A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Utilising AI: Protocol for QUADAS-AI

Ahmad Guni, Viknesh Sounderajah, Penny Whiting, Patrick Bossuyt, Ara Darzi, Hutan Ashrafian

Submitted to: JMIR Research Protocols
on: March 08, 2024

Table of Contents

A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Utilising AI: Protocol for QUADAS-AI

Ahmad Guni[1*]; Viknesh Sounderajah[1*]; Penny Whiting[2]; Patrick Bossuyt[3]; Ara Darzi[1]; Hutan Ashrafian[1]

[1]Institute of Global Health Innovation Imperial College London London GB
[2]Population Health Sciences Bristol Medical School University of Bristol Bristol GB
[3]Department of Epidemiology & Data Science Amsterdam University Medical Centres Amsterdam NL
[*]these authors contributed equally

Corresponding Author:
Hutan Ashrafian
Institute of Global Health Innovation
Imperial College London
10th Floor QEQM Building, St Mary's Hospital
Imperial College London
London
GB

Abstract

Background: QUADAS, and more recently QUADAS-2, were developed to aid the evaluation of methodological quality within primary diagnostic accuracy studies. However, its current form, QUADAS-2 does not address the unique considerations raised by artificial intelligence (AI) centred diagnostic systems. The rapid progression of the AI diagnostics field mandates suitable quality assessment tools to determine risk of bias and applicability, and subsequently evaluate translational potential for clinical practice.

Objective: We aim to develop an AI-specific quality assessment tool (QUADAS-AI), which addresses the specific challenges associated with the appraisal of AI diagnostic accuracy studies. This paper describes the processes and methods that will be used to develop QUADAS-AI.

Methods: The development of QUADAS-AI can be distilled into three broad stages. Stage 1: A project organisation phase has been undertaken, during which a Project Team and a Steering Committee were established. Following this, the scope of the project was finalised. Stage 2: An item generation process will be completed following: (1) a mapping review, (2) a meta-research study, (3) a scoping survey of international experts, and (4) a patient and public involvement and engagement (PPIE) exercise. A modified Delphi consensus methodology will be carried out to refine the tool, following which the initial QUADAS-AI tool will be drafted. A piloting phase will be carried out to identify components which are considered to be either ambiguous or missing. Stage 3: Specific dissemination strategies will be aimed towards academic, policy, regulatory, industry and public stakeholders respectively.

Results: Ethical approval to carry out the study, including the Delphi consensus process, has been granted by the Joint Research Compliance Office at Imperial College London (reference number: 21IC6664). QUADAS-AI aims to provide a consensus-derived platform upon which stakeholders may systematically appraise the methodological quality associated with AI diagnostic accuracy studies by the beginning of 2025.

Conclusions: AI-driven systems comprise an increasingly significant proportion of research in clinical diagnostics. Through this process, QUADAS-AI will aid the evaluation of studies in this domain in order to identify bias and applicability concerns. As such, QUADAS-AI may form a key part of clinical, governmental and regulatory evaluation frameworks for AI diagnostic systems globally.

(JMIR Preprints 08/03/2024:58202)
DOI: https://doi.org/10.2196/preprints.58202

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Utilising AI: Protocol for QUADAS-AI

Ahmad Guni[1, 2]*, Viknesh Sounderajah[1, 2]*, Penny Whiting[3], Patrick M Bossuyt[4], Ara Darzi[1, 2], Hutan Ashrafian[1, 2]

[1] Institute of Global Health Innovation, Imperial College London, London, UK

[2] Department of Surgery and Cancer, Imperial College London, London, UK

[3] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[4]Department of Epidemiology & Data Science, Amsterdam University Medical Centres, Amsterdam, Netherlands

*Contributed equally

Corresponding Author:

Professor Hutan Ashrafian

Institute of Global Health Innovation

10th Floor QEQM Building, St Mary's Hospital

Imperial College London

London

United Kingdom

W2 1NY

Email: h.ashrafian@imperial.ac.uk

-

Abstract

Background

QUADAS, and more recently QUADAS-2, were developed to aid the evaluation of methodological quality within primary diagnostic accuracy studies. However, its current form, QUADAS-2 does not address the unique considerations raised by artificial intelligence (AI) centred diagnostic systems. The rapid progression of the AI diagnostics field mandates suitable quality assessment tools to determine risk of bias and applicability, and subsequently evaluate translational potential for clinical practice.

Objectives

We aim to develop an AI-specific quality assessment tool (QUADAS-AI), which addresses the specific challenges associated with the appraisal of AI diagnostic accuracy studies. This paper describes the processes and methods that will be used to develop QUADAS-AI.

Methods

The development of QUADAS-AI can be distilled into three broad stages. Stage 1: A project organisation phase has been undertaken, during which a Project Team and a Steering Committee were established. The Steering Committee consists of a panel of international experts representing diverse stakeholder groups. Following this, the scope of the project was finalised. Stage 2: An item generation process will be completed following: (1) a mapping review, (2) a meta-research study, (3) a scoping survey of international experts, and (4) a patient and public involvement and engagement (PPIE) exercise. Candidate items will then be put forward to the international Delphi panel to achieve consensus for inclusion in the revised tool. A modified Delphi consensus methodology involving multiple online rounds and a final consensus meeting will be carried out to refine the tool, following which the initial QUADAS-AI tool will be drafted. A piloting phase will be carried out to identify components which are considered to be either ambiguous or missing. Stage 3: Once the Steering Committee has finalised the QUADAS-AI tool, specific dissemination strategies will be

aimed towards academic, policy, regulatory, industry and public stakeholders respectively.

Results

Ethical approval to carry out the study, including the Delphi consensus process, has been granted by the Joint Research Compliance Office at Imperial College London (reference number: 21IC6664). As of July 2024, the project organisation phase, as well as the mapping review and meta-research study, have been completed. We aim to complete the item generation, including the Delphi consensus, and finalise the tool by the end of 2024. QUADAS-AI will therefore be able to provide a consensus-derived platform upon which stakeholders may systematically appraise the methodological quality associated with AI diagnostic accuracy studies by the beginning of 2025.

Conclusions

AI-driven systems comprise an increasingly significant proportion of research in clinical diagnostics. Through this process, QUADAS-AI will aid the evaluation of studies in this domain in order to identify bias and applicability concerns. As such, QUADAS-AI may form a key part of clinical, governmental and regulatory evaluation frameworks for AI diagnostic systems globally.

Introduction

Despite much promise, the integration of artificial intelligence (AI) centred systems into clinical workflows has been limited thus far. In the current paradigm, diagnostic investigations require interpretation from expert clinicians in order to generate a diagnosis and subsequently determine management. However, diagnostic services across the world are overburdened with unmanageable workloads which exceed workforce capacity [1]. In order to address this, diagnostic AI systems have been characterised by regulators and technologists as medical devices [2] that may achieve diagnostic accuracy comparable to that of an expert clinician, whilst concurrently alleviating health-resource utilisation, helping to reduce medical errors. Indeed, the majority of healthcare-related AI systems that have reached regulatory approval belong to the field of medical diagnostics [3]. As seminal primary research studies arise in the theme of AI diagnostics [4, 5], there has been a concomitant rise in secondary research studies which amalgamate the findings of comparable studies.

Although systematic reviews serve an important role in summarising evidence, the vast majority related to AI diagnostic accuracy have been conducted in the absence of an AI specific methodological quality assessment tool [6]. AI diagnostic accuracy studies are methodologically distinct from traditional diagnostic accuracy studies as they comprise distinct methods, analyses and outcome measures which mandate specific considerations when assessing quality [7]. Currently, the most commonly used instrument for the methodological assessment of secondary research studies remains the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool [8]. It is a quality assessment tool designed for use in systematic reviews, initially developed in 2003 [9] and updated in 2011; its use is strongly encouraged by many biomedical journals. It consists of four key domains: (1) patient selection, (2) index test, (3) reference standard and (4) flow and timing. These domains allow researchers to undertake a structured appraisal of a research study's internal validity (biases) and external validity (applicability) respectively. The absence of a robust quality assessment tool in

the AI field not only hinders efficient quality appraisal at an evidence synthesis phase but has considerable downstream effects as key stakeholders, such as policy makers, regulatory officials, technologists and healthcare professionals, are unable to effectively evaluate the translational potential of these nascent technologies.

We propose an AI-specific extension, termed QUADAS-AI, that aims to provide researchers and policy makers with a framework to appraise methodological quality in systematic reviews evaluating the diagnostic accuracy of AI. This work is complementary to the STARD-AI [10] and QUADAS-3 initiatives. QUADAS-AI is being coordinated by a Project Team and Steering Committee consisting of clinician scientists, computer scientists, journal editors, EQUATOR Network representatives, regulatory leaders, epidemiologists, statisticians, industry leaders, funders, health policy makers, legal experts and bioethicists. Given the global reach of this class of technologies and the transformative potential in clinical diagnostics, we view that connecting global stakeholders is of the utmost importance for this initiative. This study aims to produce a novel quality assessment tool (QUADAS-AI) which accounts for the specific considerations required for the appraisal of AI diagnostic accuracy studies.

Methods

This protocol has benefitted from the experience and expertise from members of the Project Team and Steering Committee who have previously led upon the development of seminal quality assessment tools over the past two decades. These include QUADAS and QUADAS-2 for diagnostic accuracy studies, ROBIS for systematic reviews [11] and PROBAST [12] for prediction modelling studies. Moreover, there is shared learning from the development of AI-specific reporting guidelines and risk of bias tools, including STARD-AI [10] and PROBAST-AI [13]. The development of QUADAS-AI can be distilled into broad three stages, as previously delineated [14]. Given the pressing need for suitable quality assessment standards of diagnostic studies in this field, development is projected to finish by Q4 2024.

Stage 1

Project organisation:

QUADAS-AI is being undertaken by a Project Team and a Steering Committee. The Project Team consists of the founder of QUADAS (PW), the lead for the STARD-AI initiative (HA) and a clinician scientist (AG). The Project Team are responsible for identifying members of the Steering Committee, candidate item generation, undertaking the online surveys for the modified Delphi consensus process, organising the consensus meeting, drafting the QUADAS-AI tool and accompanying documents, coordinating the piloting of the draft QUADAS-AI tool, and leading the dissemination process.

The Steering Committee was created in order to provide diverse stakeholder guidance in this process, as well as to identify additional experts to invite for the consensus response and draft the final QUADAS-AI tool. The Steering Committee currently comprises of approximately 15 members and consists of healthcare professionals, computer scientists, epidemiologists, statisticians, regulatory

officials, health policy leaders and industry leaders. These individuals were identified through their notable work in the fields of (1) diagnostic accuracy research, (2) artificial intelligence in healthcare and (3) applied health policy.

Defining Scope

The scope of QUADAS-AI has been defined by the Project Team and Steering Committee through a discussion framed around questions previously proposed [14]. It was predetermined that QUADAS-AI, as per previous iterations of the tool, will focus upon the methodological quality of AI diagnostic accuracy studies. This study is complementary to the ongoing QUADAS-3 initiative, which is the next iteration of QUADAS and is currently led by one of the study authors and project team (PW). If a draft of the QUADAS-3 becomes available during the development of QUADAS-AI or any substantial updates are anticipated in comparison to QUADAS-2, we will base the QUADAS-AI tool on the QUADAS-3 structure; otherwise, we will instead focus on QUADAS-2. Discourse related to the (1) assessments related risk of bias (internal validity), (2) assessments related to applicability (external validity), (3) tool structure and (4) rating system is a dynamic process that will be open to adaptation throughout Stage 2 of the study.

Stage 2

Item generation

In order to generate a candidate list of items to enter the modified Delphi consensus process, the Project Team will undertake a mapping review, a meta-research study, a scoping survey with a global panel of experts, and a patient public involvement and engagement (PPIE) exercise.

Mapping review

A mapping review of both academic and non-academic literature has been undertaken in order to

identify key considerations in the development of QUADAS-AI. An electronic database search of Medical Literature Analysis and Retrieval System Online (MEDLINE) and Excerpta Medica database (EMBASE) was conducted through Ovid. This process was augmented by non-systematic searches using traditional search engines for grey literature, social networking platforms as well as personal article collections highlighted by members of the Project Team. Extracted material were broadly classified into four categories: (1) general considerations regarding diagnostic accuracy studies and artificial intelligence; (2) evidence and statements suggesting modifications to current items; (3) evidence and statements suggesting additions of items; and (4) evidence and statements suggesting the removal of specific items.

Meta-research study

As previously noted, there have been no studies examining the adherence and suitability of QUADAS-2 for the appraisal of AI diagnostic accuracy study quality. Therefore, a meta-research study was carried out to evaluate adherence of AI diagnostic accuracy systematic reviews to the existing QUADAS-2 tool (6). This study demonstrated that there is incomplete uptake of quality-assessment tools as well as inconsistent reporting of bias in AI-diagnostic accuracy systematic reviews, with just over half of the studies using QUADAS-2. This study also identified key biases and features unique to AI diagnostic accuracy studies. These will contribute to the formulation of candidate items for addition or modification.

Online scoping survey

The Project Team and Steering Committee will undertake a survey of an international panel of experts in order to identify potential further items or modifications that warrant inclusion in QUADAS-AI. A diverse and independent panel of experts will be identified by the Project Team and Steering Committee from the various stakeholder groups outlined above. They will be provided with

an information sheet describing the study and asked to participate in an online questionnaire. Participants will be asked to consider whether each item on the existing QUADAS-2 tool should be retained, removed or modified in the QUADAS-AI tool. Free-text sections will allow participate to express their thoughts on each item as well as suggest modifications or further considerations. Furthermore, participants will be asked to comment on additional candidate items or considerations produced from preceding rounds of the item generation process.

Patient public involvement and engagement (PPIE) exercise

-

Lastly, a focus group will be conducted with patients and members of the public who have expressed an interest in participating in forums related to digital health and AI. The objective of these discussions is two-fold: (1) to further identify issues not uncovered during previous evidence generation steps and (2) to gain further understanding of the perceived importance to the public of specific items that have been raised thus far. These discussions will be conducted remotely using Zoom (Zoom Video Communications, Inc., USA).

An expert facilitator will lead a discussion on the current uses of AI in healthcare, including considerations on the aims of QUADAS-AI and important items the participants deem to be important to capture during the study process. As stakeholder discussions will be conducted virtually on Zoom, anonymised post-hoc discussion transcripts will be retained.

Collation of items

The Project Team and Steering Committee will group items from the item generation phase into domains and subsequently word items as signalling questions. An online discussion among the Project Team and members of the Steering Committee will be held to further refine the domains and signalling questions into a draft tool, which will then enter the Delphi consensus process for approval

and refinement.

Modified Delphi consensus process

We will adopt a pragmatic modified Delphi consensus methodology. The Delphi consensus methodology is a well-established method of obtaining a collective opinion from a group of experts through a series of questionnaires; each one refined based upon feedback from respondents on a previous version [15]. We will conduct the Delphi consensus process in a similar way as described in the STARD-AI protocol [10].

Participants from across the world are invited to join the QUADAS-AI Consensus Group on account of their expertise as clinician scientists, computer scientists, journal editors, EQUATOR Network representatives, epidemiologists, statisticians, health technology industry leaders, funders, health policy makers, legal experts and bioethicists. The Steering Committee will identify potential participants from their wider professional network or experts who have made significant contributions to their respective fields. Invited experts will be provided with a written invite detailing the study and given a six-week timeframe to respond. Those who accept the invitation will be invited to complete each round of the modified Delphi consensus process, and will be acknowledged as an author, within a group authorship model, in the publication that arises from this study. Studies of similar scope and breadth, such as STARD-AI, recruited over 150 participants from varied backgrounds across the world. A similar number is anticipated for QUADAS-AI.

During each phase of the modified Delphi consensus process, participants will use a 5-point Likert-like scale to evaluate each item (1 – very important, 2 – important, 3 – moderately important, 4 – slightly important, 5 – not at all important). The threshold for consensus will be predefined at ≥75%. Items which achieve ≥75% ratings of 1 or 2 will be put forward for discussion in the final round, which will occur in the form of a virtual teleconference meeting. Items which achieve ≥75% ratings

of 4 or 5 will be excluded. Items that do not meet the 75% consensus threshold will advance to the next phase of the Delphi process. Participants will also have the opportunity to propose additional items that they believe warrant discussion in future rounds through open-ended responses.

In subsequent rounds, the survey will compose of items for which consensus was not achieved and any new items suggested in prior rounds. Each item will be accompanied by a reminder of the participant's last rating and the average rating from all participants in the prior round. This allows participants to reconsider their initial evaluations with the benefit of understanding the perspective of the wider group. Items which have not reached consensus will be put forward for discussion in following rounds until a consensus is reached. We will conduct descriptive statistical tests upon the results for each round (median, range, mean, percentage agreement and consensus).

Once a consensus is reached, there will be a final meeting between a small group of the Project Team and Steering Committee to finalise the structure and content of the QUADAS-AI tool based on feedback from the Delphi consensus. The primary objective is to develop a draft version of the QUADAS-AI tool. As recommended in the COMET handbook, the nominal group technique, a highly structured group interaction framework, will be utilised to aid this process [16, 17]. Following a brief introduction and explanation of the purpose of the meeting by the facilitators, participants will discuss the inclusion and exclusion of candidate items and share any comments until all contributions are exhausted. This discussion phase will be led by the facilitators to ensure that the discussion will not be dominated by any one individual and be as neutral as possible [18].

The first two rounds of the modified Delphi consensus process will be conducted as online surveys using the DelphiManager software (version 4.0), which is developed and maintained by the COMET (Core Outcome Measures in Effectiveness Trials) initiative. The final meeting to draft the QUADAS-

AI tool will be conducted using Zoom. All data is pseudo-anonymised and no identifiable data will be published.

Development of the (1) quality assessment tool, (2) statement and (3) explanation and elaboration (E&E) document

Upon completion, the Project Team will construct the initial QUADAS-AI tool. The draft tool, with an accompanying statement, will be shared amongst the wider Steering Committee in order to discuss its content and therefore allow the Steering Committee to suggest additions, subtractions or modifications as they see fit.

Piloting amongst experts and non-experts

Upon completion of the first draft of the QUADAS-AI tool, we intend to organise multiple rounds of piloting amongst expert and non-expert users (QUADAS-AI Pilot Group). The main aim of these piloting sessions is to test the tool's usability as well as identify items which are considered to be vague, ambiguous or perceived to be missing. We intend to undertake this process amongst healthcare professionals, computer scientists, expert statisticians, journal editorial boards, key industry stakeholders, regulatory leaders as well as policy experts. Interviews amongst this QUADAS-AI Pilot Group will be undertaken in order to ensure that a granular level of feedback is attained for points of discussion. Members of the pilot group will not be part of the Steering Committee or have previously participated in the consensus process in order to provide an independent opinion. We anticipate around 20 to 30 members will be recruited. Experts and non-experts within the Pilot Group will be acknowledged by name as an author, within a group authorship model, in the publications that arise from this study.

In conjunction with this piloting process, the Project Team will prepare the explanation and

elaboration (E&E) document, to provide rationale for the domains, structure and items associated with the tool.

Stage 3: Dissemination

Following the piloting phase, the final proposed amendments to QUADAS-AI will be discussed amongst the Project Team and the Steering Committee. Once consensus has been reached through e-mail correspondence, the documents will be disseminated.

We strongly anticipate that the dissemination strategy will be principally tailored towards five groups of stakeholders: (a) academia, (b) policy, (c) guidelines and regulation, (d) industry and (e) patient-representative bodies. Although a significant amount of material will cross over between stakeholders, creating stakeholder specific material is considered to be the most meaningful way of achieving impact.

Academic stakeholders

We aim to publish the QUADAS-AI tool, the accompanying statement and the E&E document in an open access format in a high-impact peer-reviewed journal. In order to further complement this, we aim to create specialty-specific discourse regarding QUADAS-AI through focused editorials in pertinent journals. These journal editors will also be actively encouraged to endorse the use of QUADAS-AI as part of their peer review process. Translations of the tool in various languages are also encouraged in order to further broaden the scope of its impact. We urge interested parties to contact the corresponding author for further information about the translation policies.

Policy stakeholders

We are in close collaboration with organisations such as Public Health England, National Health

Service (NHS) Digital, National Institute for Health and Care Excellence (NICE) and the NHS Accelerated Access Collaborative (AAC) and their wider network to ensure that the tool will form part of their health technology assessment pathways.

Guidelines and regulatory stakeholders

QUADAS-AI has been co-designed with senior figures from the United States Food and Drug Administration (FDA) and the United Kingdom Medicines and Healthcare products Regulatory Agency (MHRA). Whilst they do not represent the views of either organisation, these Steering Committee members have a high-level understanding as to how QUADAS-AI may be constructed to achieve maximal real-world impact.

Industry stakeholders

We will present QUADAS-AI to a broad range of health technology companies, ranging from start-ups, small and medium-sized enterprises to multinational corporations, so that their product pipelines may accommodate for this.

Public and non-specific stakeholders

Ensuring that the core material is available in an open access fashion, through a CC-BY licence, is paramount to achieving general impact. In addition, we aim to publish articles in mainstream media and attain distribution through non-traditional means (e.g. social networking platforms, webinars, podcast episodes and blog posts).

Ethical considerations

Ethical approval for the study has been granted by the Joint Research Compliance Office at Imperial College London (reference number: 21IC6664). Written consent will be gained for all participants in

the online scoping survey, PPIE, Delphi consensus process and checklist piloting.

Results

As of July 2024, the Project Team and Steering Committee have been established, as has the scope of the project. The study is currently in the item generation phase (Stage 2), and the mapping and meta-research reviews have been completed (6). We aim to conduct the scoping survey of experts, PPIE and Delphi consensus process by the end of 2024 and publish the statement by the first quarter of 2025 for stakeholder use.

Discussion

QUADAS-AI will be a consensus-derived quality assessment tool that will allow readers to critically appraise the risk of bias and applicability of study findings in systematic reviews of diagnostic accuracy studies using AI. By providing a framework to evaluate the methodological quality of studies, stakeholders will be in a better position to assess the evidence-base and potential for clinical translation of AI-driven diagnostic tools.

AI technology will likely be integrated into several clinical workflows within the next decade in order to enhance patient care and improve clinical outcomes. Specifically, clinical diagnostics has emerged as a key area that has gathered significant interest from global clinical, academic and industry communities. The importance of evidence synthesis becomes increasingly evident as rapidly advancing AI technology continues to be applied within the diagnostic field; this is typically achieved with systematic reviews to draw clinically relevant conclusions from summarised findings. Therefore, robust methods to evaluate evidence synthesis will be fundamental to the clinical development and implementation of AI technologies as the research community continues to harness the unique ability of AI to generate and process ever-increasing amounts of health data. However, given the notable flaws in using current quality assessment tools, there is a pressing need to develop an AI-specific quality assessment tool that can suitably assess the unique nature of AI diagnostic accuracy studies. We hope that this international, multi-stakeholder consensus approach will sufficiently address the unique considerations of AI technology, and will ultimately provide a useful tool for clinical, academic, policy, regulatory and industry stakeholders.

Acknowledgements

Infrastructure support for this research was provided by the NIHR Imperial Biomedical Research Centre (BRC).

Data Availability

Datasets generated or analysed during this study are available on reasonable request from the corresponding author.

Author contributions

A.G., V.S., P.W., P.B., A.D. and H.A. developed the concept and methodology of the study. A.G. and V.S. drafted the manuscript. All authors read and approved the manuscript.

Conflicts of interest

A.D. is Executive Chair for Preemptive Health and Medicine, Flagship Pioneering. H.A. is Chief Scientific Officer of Preemptive Health and Medicine, Flagship Pioneering. All other authors declare no competing interests.

Abbreviations

AAC: Accelerated Access Collaborative
AI: Artificial Intelligence
COMET: Core Outcome Measures in Effective Trials
E&E: explanation and elaboration
EMBASE: Excerpta Medica database
EQUATOR: enhancing the quality and transparency of health research
FDA: Food and Drug Administration
MEDLINE: Medical Literature Analysis and Retrieval System Online
MHRA: Medicines and Healthcare products Regulatory Agency
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PROBAST: prediction model risk of bias assessment tool
QUADAS: quality assessment of diagnostic accuracy studies
ROBIS: risk of bias in systematic reviews
STARD: standards for reporting diagnostic accuracy studies

References

1. Bethany Jill W, David B, Darren T. Future-proofing pathology: the case for clinical adoption of digital pathology. Journal of Clinical Pathology. 2017;70(12):1010. doi: 10.1136/jclinpath-2017-204644.
2. US Food and Drug Administration Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021; Available from: https://www.fda.gov/media/145022/download.
3. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digital Medicine. 2020 2020/09/11;3(1):118. doi: 10.1038/s41746-020-00324-0.
4. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020 2020/01/01;577(7788):89-94. doi: 10.1038/s41586-019-1799-6.
5. Milea D, Najjar RP, Zhubo J, Ting D, Vasseneix C, Xu X, et al. Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs. N Engl J Med. 2020 Apr 30;382(18):1687-95. PMID: 32286748. doi: 10.1056/NEJMoa1917130.
6. Jayakumar S, Sounderajah V, Normahani P, Harling L, Markar SR, Ashrafian H, Darzi A. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. NPJ Digit Med. 2022 Jan 27;5(1):11. PMID: 35087178. doi: 10.1038/s41746-021-00544-y.
7. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. npj Digital Medicine. 2020 2020/09/24;3(1):126. doi: 10.1038/s41746-020-00333-z.
8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. PMID: 22007046. doi: 10.7326/0003-4819-155-8-201110180-00009.
9. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003 2003/11/10;3(1):25. doi: 10.1186/1471-2288-3-25.
10. Sounderajah V, Ashrafian H, Golub RM, Shetty S, Fauw JD, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11(6):e047709. doi: 10.1136/bmjopen-2020-047709.
11. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016 Jan;69:225-34. PMID: 26092286. doi: 10.1016/j.jclinepi.2015.06.005.
12. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-8. PMID: 30596875. doi: 10.7326/m18-1376.
13. Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008. doi: 10.1136/bmjopen-2020-048008.
14. Whiting P, Wolff R, Mallett S, Simera I, Savović J. A proposed framework for developing quality assessment tools. Systematic Reviews. 2017 2017/10/17;6(1):204. doi: 10.1186/s13643-017-0604-6.
15. Brown BB. Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts. Santa Monica, CA: RAND Corporation; 1968.
16. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm. 2016 Jun;38(3):655-62. PMID: 26846316. doi: 10.1007/s11096-016-0257-x.

17. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. Trials. 2017 2017/06/20;18(3):280. doi: 10.1186/s13063-017-1978-4.
18. Harvey N, Holmes CA. Nominal group technique: an effective method for obtaining group consensus. Int J Nurs Pract. 2012 Apr;18(2):188-94. PMID: 22435983. doi: 10.1111/j.1440-172X.2012.02017.x.

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/83fa91326773eee9a5d48887a2016775.docx