-specified threshold, defined a priori within the study, derived from the data):	
2.1 Was the index test conducted and interpreted according to the recommended instructions?	Y/PY/PN/N/NI
2.2 Were the index test results interpreted without knowledge of the reference standard results?	Y/PY/PN/N/NI
2.3 Were the index test results interpreted with the same information as would be available when the test is used in practice?	Y/PY/PN/N/NI
2.4 If an index test threshold was used, was it either standard or pre-specified?	Y/PY/PN/N/NI
Risk that the conduct or interpretation of the index test has introduced bias	LOW/HIGH/II
Rationale for judgement	
B. Concerns regarding applicability to the systematic review synthesis question	
Concern that the index test, its conduct, or interpretation does not match the ideal test accuracy trial	LOW/HIGH/II
Rationale for judgement *Describe how differences between the setting in which the test was conducted and who conducted and interpreted the test and the ideal test accuracy trial defined in Table 4 have led to this judgement*	

Phase 5: Target condition

A. Risk of bias	
Define the target condition: *Describe how the target condition was assessed – what reference standard(s) was used and how was it conducted and interpreted:* *What was the time interval between index test and reference standard:*	
3.1 Does the reference standard adequately identify those with and without the target condition?	Y/PY/PN/N/NI
3.2 Was the target condition assessed in all participants?	Y/PY/PN/N/NI
3.3 Was the target condition assessed in the same way in all participants?	Y/PY/PN/N/NI
3.4 Did the reference standard avoid incorporating the index test?	Y/PY/PN/N/NI
3.5 Was the reference standard conducted and interpreted according to the recommended instructions?	Y/PY/PN/N/NI
3.6 Were the reference standard results interpreted without knowledge of the index test results?	Y/PY/PN/N/NI
3.7 If a reference standard threshold was used, was it standard or pre-specified?	Y/PY/PN/N/NI
3.8 Was there an appropriate time interval between index test and reference standard?	Y/PY/PN/N/NI
Risk that the assessment of the target condition has introduced bias	**LOW/HIGH/II**
Rationale for judgment	
B. Concerns regarding applicability to the systematic review synthesis question	
Concern that the target condition as defined by the reference standard does not match the ideal test accuracy trial	**LOW/HIGH/II**
Rationale for judgment *Describe how any differences between how the target condition was defined and the ideal test accuracy trial defined in Table 4 have led to this judgment*	

Phase 5: Analysis

A. Risk of bias	
Describe any participants who were enrolled in the study but excluded from the 2x2 table (refer to flow diagram) and reasons why they were excluded (e.g. did not receive index test, did not receive reference standard, uninterpretable index result etc.). *How was any missing 2x2 data, including missing reference standard data, accounted for?* *What do the numbers in the 2x2 table correspond to? That is, what is the unit of analysis (e.g. participants, samples, specimens)?* *Describe any non-standard statistical analyses used to generate the estimates of sensitivity and specificity being assessed (either conducted by the study authors or by you):*	
4.1 Were all participants included in the analysis?	Y/PY/PN/N/NI
4.2 Were missing data handled appropriately?	Y/PY/PN/N/NI
4.3 Does the unit of analysis match the ideal test accuracy trial?	Y/PY/PN/N/NI
4.4 Were the estimates of sensitivity and specificity calculated appropriately?	Y/PY/PN/N/NI
Risk that the analysis has introduced bias	**LOW/HIGH/II**
Rationale for judgment	

Phase 6: Overall Judgement

Risk of bias

Domain	Judgement	Rationale for judgement
1: PARTICIPANTS	LOW/HIGH/II	
2: INDEX TEST(S)	LOW/HIGH/II	
3: TARGET CONDITION	LOW/HIGH/II	
4: ANALYSIS	LOW/HIGH/II	
OVERALL	LOW/HIGH/II	

Concerns regarding applicability

Domain	Judgement	Rationale for judgement
1: PARTICIPANTS	LOW/HIGH/II	
2: INDEX TEST(S)	LOW/HIGH/II	
3: TARGET CONDITION	LOW/HIGH/II	
OVERALL	LOW/HIGH/II	

Overview of main changes

Major changes

- Assessment at the estimate (Phase 4)
- Introduction of the ideal test accuracy trial to aid with the assessment of applicability and to focus the risk-of-bias assessments (Phase 2)
- Formal section for overall judgement (Phase 6)

Changes to domains and signalling questions

- “Patient selection” renamed “Participants”
- “Reference standard” renamed “Target condition”
- “Flow and timing” domain replaced by “Analysis”
- Addition of new signalling questions and rewording of more challenging questions
- Answers to signalling questions – addition of “probably” options (PY and PN)
- Domain level judgements – replace “unclear” with insufficient information”
- Introduction of rationale for judgement
- Focus of applicability judgements on synthesis question and ideal test accuracy trial

Changes considered but not introduced

- Use of algorithms
- Integration of QUADAS-C
- Move from signalling questions to statements

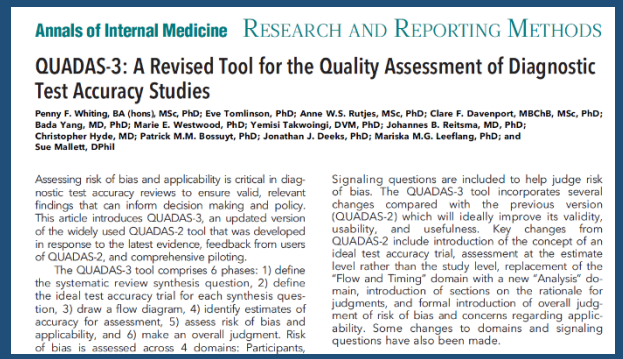
Further Information



Website

Tool, E&E

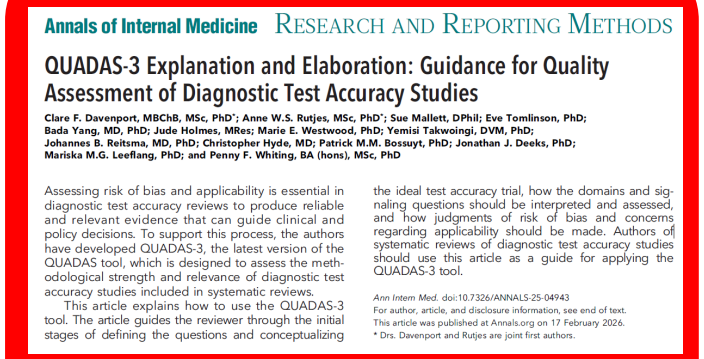
Worked Examples, Training (in development)



QUADAS-3 paper

Published in

Annals of Internal Medicine



QUADAS-3 E&E

Published in

Annals of Internal Medicine

Acknowledgement

Sue Mallett*

Patrick M.M. Bossuyt, Mariska M.G. Leeflang*

Clare Davenport*, Jonathan J. Deeks, Yemisi Takwoingi;

Eve Tomlinson*, Penny F. Whiting*; Marie E. Westwood.

Anne W.S. Rutjes*

Johannes B. Reitsma, Bada Yang*

Chris Hyde

*Indicates core group member

KEY TOOL

LATITUDES-network.org