



Adherence of randomised controlled trials using artificial intelligence in ophthalmology to CONSORT-AI guidelines: a systematic review and critical appraisal

Niveditha Pattathil ¹, Jonathan Z L Zhao,² Olapeju Sam-Oyerinde,³ Tina Felfeli ^{4,5}

To cite: Pattathil N, Zhao JZL, Sam-Oyerinde O, *et al*. Adherence of randomised controlled trials using artificial intelligence in ophthalmology to CONSORT-AI guidelines: a systematic review and critical appraisal. *BMJ Health Care Inform* 2023;**30**:e100757. doi:10.1136/bmjhci-2023-100757

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2023-100757>).

NP and JZLZ are joint first authors.

Received 01 March 2023
Accepted 15 June 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Tina Felfeli;
tina.felfeli@mail.utoronto.ca

ABSTRACT

Purpose Many efforts have been made to explore the potential of deep learning and artificial intelligence (AI) in disciplines such as medicine, including ophthalmology. This systematic review aims to evaluate the reporting quality of randomised controlled trials (RCTs) that evaluate AI technologies applied to ophthalmology.

Methods A comprehensive search of three relevant databases (EMBASE, Medline, Cochrane) from 1 January 2010 to 5 February 2022 was conducted. The reporting quality of these papers was scored using the Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) checklist and further risk of bias was assessed using the RoB-2 tool.

Results The initial search yielded 2973 citations from which 5 articles satisfied the inclusion/exclusion criteria. These articles featured AI technologies applied to diabetic retinopathy screening, ophthalmologic education, fungal keratitis detection and paediatric cataract diagnosis. None of the articles reported all items in the CONSORT-AI checklist. The overall mean CONSORT-AI score of the included RCTs was 53% (range 37%–78%). The individual scores of the articles were 37% (19/51), 39% (20), 49% (25), 61% (31) and 78% (40). All articles were scored as being moderate risk, or ‘some concerns present’, regarding potential risk of bias according to the RoB-2 tool.

Conclusion A small number of RCTs have been published to date on the applications of AI in ophthalmology and vision science. Adherence to the 2020 CONSORT-AI reporting guidelines is suboptimal with notable reporting items often missed. Greater adherence will help facilitate reproducibility of AI research which can be a stimulus for more AI-based RCTs and clinical applications in ophthalmology.

INTRODUCTION

The growing advent of artificial intelligence (AI) has sparked interest globally across all fields of medicine and healthcare.¹ Within ophthalmology, AI has been used in the analysis of fundus photographs, visual field testing, optical coherence tomography

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Within the field of ophthalmology, there is a growing interest in exploring the potential of deep learning and artificial intelligence (AI), however, the level of quality of the randomised controlled trials (RCTs) currently published on the efficacy of AI-driven interventions is not known.

WHAT THIS STUDY ADDS

⇒ This systematic review aimed to characterise the RCTs using AI within the field of ophthalmology and vision science, and to critically appraise the adherence of each included study to the Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) reporting guideline.
⇒ A small number of RCTs have been published to date on the applications of AI in ophthalmology and vision science, and adherence to the 2020 CONSORT-AI reporting guidelines is suboptimal with notable reporting items often missed.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Further studies should aim for greater adherence to reporting standards to help facilitate reproducibility and generalisability of AI research and clinical applications in ophthalmology.

and surgical skill assessment.¹ It has also been applied to improve the efficiency and robustness of detection of conditions including diabetic retinopathy,² retinopathy of prematurity,³ glaucoma,⁴ macular oedema⁵ and age-related macular degeneration.⁶ However, further expansion of AI into clinical practice requires extensive research and development.

Randomised controlled trials (RCTs) are considered the gold standard experimental design for researchers seeking to provide evidence to support the safety and

efficacy of a new intervention.⁷ However, deficiencies in reporting clarity can interfere with accurate assessment of sources of potential bias arising from inadequacies in methodologies. The Consolidated Standards of Reporting Trials (CONSORT) statement provides the minimum guidelines for reporting randomised trials, and its use has been key in ensuring transparency in the assessment of new interventions. It was originally published in 1996,⁸ revised in 2001,⁹ and most recently updated in 2010.¹⁰ A 2001 review of 24 RCTs in ophthalmology found that on average, only 33.4 out of 57 descriptors were reported adequately according to the 1996 CONSORT guidelines.¹¹ A 2014 review that assessed the compliance of 65 ophthalmological RCTs published in 2011 with the 2008 CONSORT extension for non-pharmacological treatment interventions reported a mean CONSORT score of 8.9 out of 23 criteria, or 39%.¹²

The Consolidated Standards of Reporting Trials - Artificial Intelligence (CONSORT-AI) extension is the reporting guideline for clinical trials evaluating interventions with an AI component, published in 2020.¹³ The extension includes 14 new items considered sufficiently relevant with regard to evaluation of reporting requirements for methodologies in RCTs involving assessment of AI as the intervention.¹³ With the recent rise in new initiatives using AI, adherence to reporting guidelines such as CONSORT-AI plays a critical role in guiding and standardising the conduct and reporting of AI-related trials.

This systematic review aimed to characterise the RCTs using AI within the field of ophthalmology and vision science, and to critically appraise the adherence of each included study to the CONSORT-AI reporting guideline.

METHODS

Search strategy

This systematic review was conducted in accordance to the Preferred Reporting Items for a Systematic Reviews and Meta-analyses guidelines. The protocol was prospectively registered in PROSPERO (registration number: CRD42022304021). A comprehensive search of the relevant databases MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials and Cochrane Database of Systematic Reviews was done in consultation with an experienced librarian. All English-language RCTs using AI within the field of ophthalmology and vision science from 1 January 2010 to 5 February 2022 were identified. This restriction in publication date was put in place to capture the most relevant and recent publications in light of the increasing interest in research on AI following the advent and popularisation of the computing technique ‘deep learning’, especially with regard to image analysis.¹⁴ A combination of keywords and Medical Subject Headings related to concepts of RCTs, ophthalmology and AI were used to build the search strategy (online supplemental appendix 1).

Study selection and data extraction

Two authors (NP and JZLZ) independently conducted an initial title-abstract screening followed by full-text screening of all articles. All conflicts were resolved by consensus and in consultation with a third reviewer (TF or OS). The inclusion criteria were: articles that were (1) RCTs, (2) using AI as their main intervention and (3) evaluating the AI for application within any aspects in the field of ophthalmology. Articles were excluded if they were (1) not specific to ophthalmology and/or (2) were not available in English. The authors of articles whose full-text was not available were contacted to request full-text versions directly. Data from the final set of articles included in the review were extracted and recorded in a predetermined datasheet by two authors (NP and JZLZ).

Risk of bias assessment

Risk of bias assessment was completed for each study by two independent reviewers (NP and JZLZ) using the RoB-2 tool.¹⁵ Any conflicts were resolved in consultation with a third reviewer (TF or OS). For each domain, the risk of bias was reported as ‘high’, ‘low’ or ‘some concerns present’.

CONSORT-AI checklist

The final articles were scored independently by two authors (NP and JZLZ) using the CONSORT-AI checklist.¹³ Based on previously published methods, articles were scored 1 for an item if all of the components identified in the respective criterion were reported, and 0 if any portions were missing.^{12 16–18} There are 51 criteria in the CONSORT-AI checklist. Each item was given equal weight, scoring 1 point each. The resulting mark was termed the ‘CONSORT-AI score.’ The criterion regarding providing an explanation of any interim analyses and/or stopping guidelines if applicable (7b) was not applicable to any of the articles and was therefore scored as ‘0’ for all. After initial scoring, any discrepancies were resolved by consensus. If an agreement could not be reached, a third author (TF or OS) was consulted to make the final decision.

RESULTS

The search strategy yielded a total of 2973 citations (figure 1). Following deduplication and screening, five articles met the inclusion and exclusion criteria. The characteristics of the included articles are summarised in online supplemental table 1. The final articles included in this review looked at the utility of AI in diabetic retinopathy screening,^{19 20} ophthalmologic education,²¹ detecting fungal keratitis²² and diagnosing childhood cataracts.²³ Three out of the five included articles were studies conducted in China, and the remaining two were conducted in Mexico and Rwanda. The majority (3/5) of the articles were published in 2021 or 2022,^{19 20 22} and the remaining two were published in 2019 and 2020.^{21 23}

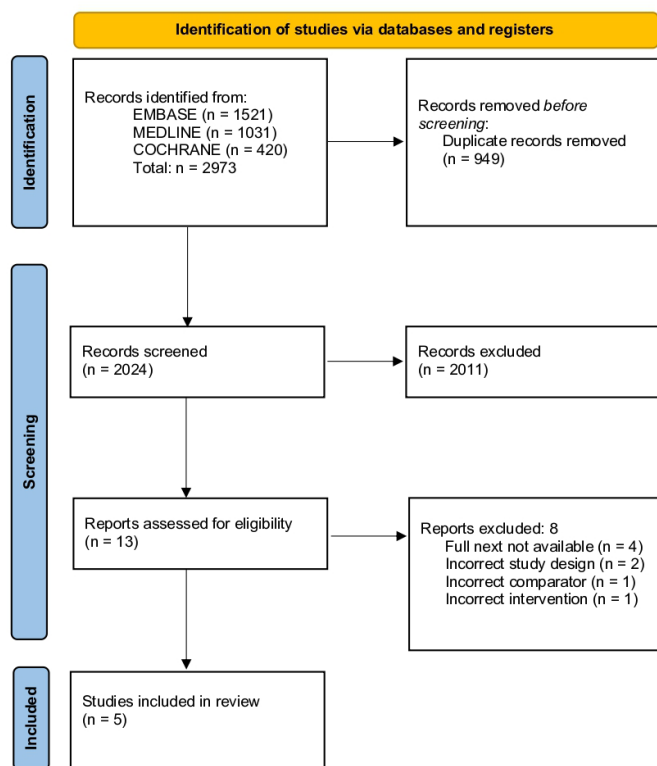


Figure 1 PRISMA, Preferred Reporting Items for a Systematic Reviews and Meta-analyses flow chart diagram for study identification and selection.³⁰

The overall mean CONSORT-AI score of the included RCTs was 53% (range 37%–78%), and the median score was 49%. The individual scores of the articles were 19/51 (37%), 20/51 (39%), 25/51 (49%), 31/51 (61%) and 40/51 (78%). Following the initial round of scoring, there was conflict on 14 items (5.49%). The inter-rater concordance for the CONSORT-AI scoring had a kappa score of 0.89.

The compliance of the included articles to each of the individual CONSORT-AI criteria is shown in online supplemental table 2. None of the articles addressed the following criteria: important changes to methods after trial commencement with reasons (3b), changes to trial outcomes after trial commenced with reasons (6b), information on why the trial ended or was stopped (14b), important harms or unintended effects in each group,¹⁹ analysis of performance errors and how errors were identified (19-i), and where the full trial protocol can be accessed.²⁴ Only one of the articles addressed the following criteria: information on which version of the AI algorithm was used (5-i), whether there was human–AI interaction in the handling of the input data and what level of expertise was required of users (5-iv), mechanism used to implement random allocation sequence,⁹ who generated random allocation sequence, who enrolled participants and who assigned participants to interventions,¹⁰ methods for additional analyses (12b), presentation of both absolute and relative effect sizes (17b), and

where and how AI intervention and/or its code can be accessed (25-i). None of the articles reported all of the items in the CONSORT-AI checklist.

Quality of evidence

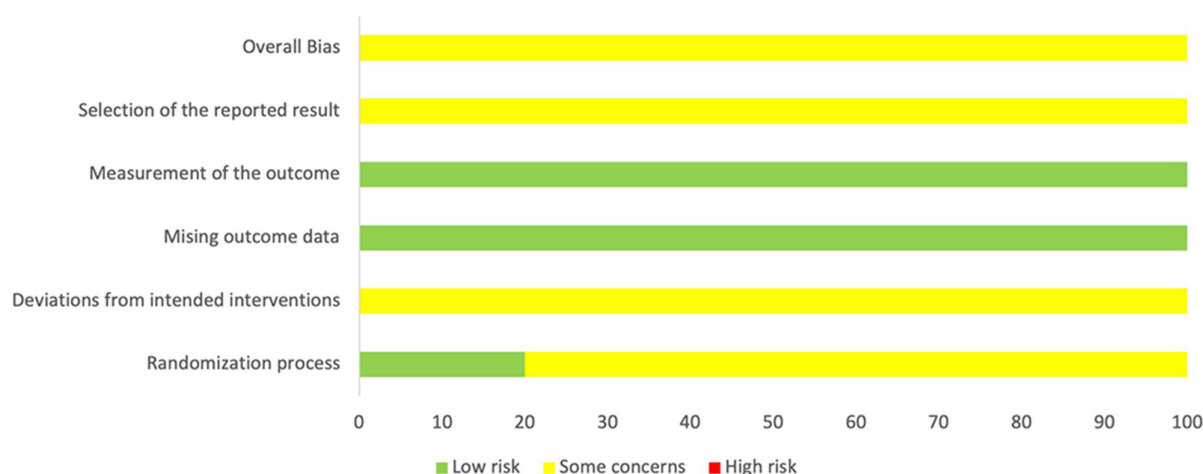
The results of the RoB-2 scoring are shown in figure 2. All included articles had an overall moderate risk of bias, with all articles having a score of ‘some concerns present’. All articles scored moderate risk for the domains of ‘selection of the reported result’ and ‘deviations from intended interventions’. All articles were scored as low risk for the domains of ‘measurement of the outcome’ and ‘missing outcome data’. For the domain of ‘randomisation process’, 80% of the articles were moderate risk and the remaining 20% were low risk. None of the articles scored high risk in any domains. The inter-rater concordance for RoB-2 scoring had a kappa score of 0.86.

DISCUSSION

Here, we aimed to evaluate the adherence of RCTs investigating the use of AI within ophthalmology to the guidelines set by CONSORT-AI checklist for reporting standards for RCTs. Our study found a total of five RCTs that evaluated AI applications in ophthalmology. These articles looked at the utility of AI in diabetic retinopathy screening,^{19 20} ophthalmologic education,²¹ detecting fungal keratitis²² and diagnosing childhood cataracts.²³ The mean CONSORT-AI score of the articles was 53% (range 37%–78%). None of the articles reported all items in the CONSORT-AI checklist, and all articles were rated as moderate risk, or ‘some concerns present’, through the RoB-2 tool assessment. All articles had moderate risk of bias for the ‘selection of the reported result’ and ‘deviations from intended interventions’ domains, and low risk of bias for ‘measurement of the outcome’ and ‘missing outcome data’ domains. Only one article had low risk of bias for their ‘randomisation process’, with the remainder having moderate risk in this domain.

The mean CONSORT score for our included studies (53%) is higher than mean score of 39% reported in the previous work by Yao *et al* in 2014 which reviewed the quality of reporting guidelines in 64 RCTs focused on ophthalmic surgery.¹² Aside from the difference in the number of reviewed articles, a potential reason for this difference in reported CONSORT-AI scores is that the articles found in our study are relatively new. The CONSORT-AI guidelines were published in 2020, and 3/5 of our articles were published in 2021 or later,^{19 20 22} which suggests that awareness of and adherence to reporting guidelines may have increased over time. Many of the items that the identified articles in our review failed to report on were also missed in studies identified by Yao *et al*.¹² These include determining adequate sample size (item 7), concern random allocation sequence generation (item 8) and its implementation (item 10).¹³ The low reporting rate of sample size calculation is a critical concern as this information is essential for protocol

A



B

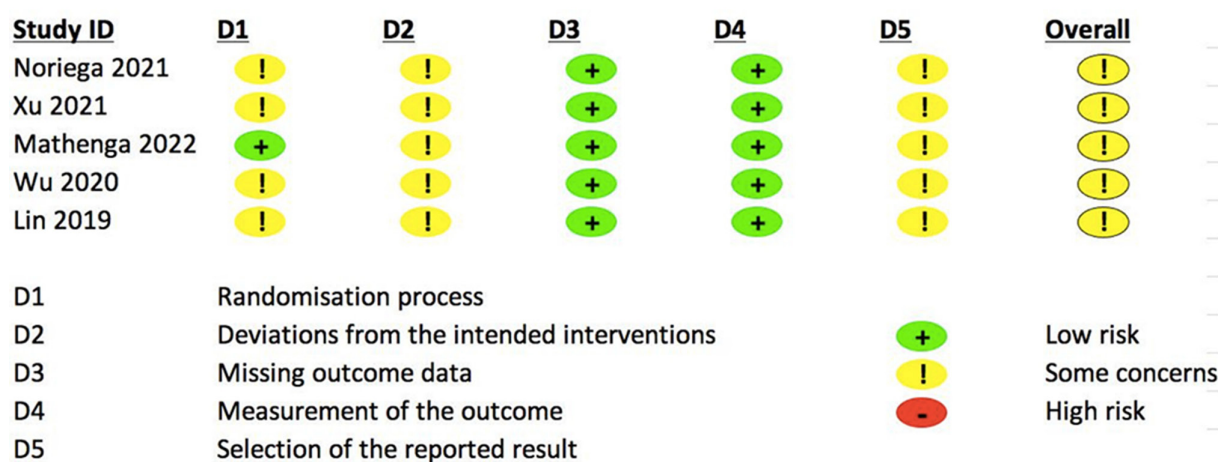


Figure 2 Risk of bias assessment using RoB2 tool for included studies displayed by means of a weighted plot for the distribution of the overall risk of bias within each bias domain (A) and traffic light plot of the risk of bias of each included clinical study (B).

development in all RCTs. There were some items that were commonly missed in Yao *et al* that were not missed in our reviewed articles, such as mentioning the term RCT in title or abstract (item 1),¹² which demonstrates the value in establishment of expected reporting standards by journals and publishing editors.

We observed some common trends in CONSORT-AI and RoB2 assessments in our study. For AI-based RCTs, it is difficult to blind both the physicians and participants to the intervention received, if the participants are humans and not images. For instance, if an RCT is comparing AI-based screening versus human-based screening, the participant may know whether they have been assigned to the AI or to a human at the time the intervention is given. One strategy to blind the participants, as seen in Noriega *et al*¹⁹ and Xu *et al*,²² is to replace human participants with human-derived data. Additionally, blinding the outcome assessors to the prescribed intervention is an important feature of the study design in RCTs, but in

three of the included studies in this review, Noriega *et al*,¹⁹ Xu *et al*,²² and Wu *et al*,²¹ did not outline these steps in their methods.

None of our included articles described where to find their initial trial protocol. Only one of the articles, by Lin *et al*, was registered on ClinicalTrials.gov.²³ This is a critical limitation as it could indicate a potential source of bias if analysis decisions were made after outcomes were measured which undermines the credibility of the RCT findings. Although outcome measurements were standard choices (eg, sensitivity and specificity for binary classification model performance), the role of an initial trial protocol cannot be overlooked as it is a key component of pretrial planning and study integrity. Furthermore, no articles other than Mathenga *et al* reported where the AI algorithm codes could be found.²⁰ This reduces transparency and may impede the reproducibility of the results as well as the progress of applying AI technologies. Siontis *et al* have found that AI RCTs across all healthcare

applications, not just ophthalmology, fail to provide the algorithm code for their AI tools.²⁴

Criteria 4b (settings and locations where data were collected), 15 (baseline demographics) and 21 (generalisability of trial findings) of the CONSORT-AI checklist were not perfectly adhered to in our five articles. Only three articles reported items 4b,^{20 22 23} three articles reported item 15^{20 21 23} and two articles reported item 21.^{20 22} Although these criteria were not the most frequently missed items, they are of utmost importance clinically, as they concern whether the results of the trial can be reasonably applied to a clinician's patient population. In a 2021 review of the development and validation pathways of AI RCTs, Siontis *et al* found that most AIs are not tested on datasets collected from patient populations outside of where the AI was developed and thus, it may be unsafe to apply these AIs to such populations.²⁴ In fact, using limited or imbalanced datasets both in development and validation stages may lead to discriminatory AI.²⁵ Therefore, special attention should be paid to these criteria.

In our review, we also found that the criterion for providing an explanation of any interim analyses and/or stopping guidelines if applicable (7b) was not reported across all articles. It could be argued that all RCTs should at least comment that an interim analysis was not planned, even if it was not applicable to the specific study design. Shahzad *et al* conducted a systematic review that also used CONSORT-AI to review the reporting quality of AI RCTs across all healthcare applications published between January 2015 and December 2021. They also found that item 7b was not reported in more than 85% of the included studies, and scored these items as non-applicable in their grading using CONSORT-AI.¹⁶

When analysing the appropriateness of analyses and the clarity of the performance assessments for each article, we found that each article chose suitable methods for their individual trials. Noriega *et al*, Xu *et al* and Lin *et al* evaluated the performance of their different comparators by calculating sensitivity and specificity among other metrics.^{19 22 23} Xu *et al* and Lin *et al* presented this information in the form of a table.^{22 23} Noriega *et al* and Xu *et al* also presented these results visually by plotting sensitivity and specificity of different comparators on a receiver operating curve which represented the performance of the AI alone.^{19 22} In Wu *et al*'s investigation of the effectiveness of AI-assisted problem based learning, ophthalmology clerks did a pre-lecture test and post-lecture test after either a traditional lecture or AI-assisted lecture.²¹ Improvement in test performance was assessed and compared between the two groups by analysing differences in the pre-lecture and post-lecture test scores using paired t-tests. A main source of bias in their study, not captured in the risk of bias assessment, is the quality of the test questions which were not made available to the readers. It is important to note that all AI-based RCTs identified in this study had no drop-outs, as all participants that enrolled in the RCT yielded valid data for analysis. This is due to the fact that

in some cases, images were subjects, in pre-collected databases and registries.

Despite the comprehensive search of the literature, a limited number of RCTs on AI were retrieved in the current study. The small number of RCTs identified prevented our study from conducting any temporal analyses or stratifying our analyses. In comparison, a literature review on the reporting guidelines of RCTs in ophthalmic surgery overall yielded 65 RCTs.¹² There are a couple of reasons that may explain the small number of RCTs investigating the efficacy of AI for ophthalmological applications. First, this small number may be an indication of the novelty of AI within the field of ophthalmology. Another reason may be the high costs and resources associated with RCTs. It is not feasible to conduct an RCT for all of the various AI tools developed for ophthalmology. Siontis *et al* found that the development and validation stages that different AI models go through before being evaluated in RCTs vary widely between papers.²⁴ The increasing number of standard guidelines for the reporting and quality assessment of AI, including DECIDE-AI,²⁶ PROBAST-AI,²⁷ QUADAS-AI,²⁸ STARD-AI²⁹ and TRIPOD-AI²⁷ are suggestive of the shift towards standardised assessment of AI tools. Another step that may aid in better assessment of AI tools in RCTs is determining performance metric thresholds that must be met at each stage of development and validation, although justifying these cutoffs may be difficult and subjective, and does not automatically imply high reliability for the RCT results.

CONCLUSIONS

AI is a growing field within ophthalmology that holds great promise for its applications in wide-reaching areas. Our findings suggest that there are a limited number of RCTs on applications of AI in ophthalmology, and adherence to some aspects of the 2020 CONSORT-AI reporting guidelines is suboptimal. It is essential that future trials provide information on protocol registration, a clear explanation for sample size calculations and details on the method of randomisation (i.e. type of randomisation, how it was implemented, who it was implemented by). Open access to the AI algorithm codes as well as further details about the software and version number used will enhance reproducibility of research efforts. Attention should be paid to blinding participants, physicians and the outcome assessors whenever possible. Finally, it is critical to report information that allows the readers to assess the generalisability of the trial results, such as baseline demographics of patients and settings where the trial data are collected.

It is recommended that future authors, funding organisations, peer-reviewers and others involved in the ophthalmological research process collaborate and place emphasis on adherence and integration of the CONSORT-AI checklist within the RCT development and publication process. This may facilitate the reproducibility of AI research which can in turn be a stimulus

for more AI-based RCTs and its clinical application in ophthalmology.

Author affiliations

¹Queen's University School of Medicine, Faculty of Health Sciences, Kingston, Ontario, Canada

²Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

³Department of Ophthalmology, University College London, London, UK

⁴Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Ontario, Canada

⁵University Health Network, Toronto, Ontario, Canada

Twitter Tina Felfeli @TinaFelfeli

Funding This research was in-part funded by Fighting Blindness Canada awarded to Dr. Tina Felfeli.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study. All data relevant to the study are included in the article or uploaded as online supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Niveditha Pattathil <http://orcid.org/0000-0002-7583-2640>

Tina Felfeli <http://orcid.org/0000-0002-0927-3086>

REFERENCES

- Ting DSW, Pasquale LR, Peng L, *et al*. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167–75.
- Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017;124:962–9.
- Brown JM, Campbell JP, Beers A, *et al*. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol* 2018;136:803–10.
- Ting DSW, Cheung CY-L, Lim G, *et al*. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;318:2211–23.
- Lee CS, Tying AJ, Deruyter NP, *et al*. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express* 2017;8:3440–8.
- Keel S, Li Z, Scheetz J, *et al*. Development and validation of a deep-learning algorithm for the detection of neovascular age-related macular degeneration from colour fundus photographs. *Clin Exp Ophthalmol* 2019;47:1009–18.
- Centre for Evidence-Based Medicine (CEBM). Oxford centre for evidence-based medicine: levels of evidence (March 2009). University of Oxford. Available: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009> [Accessed 17 Sep 2022].
- Begg C, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637–9.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191–4.
- Schulz KF, Altman DG, Moher D, *et al*. Statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:726–32.
- Sánchez-Thorin JC, Cortés MC, Montenegro M, *et al*. The quality of reporting of randomized clinical trials published in ophthalmology. *Ophthalmology* 2001;108:410–5.
- Yao AC, Khajuria A, Camm CF, *et al*. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. *Eye (Lond)* 2014;28:1341–9.
- Liu X, Rivera SC, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ* 2020;370:m3164.
- Varghese J. Artificial intelligence in medicine: chances and challenges for wide clinical adoption. *Visc Med* 2020;36:443–9.
- Sterne JAC, Savović J, Page MJ, *et al*. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.
- Shahzad R, Ayub B, Siddiqui MAR. Quality of reporting of randomised controlled trials of artificial intelligence in healthcare: a systematic review. *BMJ Open* 2022;12:e061519.
- Wang J, Wu S, Guo Q, *et al*. Investigation and evaluation of randomized controlled trials for interventions involving artificial intelligence. *Intell Med* 2021;1:61–9.
- Kothari R, Chiu C, Moukheiber M, *et al*. A descriptive appraisal of quality of reporting in a cohort of machine learning studies in anesthesiology. *Anaesth Crit Care Pain Med* 2022;41:101126.
- Noriega A, Meizner D, Camacho D, *et al*. Screening diabetic retinopathy using an automated retinal image analysis system in independent and assistive use cases in Mexico: randomized controlled trial. *JMIR Form Res* 2021;5:e25290.
- Mathenge W, Whitestone N, Nkurikiye J, *et al*. Impact of artificial intelligence assessment of diabetic retinopathy on referral service uptake in a low-resource setting: the RAIDERS randomized trial. *Ophthalmol Sci* 2022;2:100168.
- Wu D, Xiang Y, Wu X, *et al*. Artificial intelligence-tutoring problem-based learning in ophthalmology clerkship. *Ann Transl Med* 2020;8:700.
- Xu F, Jiang L, He W, *et al*. The clinical value of explainable deep learning for diagnosing fungal Keratitis using in vivo confocal microscopy images. *Front Med* 2021;8:797616.
- Lin H, Li R, Liu Z, *et al*. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine* 2019;9:52–9.
- Siontis GCM, Sweda R, Noseworthy PA, *et al*. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. *BMJ Health Care Inform* 2021;28:e100466.
- Leslie D, Mazumder A, Peppin A, *et al*. Does "AI" stand for augmenting inequality in the era of COVID-19 healthcare. *BMJ* 2021;372:n304.
- DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med* 2021;27:186–7.
- Collins GS, Dhiman P, Andaur Navarro CL, *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:e048008.
- Sunderajah V, Ashrafian H, Rose S, *et al*. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nat Med* 2021;27:1663–5.
- Sunderajah V, Ashrafian H, Golub RM, *et al*. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open* 2021;11:e047709.
- Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.