

## Appendix 1 Keywords and the datasets search

DATASETS	KEYWORDS	RESULTS
MIDLINE/PUBMED	((("osteoarthritis"[MeSH Terms] OR "Knee osteoarthritis"[All Fields]) AND "MRI"[All Fields] AND "Knee"[All Fields]) OR "magnetic resonance imaging"[MeSH Terms]) AND ("sensitivity and specificity"[MeSH Terms] AND ("sensitivity"[All Fields] AND "specificity"[All Fields]) OR "sensitivity and specificity"[All Fields]) AND (("diagnosis"[MeSH Terms] OR "diagnosis"[All Fields] OR "diagnostic"[All Fields]) AND accuracy[All Fields]) AND ("open access"[filter] AND medline[sb])	5,428
SCOPUS	(ALL (knee AND mri) OR ALL (knee AND magnetic AND resonance AND imaging) AND ABS (sensitivity AND specificity) AND ALL (osteoarthritis AND knee))	744
WEB OF SCIENCE	Osteoarthritis (Topic) AND Knee (All Fields) AND Magnetic Resonance Imaging (All Fields) AND MRI (All Fields) AND Sensitivity and Specificity (All Fields) OR knee osteoarthritis diagnosis accuracy (Title)	134
SPORT-DISCUS CINAHL	"osteoarthritis" AND "Knee" AND "Magnetic Resonance Imaging" AND "MRI" AND "Sensitivity and Specificity"	750
GOOGLE SCHOLAR	MRI Knee AND OA diagnosis AND MRI accuracy Knee OR Osteoarthritis OR diagnosis OR MRI OR accuracy "systematic review and meta-analysis"	1,530

## Appendix 2 List of articles excluded after full-text reading 70

## Exclusion reasons:

- Not meet our criteria or not relevant to our topic: 67
- Not available: 2
- Duplication: 1

## Articles excluded:

Exclusion reason; Not meeting our inclusion criteria:

Title	DOI	Exclusion Reason
Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis	10.1016/j.joca.2017.11.018	Not meeting our inclusion criteria
Longitudinal MRI-defined cartilage loss and radiographic joint space narrowing following intra-articular corticosteroid injection for knee osteoarthritis	10.1016/j.ostima.2023.100157	Not meeting our inclusion criteria
Prevalence of knee osteoarthritis features on magnetic resonance imaging in asymptomatic uninjured adults	10.1136/bjsports-2018-099257	Not meeting our inclusion criteria
Association of Sports Participation With Osteoarthritis	10.1007/s11999-010-1321-z	Not meeting our inclusion criteria
MRI T2 and T1ρ relaxation in patients at risk for knee osteoarthritis	10.1186/s12891-019-2547-7	Not meeting our inclusion criteria
Diagnostic Accuracy of Ultrasound for Assessment of Synovial Abnormalities Among Patients With Knee Pain	10.1002/acr.25205	Not meeting our inclusion criteria
Diagnostic accuracy of grayscale power Doppler and contrast-enhanced ultrasound compared with contrast-enhanced MRI in the visualization of synovitis	10.1016/j.ejrad.2020.109392	Not meeting our inclusion criteria
Risk factors for knee osteoarthritis after traumatic knee injury: a systematic review and meta-analysis of randomised controlled trials and cohort studies for the OPTIKNEE Consensus	10.1136/bjsports-2022-105496	Not meeting our inclusion criteria
The compartmental distribution of knee osteoarthritis – a systematic review and meta-analysis	10.1016/j.joca.2020.10.011	Not meeting our inclusion criteria
Running and Knee Osteoarthritis: A Systematic Review and Meta-analysis	10.1177/03635465166575	Not meeting our inclusion criteria
The prevalence of patellofemoral osteoarthritis: a systematic review and meta-analysis	10.1016/j.joca.2016.05.011	Not meeting our inclusion criteria
The prevalence of radiographic and MRI-defined patellofemoral osteoarthritis and structural pathology: a systematic review and meta-analysis	10.1136/bjsports-2017-097515	Not meeting our inclusion criteria
Diagnostic accuracy of ultrasonography MRI and MR arthrography in the characterization of rotator cuff disorders: a systematic review and meta-analysis	10.1136/bjsports-2014-094148	Not meeting our inclusion criteria
Arthroscopy vs. MRI for a detailed assessment of cartilage disease in osteoarthritis: diagnostic value of MRI in clinical practice	10.1186/1471-2474-11-75	Not meeting our inclusion criteria
Association between quadriceps fat pad edema and patellofemoral osteoarthritis: a quantitative Q-Dixon-based magnetic resonance imaging (MRI) analysis	10.21037/qims-23-1730	Not meeting our inclusion criteria
The role of SPECT/CT in painful noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review	10.1186/s13018-023-03687-8	Not meeting our inclusion criteria
Comparison of 1.5- and 3.0-T magnetic resonance imaging for evaluating lesions of the knee	10.1097/MD.00000000000012401	Not meeting our inclusion criteria
Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence	10.1016/j.joca.2010.10.030	Not meeting our inclusion criteria
The use of radiomic analysis of magnetic resonance imaging findings in predicting features of early osteoarthritis of the knee—a systematic review and meta-analysis	10.1007/s11845-024-03714-5	Not meeting our inclusion criteria
The relation between the biochemical composition of knee articular cartilage and quantitative MRI: a systematic review and meta-analysis	10.1016/j.joca.2021.10.016	Not meeting our inclusion criteria
Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis	10.1016/j.joca.2014.11.019	Not meeting our inclusion criteria
Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis	10.1016/j.joca.2009.08.010	Not meeting our inclusion criteria
Effects of long-term exercise therapy on knee joint structure in people with knee osteoarthritis: A systematic review and meta-analysis	10.1016/j.semarthrit.2018.10.014	Not meeting our inclusion criteria
Knee osteoarthritis risk is increased 4-6 fold after knee injury—a systematic review and meta-analysis	10.1136/bjsports-2018-100022	Not meeting our inclusion criteria
Radiographic severity of knee osteoarthritis and its relationship to outcome post total knee arthroplasty: a systematic review	10.1111/ans.15847	Not meeting our inclusion criteria
Knee joint distraction as treatment for osteoarthritis results in clinical and structural benefit: a systematic review and meta-analysis of the limited number of studies and patients available	10.1177/19476035211014997	Not meeting our inclusion criteria
Exercise for osteoarthritis of the knee: a Cochrane systematic review	10.1136/bjsports-2015-095424	Not meeting our inclusion criteria
Risk factors for falls in adults with knee osteoarthritis: a systematic review	10.1002/pmrj.12157	Not meeting our inclusion criteria
Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies	10.1002/art.23188	Not meeting our inclusion criteria
Relationship Between Knee Biomechanics and Pain in People With Knee Osteoarthritis: A Systematic Review and Meta-Analysis	10.1002/acr.24975	Not meeting our inclusion criteria
Fatty infiltration in the thigh muscles in knee osteoarthritis: a systematic review and meta-analysis	10.1007/s00296-019-04368-y	Not meeting our inclusion criteria
Occupational risk in knee osteoarthritis: a systematic review and meta-analysis of observational studies	10.1002/acr.24190	Not meeting our inclusion criteria
Knee extensor strength and risk of structural symptomatic and functional decline in knee osteoarthritis: a systematic review and meta-analysis	10.1002/acr.22940	Not meeting our inclusion criteria
Impact of exercise on articular cartilage in people at risk of or with established knee osteoarthritis: a systematic review of randomised controlled trials	10.1136/bjsports-2018-100056	Not meeting our inclusion criteria
Diagnostic accuracy of dual-energy CT for bone marrow edema in patients with acute knee injury: a systematic review and meta-analysis	10.1186/s13018-023-04151-3	Excluded: Not relevant to our topic
Knee meniscal retears after repair: A systematic review comparing diagnostic imaging modalities	10.1080/19932820.2022.2030024	Not meeting our inclusion criteria
MRI as Diagnostic Modality for Analyzing the Problematic Knee Arthroplasty: A Systematic Review	10.1002/jmri.26874	Excluded: Not relevant to our topic
Ramp lesions: a systematic review of MRI diagnostic accuracy and treatment efficacy	10.1186/s40634-020-00287-x	Excluded: Not relevant to our topic
Diagnostic accuracy of administrative data algorithms in the diagnosis of osteoarthritis: a systematic review	10.1186/s12911-016-0319-y	Not meeting our inclusion criteria
The diagnostic accuracy of clinical tests for anterior cruciate ligament tears are comparable but the Lachman test has been previously overestimated: a systematic review and meta-analysis	10.1007/s00167-022-06898-4	Not meeting our inclusion criteria
This systematic review aimed to assess the advantages of biosensors in detecting biomarkers for the early diagnosis of osteoarthritis (OA)	10.3390/bios11020031	Not meeting our inclusion criteria

Diagnostic Accuracy of Clinical Tests Assessing Ligamentous Injury of the Talocrural and Subtalar Joints: A Systematic Review With Meta-Analysis	10.1177/19417381211029953	Excluded: Not relevant to our topic
Accuracy of clinical tests in the diagnosis of anterior cruciate ligament injury: a systematic review	10.1186/s12998-014-0025-8	Not meeting our inclusion criteria
Three-Dimensional Fast Spin-Echo Imaging without Fat Suppression of the Knee: Diagnostic Accuracy Comparison to Fat-Suppressed Imaging on 1.5T MRI	10.3349/ymj.2017.58.6.1186	Excluded: Not relevant to our topic
Knee meniscal retears after repair: A systematic review comparing diagnostic imaging modalities	10.1080/19932820.2022.2030024	Excluded: Not relevant to our topic
A Systematic Summary of Systematic Reviews on the Topic of the Anterior Cruciate Ligament	10.1177/2325967116634074	Excluded: Not relevant to our topic
Clinical assessment of effusion in knee osteoarthritis—A systematic review	10.1016/j.semarthrit.2015.10.004	Not meeting our inclusion criteria
Classification Systems for Knee Osteochondritis Dissecans: A Systematic Review	10.1177/19476035221121789	Excluded: Not relevant to our topic
Diagnostic accuracy of administrative data algorithms in the diagnosis of osteoarthritis: a systematic review	10.1186/s12911-016-0319-y	Excluded: Not relevant to our topic
The Clinical Utility and Diagnostic Performance of Magnetic Resonance Imaging for Identification of Early and Advanced Knee Osteoarthritis: A Systematic Review	10.1177/0363546511407612	Excluded: Not relevant to our topic
Definition of osteoarthritis on MRI: results of a Delphi exercise	10.1016/j.joca.2011.04.017	Excluded: Not relevant to our topic
MRI as Diagnostic Modality for Analyzing the Problematic Knee Arthroplasty: A Systematic Review	10.1002/jmri.26874	Excluded: Not relevant to our topic
Systematic review of imaging tests to predict the development of rheumatoid arthritis in people with unclassified arthritis	10.1016/j.semarthrit.2021.10.003	Excluded: Not relevant to our topic
Defining and predicting radiographic knee osteoarthritis progression: a systematic review of findings from the osteoarthritis initiative	10.1007/s00167-021-06768-5	Excluded: Not relevant to our topic
Diagnostic accuracy of physical examination for anterior knee instability: a systematic review	10.1007/s00167-015-3563-2	Excluded: Not relevant to our topic
Diagnostic Test Accuracy of Physical Examination Tests in Suspected Patellofemoral Osteoarthritis: A Systematic Review	10.6890/IJGE	Excluded: Not relevant to our topic
Clinical assessment of effusion in knee osteoarthritis-A systematic review	10.1016/j.semarthrit.2015.10.004	Not meeting our inclusion criteria
Can MRI visualise mechanical knee load in osteoarthritic knees? A systematic review with meta-analysis	doi.org/10.1016/j.physio.2015.03.279	Excluded: Not relevant to our topic
Magnetic resonance tomography of the knee joint	10.1007/s00256-015-2178-5	Not meeting our inclusion criteria
MRI T2 and T1ρ relaxation in patients at risk for knee osteoarthritis: a systematic review and meta-analysis	doi: 10.1186/s12891-019-2547-7	Not meeting our inclusion criteria
Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review	DOI: 10.1136/ard.2010.131904	Not meeting our inclusion criteria
Which patellofemoral joint imaging features are associated with patellofemoral pain? Systematic review and meta-analysis	DOI: 10.1016/j.joca.2015.09.004	Not meeting our inclusion criteria
Are contrast-enhanced and non-contrast MRI findings reflecting synovial inflammation in knee osteoarthritis: a meta-analysis of observational studies	DOI: 10.1016/j.joca.2019.10.008	Not meeting our inclusion criteria
Diagnostic performance of three-dimensional MRI for depicting cartilage defects in the knee: A meta-analysis	DOI: 10.1148/radiol.2018180426	Not meeting our inclusion criteria
Sensitivity of Magnetic Resonance Imaging for Detection of Patellofemoral Articular Cartilage Defects	10.1016/j.arthro.2012.03.018	Not meeting our inclusion criteria
Magnetic resonance imaging is able to detect patellofemoral focal cartilage injuries: a systematic review with meta-analysis	10.1007/s00167-022-07203-z	Not meeting our inclusion criteria
Clinical value of MRI in assessing the stability of osteochondritis dissecans lesions: a systematic review and meta-analysis	DOI: 10.2214/AJR.18.20710	Not meeting our inclusion criteria
Quantitative magnetic resonance imaging had higher sensitivity in diagnosing chondral lesions of the knee: a systematic review and meta-analysis	DOI: 10.1016/j.arthro.2024.01.035	Not Available
Accuracy of magnetic resonance imaging in grading knee chondral defects	doi.org/10.1016/j.arthro.2012.04.138	Not Available
Diagnostic Accuracy of Magnetic Resonance Images and Weight-Bearing Radiographs in Patients With Arthroscopic-Proven Medial Osteoarthritis of the Knee	doi.org/10.1177/11795441209383	

**Appendix 3 Clarifications regarding the comparison between the AMSTAR 2 and modified checklist tool for systematic reviews assessment, through 16 items, with a focus on methodological rigor and potential biases**

Original AMSTAR 2 Statements and Scoring	Modified AMSTAR 2 Items	Implement and *Critical Items Adjustment
<p><b>1. Did the research questions and inclusion criteria for the review include the components of PICO?</b></p> <p>Options: Yes, No</p> <p><b>For Yes:</b></p> <ul style="list-style-type: none"> <li>• Population</li> <li>• Intervention</li> <li>• Comparator group</li> <li>• Outcome</li> </ul> <p>Optional (recommended)</p> <ul style="list-style-type: none"> <li>• Timeframe for follow-up</li> <li>• Yes</li> <li>• No</li> </ul>	<p><b>1. Did the research questions and inclusion criteria for the systematic reviews of diagnostic test accuracy include relevant question components?</b></p> <p>Options: Yes, Partial Yes, No</p> <p>Yes: Must include all the following:</p> <ul style="list-style-type: none"> <li>• Detailed description of the index test, including how it is performed and what it measures.</li> <li>• Clearly specify the reference standard against which the index test is compared, providing details on its application and rationale for its use.</li> </ul> <p>Partial Yes: Includes some, but not all, of the above components:</p> <ul style="list-style-type: none"> <li>• Describes the index test being evaluated but lacks detail.</li> <li>• Specifies the reference standard used but without comprehensive details.</li> </ul> <p>No: Does not include most or any of the above components.</p>	<p><b>Research Questions and Inclusion Criteria:</b> High: Comprehensive clear diagnostic index test reference standard and accuracy measures. Moderate: Missing some details but overall adequate. Low/Critically Low: Lacking key components or poorly described.</p> <p>* The modified version specifies the inclusion of “relevant diagnostic components” for systematic reviews of diagnostic test accuracy instead of the general PICO components.</p>
<p><b>2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</b></p> <p>Options: Yes, Partial Yes, No</p> <p>For Partial Yes:</p> <p>The authors state that they had a written protocol or guide that included ALL the following:</p> <ul style="list-style-type: none"> <li>• review question(s)</li> <li>• a meta-analysis/synthesis plan, if appropriate, and</li> <li>• a search strategy</li> <li>• inclusion/exclusion criteria</li> <li>• a risk of bias assessment</li> </ul> <p>For Yes:</p> <p>As for partial yes, plus the protocol should be registered and should also have specified:</p> <ul style="list-style-type: none"> <li>• a plan for investigating causes of heterogeneity: Yes</li> </ul>	<p><b>2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</b></p> <p>Options: Yes, Partial Yes, No</p> <p>For Partial Yes:</p> <p>The authors state that they had a written protocol or guide that included ALL the following:</p> <ul style="list-style-type: none"> <li>• review question(s)</li> <li>• a meta-analysis/synthesis plan, if appropriate, and</li> <li>• a search strategy</li> <li>• inclusion/exclusion criteria</li> <li>• a risk of bias assessment</li> </ul> <p>For Yes:</p> <p>As for partial yes, plus the protocol should be registered and should also have specified:</p> <ul style="list-style-type: none"> <li>• a plan for investigating causes of heterogeneity: Yes</li> </ul>	<p><b>Review Protocol:</b> High: Protocol registered with meta-analysis/synthesis plan and plan for investigating heterogeneity. Partial Yes: Written protocol including review questions, search strategy, inclusion/exclusion criteria, and risk of bias assessment. Low/Critically Low: No protocol or major elements missing.</p>
<p><b>3. Did the review authors explain their selection of the study designs for inclusion in the review?</b></p> <p>Options: Yes, No</p> <p>For Yes, the review should satisfy ONE of the following:</p> <ul style="list-style-type: none"> <li>• Explanation for including only RCTs: Yes</li> <li>• OR Explanation for including only NRSI</li> <li>• OR Explanation for including both RCTs and NRSI</li> </ul>	<p><b>3. Did the review authors explain their selection of the study designs for inclusion in the review?</b></p> <p>Options: Yes, No</p> <p>For Yes, the review should satisfy ONE of the following:</p> <ul style="list-style-type: none"> <li>• Describe how the reference standard providing adequate classification and justified the use of the reference standard which potentially relevant study outcomes.</li> <li>• Consider multiple index tests, provide adequate analysis and or classification.</li> <li>• Provide an explanation for including or only one type of different study design</li> </ul>	<p><b>Study Design Selection:</b> High: Clear rationale for study design choices. Moderate: Partially explained. Low/Critically Low: Poorly explained or unjustified.</p> <p>*The criteria focusing on the index tests, and on the providing an adequate analysis and or classification.</p>
<p><b>4. Did the review authors use a comprehensive literature search strategy?</b></p> <p>Options: Yes, No</p> <p>For Partial Yes (all the following):</p> <ul style="list-style-type: none"> <li>• searched at least 2 databases (relevant to research question)</li> <li>• provided key word and/or search strategy</li> <li>• justified publication restrictions (e.g. language)</li> </ul> <p>For Yes, should also have (all the following):</p> <ul style="list-style-type: none"> <li>• searched the reference lists / bibliographies of included studies: Yes</li> <li>• searched trial/study registries</li> <li>• included/consulted content experts in the field</li> <li>• where relevant, searched for grey literature</li> <li>• conducted search within 24 months of completion of the review</li> </ul>	<p><b>4. Did the review authors use a comprehensive literature search strategy?</b></p> <p>Options: Yes, Partial Yes, No</p> <p>Yes: Should also have (all the following):</p> <ul style="list-style-type: none"> <li>• Searched the reference lists/bibliographies of included studies.</li> <li>• Searched trial/study registries.</li> <li>• Included/consulted content experts in the field.</li> <li>• Where relevant, searched for grey literature.</li> </ul> <p>Partial Yes (all the following):</p> <ul style="list-style-type: none"> <li>• Searched at least two databases (relevant to the research question).</li> <li>• Provided keyword and/or search strategy.</li> <li>• Justified publication restrictions (study design, language).</li> </ul>	<p><b>Literature Search Strategy:</b> High: Comprehensive search including databases registries grey literature. Moderate: Missing some search components. Low/Critically Low: Incomplete or poorly conducted search.</p>
<p><b>5. Did the review authors perform study selection in duplicate?</b></p> <p>Options: Yes, No</p> <p>For Yes, either ONE of the following:</p> <ul style="list-style-type: none"> <li>• at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include: Yes</li> <li>• OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer.</li> </ul>	<p><b>5. Did the review authors perform study selection in duplicate?</b></p> <p>Options: Yes, No</p> <p>For Yes, either ONE of the following:</p> <ul style="list-style-type: none"> <li>• at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include: Yes</li> <li>• OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer.</li> </ul>	<p><b>Study Selection:</b> High: Performed in duplicate with high agreement. Moderate: Partial duplication or lower agreement. Low/Critically Low: Single reviewer or poor agreement.</p>
<p><b>6. Did the review authors perform data extraction in duplicate?</b></p> <p>Options: Yes, No</p> <p>For Yes, either ONE of the following:</p> <ul style="list-style-type: none"> <li>• at least two reviewers achieved consensus on which data to extract from included studies: Yes</li> <li>• OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer.</li> </ul>	<p><b>6. Did the review authors perform data extraction in duplicate?</b></p> <p>Options: Yes, No</p> <p>For Yes, either ONE of the following:</p> <ul style="list-style-type: none"> <li>• at least two reviewers achieved consensus on which data to extract from included studies: Yes</li> <li>• OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer.</li> </ul>	<p><b>Study Data Extraction:</b> High: Performed in duplicate with high agreement. Moderate: Partial duplication or lower agreement. Low/Critically Low: Single reviewer or poor agreement.</p>
<p><b>7. Did the review authors provide a list of excluded studies and justify the exclusions?</b></p> <p>Options: Yes, No</p> <p>For Partial Yes:</p> <ul style="list-style-type: none"> <li>• provided a list of all potentially relevant studies that were read in full-text form but excluded from the review</li> </ul> <p>For Yes, must also have:</p> <ul style="list-style-type: none"> <li>• Justified the exclusion from the review of each potentially relevant study: Yes</li> </ul>	<p><b>7. Did the review authors provide a list of excluded studies and justify the exclusions?</b></p> <p>Options: Yes,</p> <p>Partial Yes, No</p> <p>For Partial Yes:</p> <ul style="list-style-type: none"> <li>• provided a list of all potentially relevant studies that were read in full-text form but excluded from the review</li> </ul> <p>For Yes, must also have:</p> <ul style="list-style-type: none"> <li>• Justified the exclusion from the review of each potentially relevant study: Yes</li> </ul>	<p><b>Excluded Studies:</b> High: List provided with justification. Moderate: List provided without full justification. Low/Critically Low: No list or justification.</p>

<p><b>8. Did the review authors describe the included studies in adequate detail?</b></p> <p>Options: Yes, No  For Partial Yes (ALL the following):</p> <ul style="list-style-type: none"> <li>described populations</li> <li>described interventions</li> <li>described comparators</li> <li>described outcomes</li> <li>described research designs</li> </ul> <p>For Yes, should also have ALL the following:</p> <ul style="list-style-type: none"> <li>described population in detail: Yes</li> <li>described intervention in detail (including doses where relevant)</li> <li>described comparator in detail (including doses where relevant)</li> <li>described study's setting</li> <li>timeframe for follow-up</li> </ul>	<p><b>8. Did the review authors describe the included studies in adequate details related to diagnostic components?</b></p> <p>Options: Yes, Partial Yes, No  Yes: Must meet all the following ideal criteria:  Consecutive recruitment of eligible patients.  Index test measurement parameter and cut-off reported.</p> <ul style="list-style-type: none"> <li>Measurement parameter and cut-off accepted standard.</li> <li>Blinding between index test and reference standard.</li> <li>All those having index test had reference standard.</li> </ul> <p>Partial Yes:</p> <ul style="list-style-type: none"> <li>Non-consecutive recruitment of eligible patients.</li> <li>Measurement parameter and cut-off partially reported.</li> <li>Measurement parameter and cut-off partially accepted as standard.</li> <li>Not blind between index test and reference standard.</li> <li>Reference standard based on test result.</li> </ul>	<p>Study Description: High: Detailed descriptions with all key criteria met. Moderate: Some descriptions missing or partially reported. Low/Critically Low: Inadequate descriptions.</p> <p>*The modified version emphasizes the need for details related to diagnostic components.</p>
<p><b>9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) of individual studies that were included in the review?</b></p> <p>Options: Yes, No  RCTs:</p> <ul style="list-style-type: none"> <li>For Partial Yes, must have assessed RoB from: <ul style="list-style-type: none"> <li>unconcealed allocation, and</li> <li>lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality)</li> </ul> </li> <li>For Yes, must also have assessed RoB from: <ul style="list-style-type: none"> <li>allocation sequence that was not truly random, and</li> <li>selection of the reported result from among multiple measurements or analyses of a specified outcome: Yes</li> </ul> </li> </ul> <p>NRSI:</p> <ul style="list-style-type: none"> <li>For Partial Yes, must have assessed RoB: <ul style="list-style-type: none"> <li>from confounding, and</li> <li>from selection bias</li> </ul> </li> <li>For Yes, must also have assessed RoB: <ul style="list-style-type: none"> <li>methods used to ascertain exposures and outcomes, and</li> <li>selection of the reported result from among multiple measurements or analyses of a specified outcome: Yes</li> </ul> </li> </ul>	<p><b>9. Did the review authors assess the risk of bias (RoB) using QUADAS-2 for the included studies?</b></p> <p>Options:  Yes  Partial Yes  No</p> <p>Yes</p> <ul style="list-style-type: none"> <li>The review must have comprehensively assessed all the QUADAS-2 domains for each included study:</li> </ul> <ul style="list-style-type: none"> <li>Patient Selection: Fully describes the methods of patient selection and confirms whether a consecutive or random sample of patients was enrolled.</li> <li>Index Test: Provides detailed information on the execution of the index test, including its conduct and interpretation.</li> <li>Reference Standard: Thoroughly explains the reference standard, including its conduct and interpretation.</li> <li>Flow and Timing: Fully describes the flow of patients through the study and the timing of the index test(s) and reference standard(s).</li> </ul> <p>Partial Yes</p> <ul style="list-style-type: none"> <li>The review must have assessed some, but not all, of the QUADAS-2 domains:</li> </ul> <ul style="list-style-type: none"> <li>Patient Selection: Partially describes the methods of patient selection but lacks complete details.</li> <li>Index Test: Provides some information on the execution of the index test, but not comprehensively.</li> <li>Reference Standard: Partially explains the reference standard, with some details missing.</li> <li>Flow and Timing: Describes the flow of patients and timing to some extent, but not fully detailed.</li> </ul>	<p>Risk of Bias Assessment: High: Comprehensive assessment using all QUADAS-2 domains. Moderate: Partial assessment. Low/Critically Low: Incomplete or missing assessment.</p> <p>*The modified version specifies the use of the QUADAS-2 domains tool for assessing risk of bias.</p>
<p><b>10. Did the review authors report on the sources of funding for the studies included in the review?</b></p> <p>Options: Yes, No  For Yes:</p> <ul style="list-style-type: none"> <li>Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information, but it was not reported by study authors also qualifies: Yes</li> </ul>	<p><b>10. Did the review authors report on the sources of funding for the studies included in the review?</b></p> <p>Options: Yes, No  For Yes:</p> <ul style="list-style-type: none"> <li>Must have reported on the sources of funding for individual studies included in the review. Note: Reporting that the reviewers looked for this information, but it was not reported by study authors also qualifies: Yes</li> </ul>	<p>Sources of Funding: High: All sources reported. Low/Critically Low: Sources not reported.</p>
<p><b>11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?</b></p> <p>Options: Yes, No, No meta-analysis conducted  RCTs:</p> <p>For Yes:</p> <ul style="list-style-type: none"> <li>The authors justified combining the data in a meta-analysis: Yes</li> <li>AND they used an appropriate weighted technique to combine study results and adjusted for heterogeneity if present.</li> <li>AND investigated the causes of any heterogeneity</li> </ul> <p>For NRSI:</p> <p>For Yes:</p> <ul style="list-style-type: none"> <li>The authors justified combining the data in a meta-analysis: Yes</li> <li>AND they used an appropriate weighted technique to combine study results, adjusting for heterogeneity if present</li> <li>AND they statistically combined effect estimates from NRSI that were adjusted for confounding, rather than combining raw data, or justified combining raw data when adjusted effect estimates were not available</li> <li>AND they reported separate summary estimates for RCTs and NRSI separately when both were included in the review</li> </ul>	<p><b>11. Did the review authors use appropriate methods for the statistical combination of results (meta-analysis) if performed?</b></p> <p>Options: Yes, Partial Yes, No  Partial Yes:</p> <ul style="list-style-type: none"> <li>Used meta-analysis techniques appropriate for diagnostic accuracy data.</li> <li>Defines some, but not all, of the accuracy measures, or provides incomplete calculations.</li> </ul> <p>Yes: Must also have:</p> <ul style="list-style-type: none"> <li>Addressed heterogeneity and potential biases through subgroup analyses or meta-regression.</li> <li>Define the accuracy measures, such as sensitivity, specificity, confidence intervals (CI), positive predictive value (PPV), and negative predictive value (NPV).</li> <li>Definition and calculation of each accuracy measure used in the review.</li> </ul>	<p>Statistical Combination of Results (Meta-analysis): High: Appropriate methods addressing heterogeneity. Moderate: Adequate methods partial heterogeneity discussion. Low/Critically Low: Inappropriate methods no heterogeneity discussion.</p> <p>*We provide detailed considerations for addressing heterogeneity in meta-analysis.  * The modified version specifies the inclusion of "relevant diagnostic components and measures " for systematic reviews of diagnostic test accuracy".</p>
<p><b>12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?</b></p> <p>Options: Yes, No, No meta-analysis conducted  For Yes:</p> <ul style="list-style-type: none"> <li>Included only low risk of bias RCTs: Yes</li> <li>OR, if the pooled estimate was based on RCTs and/or NRSI at variable RoB, the authors performed analyses to investigate possible impact of RoB on summary estimates of effect.</li> <li>No meta-analysis conducted</li> </ul>	<p><b>12. If meta-analysis was performed did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?</b></p> <p>Options: Yes, No, No meta-analysis conducted</p> <p>For Yes:</p> <ul style="list-style-type: none"> <li>Included only low risk of bias: Yes</li> <li>Or the authors performed analyses to investigate possible impact of RoB on summary estimates of effect.</li> </ul> <p>No meta-analysis conducted</p>	<p>Impact of Risk of Bias on Meta-analysis: High: Assessed and discussed. Low/Critically Low: Not assessed or discussed.</p>

<p><b>13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?</b></p> <p>Options: Yes, No For Yes:</p> <ul style="list-style-type: none"> <li>• included only low risk of bias RCTs: Yes</li> <li>• OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results</li> </ul>	<p><b>13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?</b></p> <p>Options: Yes, No</p> <p>Included study provide adequate analysis and only low risk of bias Article: Yes OR, if article with moderate or high RoB, provided a discussion of the likely impact of RoB on the results</p>	<p>Interpretation of Results Considering Risk of Bias: High: Thorough discussion. Low/Critically Low: Lacking discussion.</p> <p>*We provide detailed interpreting/discussing about the included study designs in the way provided good comparison related to the likely impact of RoB on the results.</p>
<p><b>14. Did the review authors provide a satisfactory explanation for and discussion of any heterogeneity observed in the results of the review?</b></p> <p>Options: Yes, No For Yes:</p> <ul style="list-style-type: none"> <li>• There was no significant heterogeneity in the results: Yes</li> <li>• OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review</li> </ul>	<p><b>14. Did the review authors provide a satisfactory explanation for and discussion of any heterogeneity observed in the results of the review?</b></p> <p>Options: Yes, No For Yes:</p> <ul style="list-style-type: none"> <li>• There was no significant heterogeneity in the results: Yes</li> <li>• OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review</li> </ul>	<p>Explanation of Heterogeneity: High: Comprehensive explanation. Moderate: Partial explanation. Low/Critically Low: Poor or no explanation.</p>
<p><b>15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?</b></p> <p>Options: Yes, No, No meta-analysis conducted For Yes:</p> <ul style="list-style-type: none"> <li>• performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias: Yes</li> <li>• No meta-analysis conducted</li> </ul>	<p><b>15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?</b></p> <p>Options: Yes, No, No meta-analysis conducted For Yes:</p> <ul style="list-style-type: none"> <li>• performed graphical or statistical tests for publication bias and discussed the likelihood and magnitude of impact of publication bias: Yes</li> <li>• No meta-analysis conducted</li> </ul>	<p>Publication Bias Investigation: High: Conducted and discussed. Low/Critically Low: Not conducted.</p>
<p><b>16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?</b></p> <p>Options: Yes, No For Yes:</p> <ul style="list-style-type: none"> <li>• The authors reported no competing interests OR</li> <li>• The authors described their funding sources and how they managed potential conflicts of interest: Yes</li> </ul>	<p><b>16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?</b></p> <p>Options: Yes, No For Yes:</p> <ul style="list-style-type: none"> <li>• The authors reported no competing interests OR</li> <li>• The authors described their funding sources and how they managed potential conflicts of interest: Yes</li> </ul>	<p>Conflict of Interest Reporting: High: Fully reported. Low/Critically Low: Not reported.</p>

1. Rogozińska, E. & Khan, K. Grading evidence from test accuracy studies: what makes it challenging compared with the grading of effectiveness studies? *Evidence Based Medicine* 22, 81–84 (2017).

**Note:** The adjustment and solutions were adopted following the addressed challenges for judicious estimation of the strength of test accuracy evidence to develop the guideline

### Modified Checklist Items

#### 1. Research Questions and Inclusion Criteria:

- **High:** Comprehensive, clear diagnostic index test, reference standard, and accuracy measures.
- **Moderate:** Missing some details but overall adequate.
- **Low/Critically Low:** Lacking key components or poorly described.

#### 2. Review Protocol:

- **High:** Protocol registered with meta-analysis/synthesis plan and plan for investigating heterogeneity.
- **Moderate:** Written protocol including review questions, search strategy, inclusion/exclusion criteria, and risk of bias assessment.
- **Low/Critically Low:** No protocol or major elements missing.

#### 3. Study Design Selection:

- **High:** Clear rationale for study design choices.
- **Moderate:** Partially explained.
- **Low/Critically Low:** Poorly explained or unjustified.

#### 4. Literature Search Strategy:

- **High:** Comprehensive search including databases, registries, and grey literature.
- **Moderate:** Missing some search components.
- **Low/Critically Low:** Incomplete or poorly conducted search.

#### 5. Study Selection:

- **High:** Performed in duplicate with high agreement.
- **Moderate:** Partial duplication or lower agreement.
- **Low/Critically Low:** Single reviewer or poor agreement.

#### 6. Study Data Extraction:

- **High:** Performed in duplicate with high agreement.
- **Moderate:** Partial duplication or lower agreement.
- **Low/Critically Low:** Single reviewer or poor agreement.

#### 7. Excluded Studies:

- **High:** List provided with justification.
- **Moderate:** List provided without full justification.
- **Low/Critically Low:** No list or justification.

#### 8. Study Description:

- **High:** Detailed descriptions with all key criteria met.
- **Moderate:** Some descriptions missing or partially reported.
- **Low/Critically Low:** Inadequate descriptions.

#### 9. Risk of Bias Assessment:

- **High:** Comprehensive assessment using all QUADAS-2 domains.
- **Moderate:** Partial assessment.
- **Low/Critically Low:** Incomplete or missing assessment.

#### 10. Sources of Funding:

- **High:** All sources reported.
- **Low/Critically Low:** Sources not reported.

#### 11. Statistical Combination of Results (Meta-analysis):

- **High:** Appropriate methods addressing heterogeneity.
- **Moderate:** Adequate methods, partial heterogeneity discussion.
- **Low/Critically Low:** Inappropriate methods, no heterogeneity discussion.

**12. Impact of Risk of Bias on Meta-analysis:**

- **High:** Assessed and discussed.
- **Low/Critically Low:** Not assessed or discussed.

**13. Interpretation of Results Considering Risk of Bias:**

- **High:** Thorough discussion.
- **Low/Critically Low:** Lacking discussion.

**14. Explanation of Heterogeneity:**

- **High:** Comprehensive explanation.
- **Moderate:** Partial explanation.
- **Low/Critically Low:** Poor or no explanation.

**15. Publication Bias Investigation:**

- **High:** Conducted and discussed.
- **Low/Critically Low:** Not conducted.

**16. Conflict of Interest Reporting:**

- **High:** Fully reported.
- **Low/Critically Low:** Not reported.

**Determining Overall Confidence**

**1. Identify Critical Items:**

- Items marked as critical in systematic reviews (such as; Literature Search Strategy, Risk of Bias Assessment).

**2. Assess for Flaws:**

- Determine the number of critical and non-critical flaws.

**3. Apply Overall Rating Criteria:**

- **High Confidence:** No or one non-critical weakness.
- **Moderate Confidence:** More than one non-critical weakness.
- **Low Confidence:** One critical flaw with or without non-critical weaknesses.
- **Critically Low Confidence:** More than one critical flaw with or without non-critical weaknesses.