

KEY MESSAGE

AI-based retinal image analysis showed excellent accuracy for diabetic retinopathy in primary care, supporting its potential to augment screening, although prospective real-world validation remains essential.

Background

- Diabetic retinopathy (DR) remains a leading cause of preventable sight loss globally, particularly in working-age adults.
- Primary care screening programmes are established, but manual image grading places pressure on finite workforce capacity.
- Artificial intelligence (AI) offers scalable automated retinal image analysis with potential efficiency and cost-effectiveness benefits -- 3 devices are currently FDA approved

Purpose

- Synthesise current evidence on the diagnostic accuracy of AI models for detecting referable and vision threatening DR in primary care.
- To quantify pooled sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and SROC AUC.

Methods

Search period
Jan 2019 – Jan 2026

Studies included
11 of 55 identified

PubMed

Cochrane Library

Google Scholar

IEEE Xplore

Population:
Primary care DR screening;
all DR grades eligible

Outcomes:
Sensitivity, Specificity, PPV,
NPV and AUC

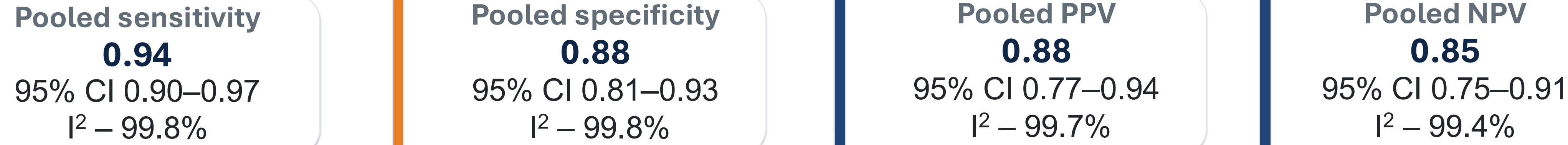
Inclusion criteria

Intervention / Comparator:
AI model detecting DR from
fundus photographs; trained
image grader / ophthalmologist

Study design:
RCTs; observational studies;
diagnostic accuracy studies;
real-world implementation
studies

- Conducted in accordance with PRISMA-DTA guidance.
- Two reviewers screened studies and extracted data
- Random-effects meta-analysis performed in RStudio; heterogeneity assessed by I^2 statistic; risk of bias assessed using QUADAS-2

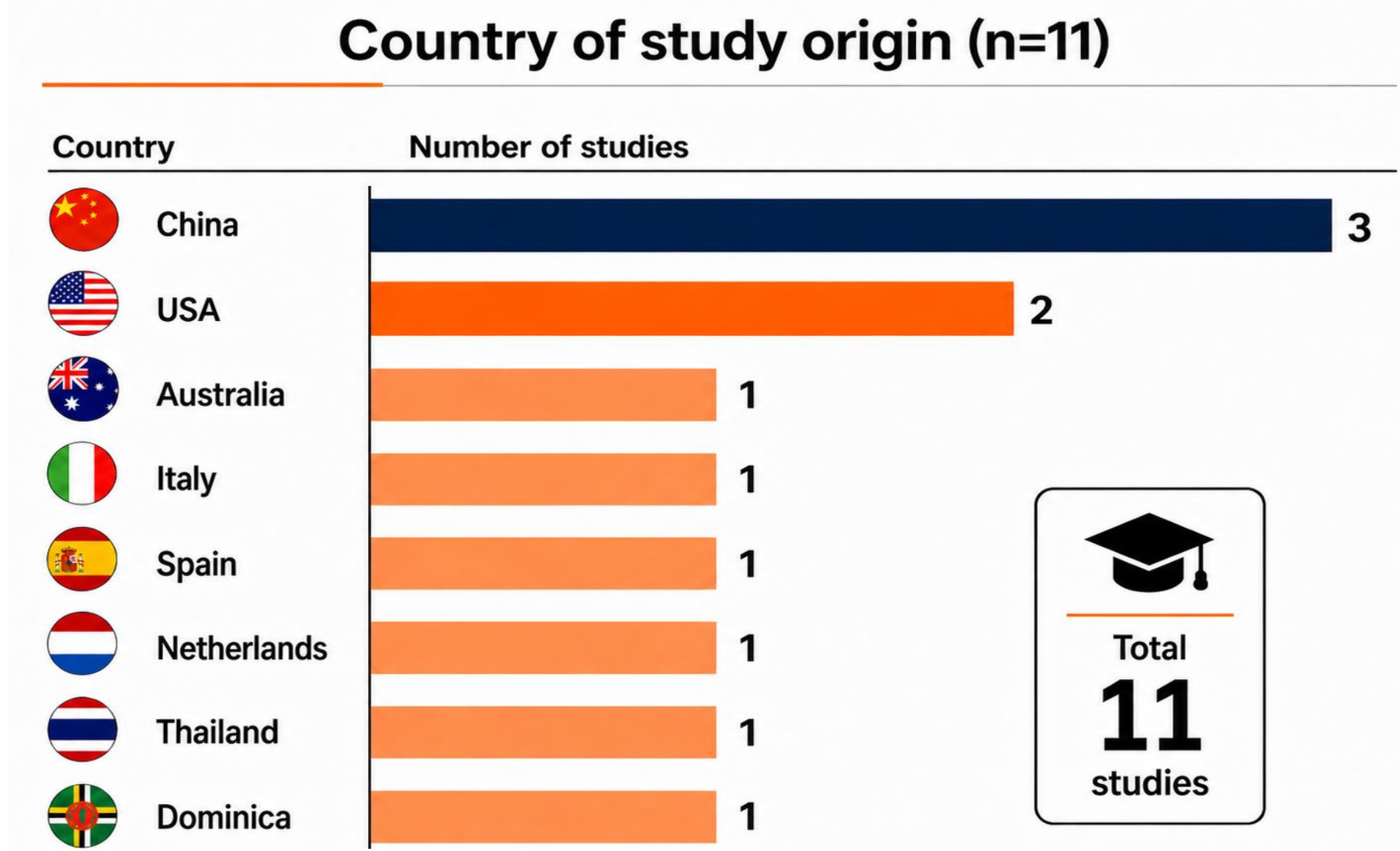
Results



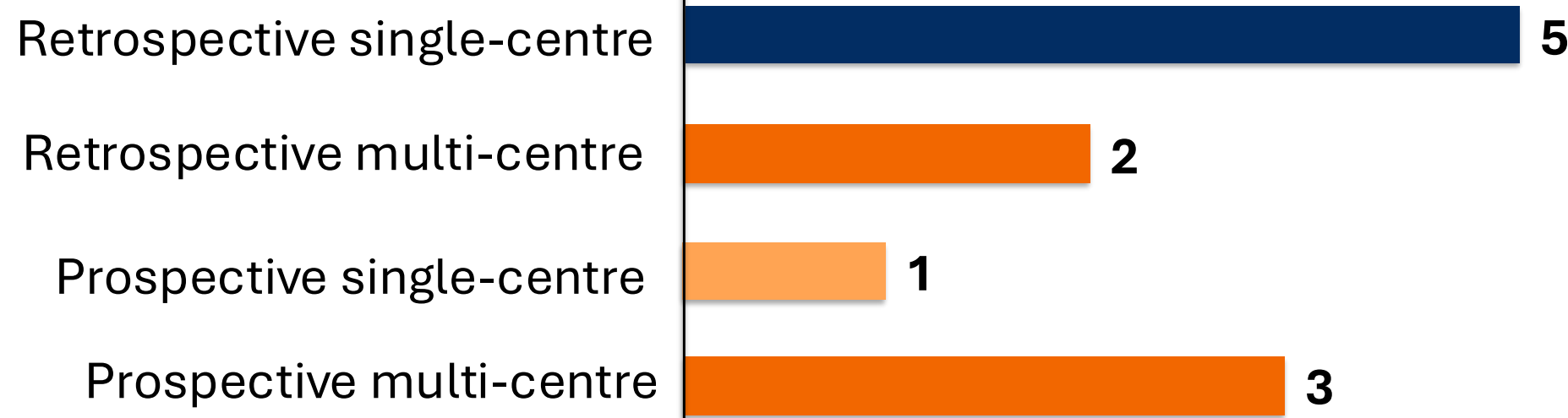
All 11 studies reported sensitivity and specificity, 5 studies reported PPV and NPV

SROC AUC = 0.971, indicating excellent discriminative ability for detecting referable diabetic retinopathy

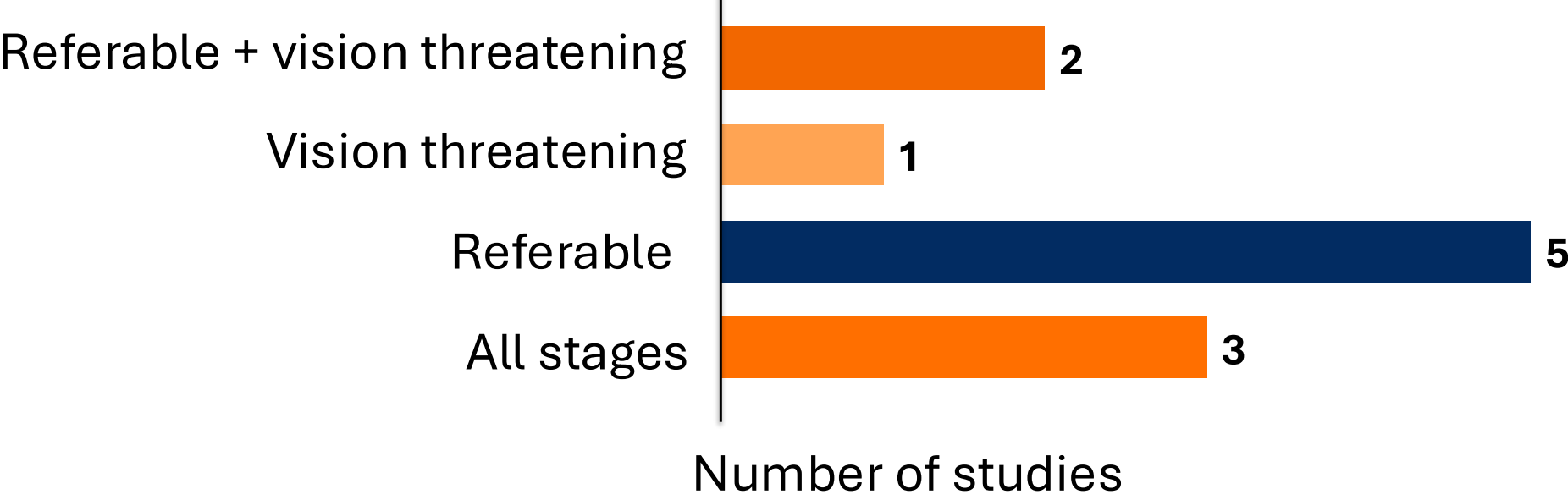
Study characteristics



Study design

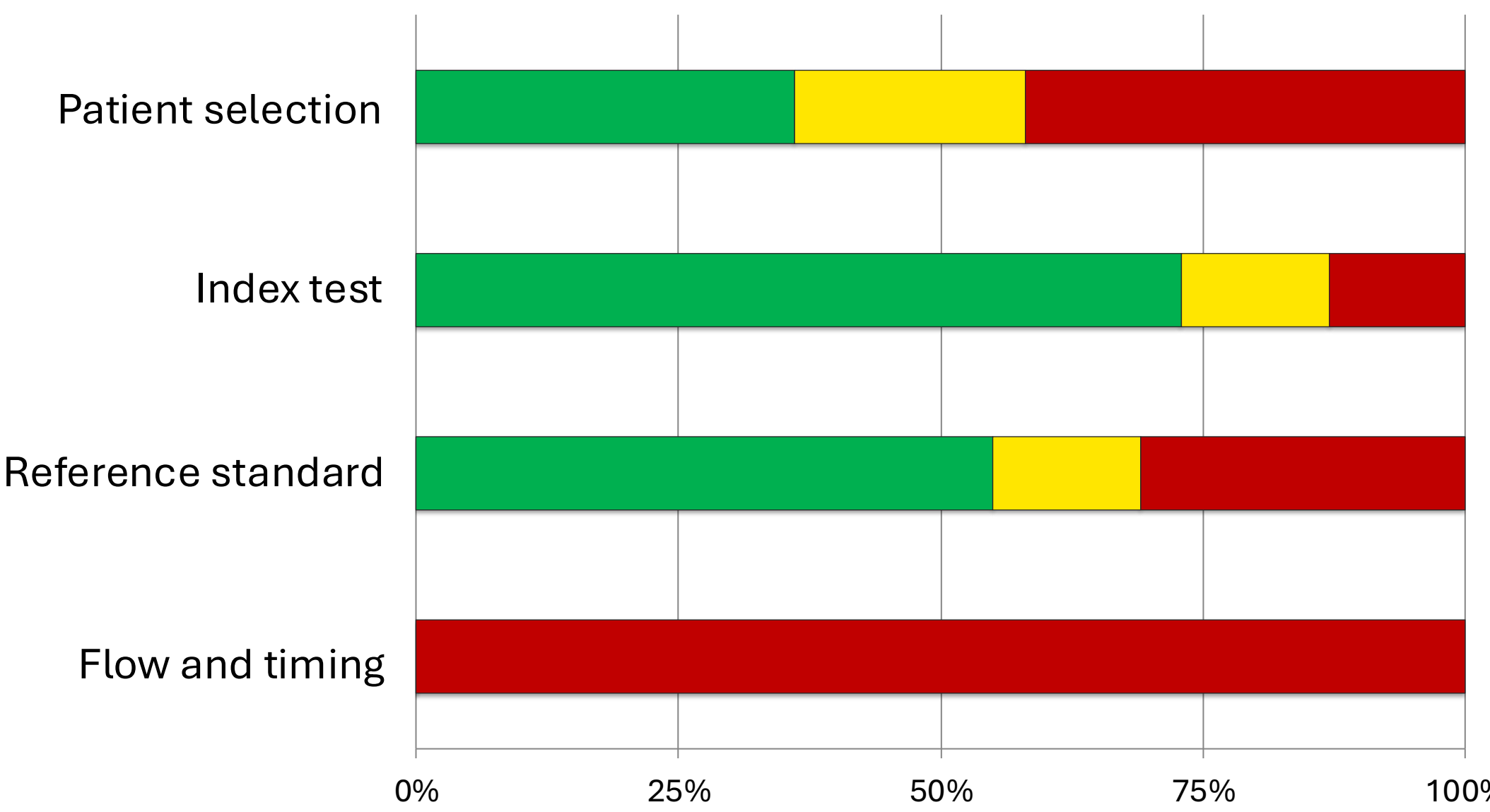


Stage of DR assessed

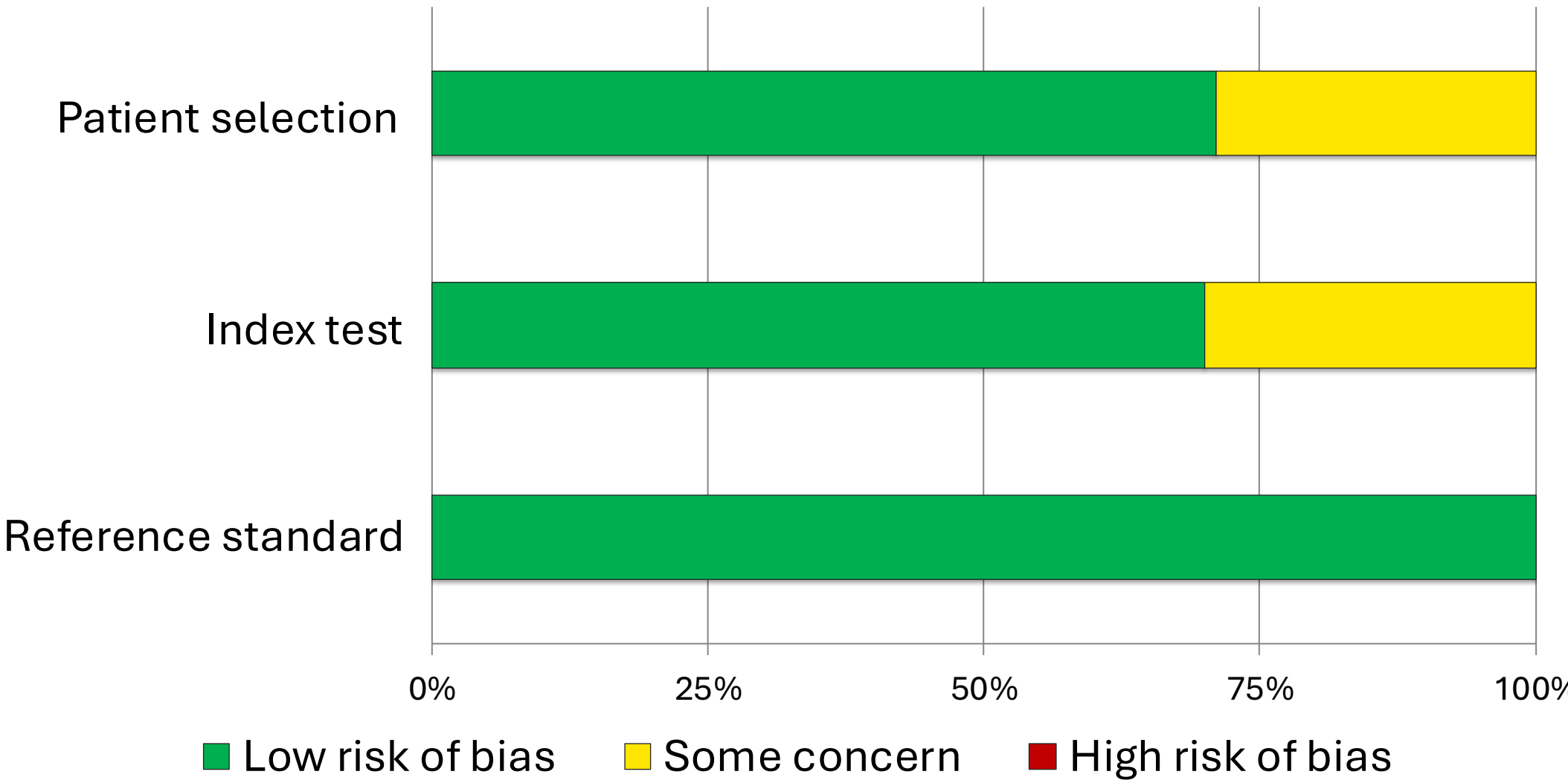


QUADAS-2

Risk of Bias Summary



Applicability Concerns Summary



Discussion

- AI-based models demonstrate excellent diagnostic accuracy for detecting referable and vision threatening DR in primary care.
- Pooled performance appears comparable to published human expert grader performance, supporting a role for AI as an augmentation tool.
- The British Diabetic Association stipulates a minimum sensitivity of 80% and specificity of 95% for any DR screening test
- Potential benefits include improved scalability, screening efficiency and support for resource-limited settings.
- However, inter-study heterogeneity, universal risk of bias and absence of randomised controlled trials limit immediate generalisability.
- Implementation decisions should prioritise regulatory-approved systems evaluated prospectively within real-world primary care pathways.

Limitations

- Substantial heterogeneity was present across all outcomes ($I^2 \geq 99.0\%$).
- AI models may perform less well in real-world settings due to variation in image quality, lighting, and camera specifications.
- Imaging protocols differed between studies, including both mydriatic and non-mydriatic fundus photography.
- Definitions of referable diabetic retinopathy varied substantially, ranging from mild non-proliferative DR to vision-threatening disease.

Next steps

- Undertake prospective multicentre studies evaluating regulatory-approved AI systems
- Standardise DR grading definitions and imaging protocols across studies to reduce heterogeneity
- Develop implementation and comparative-effectiveness studies against current screening pathways.
- Assess cost-effectiveness, workflow integration, image quality failure rates and equity across diverse populations.
- Trial deployment in national screening programmes

References

