

Reporting guidelines for artificial intelligence in healthcare research

Ibrahim, Hussein; Liu, Xiaoxuan; Denniston, Alastair K

DOI:

[10.1111/ceo.13943](https://doi.org/10.1111/ceo.13943)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Ibrahim, H, Liu, X & Denniston, AK 2021, 'Reporting guidelines for artificial intelligence in healthcare research', *Clinical and Experimental Ophthalmology*, vol. 49, no. 5, pp. 470-476. <https://doi.org/10.1111/ceo.13943>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.


When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Reporting guidelines for artificial intelligence in healthcare research

Hussein Ibrahim MBChB^{1,2,3}  | Xiaoxuan Liu MBChB^{1,2,3,4,5}  |
Alastair K. Denniston PhD^{1,2,3,4,5,6,7} 

¹Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK

²Department of Ophthalmology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

³Centre for Regulatory Science and Innovation, Birmingham Health Partners, Birmingham, UK

⁴NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and University College London, London, UK

⁵Health Data Research UK, London, UK

⁶Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, UK

⁷National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, University College London Institute of Ophthalmology, London, UK

Correspondence

Alastair K. Denniston, Academic Unit of Ophthalmology, Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK.
Email: a.denniston@bham.ac.uk

Abstract

Reporting guidelines are structured tools developed using explicit methodology that specify the minimum information required by researchers when reporting a study. The use of artificial intelligence (AI) reporting guidelines that address potential sources of bias specific to studies involving AI interventions has the potential to improve the quality of AI studies, through improvements in their design and delivery, and the completeness and transparency of their reporting. With a number of guidance documents relating to AI studies emerging from different specialist societies, this Review article provides researchers with some key principles for selecting the most appropriate reporting guidelines for a study involving an AI intervention. As the main determinants of a high-quality study are contained within the methodology of the study design rather than the intervention, researchers are recommended to use reporting guidelines that are specific to the study design, and then supplement them with AI-specific guidance contained within available AI reporting guidelines.

KEYWORDS

artificial intelligence, checklist, guidelines, machine learning, research design, research report

1 | INTRODUCTION

Artificial intelligence (AI) is one of the most exciting and rapidly evolving areas of healthcare research in the 21st century. The number of AI-related academic publications has

risen exponentially in recent years and scientific advances have revealed many potential AI healthcare applications.¹ Examples are wide-ranging and include AI algorithms for screening and triage,²⁻⁴ diagnosis,⁵⁻⁷ prognostication,^{8,9} decision support¹⁰ and treatment recommendation.¹¹

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Clinical & Experimental Ophthalmology* published by John Wiley & Sons Australia, Ltd on behalf of Royal Australian and New Zealand College of Ophthalmologists.



Naturally, the potential for AI to transform healthcare – by, for example, offering earlier and more accurate diagnoses, providing novel insights for the understanding of diseases, and enabling more efficient service delivery – has generated an enormous amount of excitement amongst patients, the public, politicians, and healthcare professionals. Concerns have been raised, however, that the potential impact of AI on healthcare may be overhyped.¹²

Two robust systematic reviews and meta-analyses of AI medical imaging studies have confirmed that concerns raised regarding AI hype were well-founded.^{13,14} Both of the reviews revealed that poor study design, delivery, and reporting were endemic in the field,^{13,14} and, in fact, one of the reviews showed that <1% of the 20 000 AI medical imaging studies included were of sufficient quality to provide a trustworthy evaluation of the AI algorithm versus a human reader.¹³

There has since been a collective response calling for better design, delivery and reporting of AI studies, and guidance and tools to support this. Reporting guidelines are tools that specify the minimum information required when reporting a study.¹⁵ The use of AI reporting guidelines has the potential to improve the quality of such studies, through improvements in their design and delivery, and the completeness and transparency of their reporting.

The speciality of ophthalmology is at the forefront of AI healthcare research, with notable advances including the use of machine learning techniques such as deep learning to diagnose diabetic retinopathy,¹⁶ detect papilloedema¹⁷ and predict cardiovascular risk factors using fundus photographs.¹⁸ In fact, IDx-DR, an AI system for diabetic retinopathy screening, was the first U.S. Food and Drug Administration-approved diagnostic AI algorithm.¹⁶ As such, the speciality of ophthalmology has an opportunity to lead by example by adopting and endorsing AI reporting guidelines.

This Review article provides an overview of AI reporting guidelines and their application in healthcare research to help researchers across all medical specialities, not only in the speciality of ophthalmology, improve the design, delivery, reporting, and ultimately, the quality of their work in the AI era.

2 | WHAT ARE REPORTING GUIDELINES?

Reporting guidelines are structured tools developed using explicit methodology that specify the minimum information required by researchers when reporting a study.¹⁵ The aims of reporting guidelines are to ensure that studies can be understood by readers and reviewers, replicated by other researchers, used by healthcare professionals to make clinical decisions and included in systematic reviews and meta-analyses.¹⁵

By highlighting key considerations relevant to different study types and demonstrating “what good looks like,”

reporting guidelines help improve the design, delivery and, ultimately, the quality of studies; and, by providing a clear list of minimum content that should appear in a paper, they help improve the completeness and transparency of their reporting, making areas of potential bias more visible and thus enabling more effective evaluation of studies.

The EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network is an international initiative that seeks to improve the quality of healthcare research by promoting the development and use of robust reporting guidelines.¹⁹ The network provides Toolkits to support the development, selection and use of reporting guidelines;²⁰ and has a Library that contains an up-to-date collection of reporting guidelines.²¹

3 | WHY DO WE NEED AI-SPECIFIC REPORTING GUIDELINES?

The reporting guidelines developed with and published by the EQUATOR Network are organised according to study type (i.e. separate guidelines for clinical trials, diagnostic accuracy studies, observational studies etc.). This promotes a consistent approach for addressing the same study type, regardless of the speciality area, and indeed across most types of interventions. It has been recognised, however, that specific interventions (e.g. social and psychological interventions),²² outcomes (e.g. patient-reported outcomes)^{23,24} and scenarios (e.g. cluster trials)²⁵ require specific extensions to these reporting guidelines. For example, recognition of the potential sources of bias specific to studies involving social and psychological interventions has led to the development of social and psychological intervention-specific extensions to existing EQUATOR reporting guidelines.²² In the same way, recognition of the potential sources of bias specific to studies involving AI interventions has led to the development of AI-specific extensions to existing EQUATOR reporting guidelines.

In parallel to the work of the EQUATOR Network, a number of experts and institutions have also developed their own recommendations for reporting AI studies. Unlike the AI reporting guidelines developed with and published by the EQUATOR Network, which start with study design and focus on clinical evaluation, these start with the intervention – AI – and usually have a broader scope, including algorithm development, data transparency, ethical standards and utility.

4 | WHICH AI REPORTING GUIDELINES SHOULD I USE FOR MY STUDY?

All research studies, including those involving AI interventions, should be reported using the most appropriate



reporting guidelines available. The range of different reporting guidelines that have been proposed in the field of AI can, however, make it difficult for researchers to determine which AI reporting guidelines to use for their study. Whilst researchers may be primarily interested in the intervention, particularly a complex intervention such as an AI system, they should bear in mind that it is only by considering every other component of a study that they can be confident that it will provide a trustworthy evaluation of the intervention.

Most of the factors that determine the quality of a study are not specific to the intervention, but are related to the study design. The EQUATOR Network has developed reporting guidelines based around the study design, which are deliberately “speciality neutral” and, indeed, “intervention neutral,” except where an intervention has distinct attributes or risks of bias that require additional explicit reporting requirements. Fundamentally, whilst there may be some distinct characteristics of AI interventions, clinical evaluation for AI should not be overly “exceptionalised,” but should use well-established methodology including good study design, delivery and reporting, as would be undertaken for other health technologies.

5 | WHICH AI REPORTING GUIDELINES SHOULD I USE FOR A CLINICAL TRIAL?

The strongest evidence for the safety, clinical effectiveness and cost-effectiveness of an AI intervention requires evaluation in the context of one or more well-designed, well-delivered and well-reported clinical trials.²⁶ This is because two fundamental characteristics of clinical trials, namely randomisation and a control arm, allow researchers to make causal inferences about interventions and their outcomes.²⁶

Although most AI interventions have not yet been evaluated in clinical trials, this is likely to be an area of rapid expansion as the field evolves. Such studies are particularly important because they will potentially be a key part of the evidence that regulators, payers and policymakers use to decide whether an AI intervention is sufficiently safe and effective to be approved and commissioned for use.

The risk to patients and the public of an AI algorithm being approved and commissioned for clinical use based on potentially incomplete information highlighted the need for reporting guidelines specific to AI interventions. To address this, the SPIRIT-AI and CONSORT-AI Steering Group announced in October 2019 an initiative to develop the first reporting guidelines for clinical trials involving AI interventions.²⁷ This work was undertaken

with the EQUATOR network, and formed AI-specific extensions to their standard reporting guidelines that are in widespread use and mandated by many leading journals.

The AI-specific extensions to the SPIRIT 2013²⁸ (Standard Protocol Items: Recommendations for Interventional Trials) and CONSORT 2010²⁹ (Consolidated Standards of Reporting Trials) guidelines for clinical trial protocols and reports were developed in accordance with the EQUATOR Network framework using a Delphi methodology with an international multidisciplinary consortium. The SPIRIT-AI³⁰⁻³² and CONSORT-AI³³⁻³⁵ guidelines were co-published in September 2020, representing the first global standards for reporting AI studies.

The SPIRIT-AI³⁰⁻³² and CONSORT-AI³³⁻³⁵ guidelines are extensions to, rather than replacements for, the original SPIRIT 2013²⁸ and CONSORT 2010²⁹ guidelines for clinical trial protocols and reports; and contain 15 (12 extensions, 3 elaborations) and 14 (11 extensions, 3 elaborations) new items, respectively, that should be routinely reported by researchers in addition to the 33 core SPIRIT 2013²⁸ and 25 core CONSORT 2010²⁹ items.

Elements of clinical trials of AI interventions that require detailed and specific reporting according to the SPIRIT-AI³⁰⁻³² and CONSORT-AI³³⁻³⁵ guidelines are similar across both clinical trial protocols and reports. Examples include, but are not limited to: the algorithm version; the procedure for acquiring, selecting and pre-processing the input data; and the criteria for inclusion at the level of the input data in addition to the level of participants.

For example, in a clinical trial evaluating an AI intervention for diagnosing diabetic retinopathy using fundus photographs, researchers should clearly specify which version of the AI algorithm they intend on using in the clinical trial protocol; and then state which version of the AI algorithm they actually used, and whether the algorithm version changed throughout the course of the trial, in the clinical trial report.

Researchers should also clearly and comprehensively outline their proposed plan for acquiring, selecting and pre-processing the fundus photographs prior to analysis by the AI algorithm in the clinical trial protocol; and then describe how the fundus photographs were actually acquired, selected and pre-processed prior to analysis by the AI algorithm in the clinical trial report.

Additionally, the eligibility criteria should be clearly and comprehensively reported by researchers at both the level of participants, such as patient age, and input data, such as fundus photograph image quality. This is important as it enables reviewers to differentiate between AI interventions that only work in ideal conditions and those



that are more robust and suitable for real-world settings, such as large scale, national screening programmes.

6 | WHICH AI REPORTING GUIDELINES SHOULD I USE FOR A DIAGNOSTIC ACCURACY STUDY?

Diagnostic AI algorithms for detecting diseases, such as IDx-DR for diabetic retinopathy,¹⁶ promise to achieve diagnostic accuracies similar to those provided by expert clinicians whilst simultaneously reducing healthcare-resource use.³⁶ At present, a significant proportion of potential AI healthcare applications are diagnostic AI algorithms, but much of the evidence supporting their use has been disseminated in the absence of AI-specific reporting guidelines.³⁶

In terms of study design, diagnostic accuracy studies should be reported according to the STARD 2015³⁷ (Standards for Reporting Diagnostic Accuracy Studies) guidelines. An AI-specific extension, STARD-AI,³⁶ is, at the time of writing, under development for use alongside the STARD 2015³⁷ guidelines. Other AI-specific guidelines such as MI-CLAIM³⁸ may have particular value here, addressing important AI-specific elements that are not currently covered by the STARD 2015³⁷ guidelines.

Reporting of the non-AI components of the design and delivery of diagnostic accuracy studies should continue to adhere to the original STARD 2015³⁷ guidelines. They are designed to ensure that important information relevant to the design and delivery of diagnostic accuracy studies, including information relating to the participants (e.g. how, where and when potentially eligible participants were identified and what criteria were used to establish their eligibility), test methods (e.g. details of, rationale for choosing and order of the index test and reference standard) and data analysis (e.g. how indeterminate test results and missing data were handled) are clearly and comprehensively reported.

7 | WHICH AI REPORTING GUIDELINES SHOULD I USE FOR A PREDICTION MODEL STUDY?

Clinical prediction models estimate the likelihood of an individual having (diagnostic) or developing (prognostic) disease using predictor variables (risk factors such as age, sex and biomarkers).³⁹ The ability of AI to analyse large and complex datasets of predictor variables has led to the development of several potential AI prediction models, such as AI algorithms for predicting conversion to wet age-related macular degeneration.⁸

The most widely accepted EQUATOR reporting guidelines for prediction model studies are the TRIPOD 2015⁴⁰ (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) guidelines. Many aspects of these guidelines are applicable to AI prediction models, but despite this, for various reasons – such as differences in terminology, differences in the statistical basis on which AI prediction models and non-AI prediction models are built, and the fact that predictors are often hidden in the “black box” of AI prediction models – their use for disseminating AI prediction model studies has been poor.⁴¹ To address this, an AI-specific extension, TRIPOD-AI,⁴¹ is, at the time of writing, under development for use alongside the TRIPOD 2015⁴⁰ guidelines.

Reporting of the non-AI elements of the design and delivery of prediction model studies should continue to adhere to the original TRIPOD 2015⁴⁰ guidelines. They are designed to ensure that important information relevant to the design and delivery of prediction model studies, including information relating to the participants (e.g. the eligibility criteria, participant characteristics and the flow of participants through the study), outcomes (e.g. definition of the outcome that is predicted by the prediction model including how and when it was assessed), predictors (e.g. definition of the all the predictors used by the prediction model including how and when they were assessed) and model (e.g. how the prediction model was developed, how the prediction model was used and how the prediction model performed) are completely and transparently reported.⁴⁰ As with diagnostic accuracy studies, these guidelines may be supplemented by the valuable AI-specific guidance found in guidelines such as MI-CLAIM.³⁸

8 | WHAT OTHER AI-SPECIFIC GUIDELINES SHOULD I BE AWARE OF?

The MI-CLAIM³⁸ (Minimum Information about CLinical Artificial Intelligence Modelling) guidelines are EQUATOR reporting guidelines that were published in September 2020 to improve the reporting of information regarding clinical AI algorithms. The guidelines are designed to inform readers and users about an AI algorithm by ensuring that information about how it was developed and validated is clearly and comprehensively reported. They are split into six parts: (1) study design; (2) separation of data into partitions for model training and model testing; (3) optimisation and final model selection; (4) performance evaluation; (5) model examination and (6) reproducible pipeline. The guidelines are distinct

relative to the AI reporting guidelines discussed above in that there is considerable attention given to the reporting of information relating to the AI algorithm itself rather than how it was used in the context of a specific study. In this way, their use may provide value alongside the use of AI-specific extensions such as the SPIRIT-AI,³⁰⁻³² CONSORT-AI,³³⁻³⁵ STARD-AI³⁶ and TRIPOD-AI⁴¹ guidelines.

The MINIMAR⁴² (MINimum Information for Medical AI Reporting) guidelines are non-EQUATOR reporting guidelines published in June 2020 by the American Medical Informatics Association. The guidelines are designed for studies reporting the use of AI systems in healthcare. Their purpose is to ensure that the minimum information required to adequately understand the intended predictions, target populations and potential biases of an AI algorithm are clearly and comprehensively reported. Unlike other reporting guidelines, which provide a checklist of items that require reporting by researchers, these guidelines provide suggestions for reporting information across four key areas of clinical AI studies: (1) study population and setting; (2) patient demographic characteristics; (3) model architecture and (4) model evaluation. There is overlap between the MINIMAR⁴² guidelines and the MI-CLAIM³⁸ guidelines which both focus on AI algorithms and how they were developed and validated.

The DECIDE-AI⁴³ (Developmental and Exploratory Clinical Investigation of DEcision-support systems driven by Artificial Intelligence) guidelines are EQUATOR reporting guidelines that, at the time of writing, are under development. These guidelines are distinct in that their intended purpose is to improve the evaluation and reporting of human factors in clinical AI studies. The DECIDE-AI⁴³ guidelines will address the essential role that human factors will have in how a clinical AI algorithm performs; this also brings in the important distinction between AI assessed in isolation versus an AI intervention assessed as part of an AI/human system; commentators sometimes refer to the potentially improved performance over the human system as “augmented intelligence.” The guidelines are intended to be used in early-stage, small-scale clinical studies of AI interventions, when the intervention itself and the human-machine interaction may still be being refined prior to fuller evaluation. Such studies, if conducted, would take place after development and technical validation (in diagnostic accuracy studies or prediction model studies that are covered by the STARD-AI³⁶ and TRIPOD-AI⁴¹ guidelines respectively), but before clinical validation (in clinical trials that are covered by the SPIRIT-AI³⁰⁻³² and CONSORT-AI³³⁻³⁵ guidelines). By ensuring that adequate attention is placed on the human-AI interaction during the development and evaluation of clinical

AI algorithms at this stage of the translational pipeline, the developers of the DECIDE-AI⁴³ guidelines argue that their use will ultimately enable more efficient translation of AI algorithms from code to clinic.

9 | CONCLUSION

This Review article has provided an overview of AI reporting guidelines and their application in healthcare research. The article should help researchers across all medical specialities, including in the speciality of ophthalmology, better understand, select, and use AI reporting guidelines. In this way, this article should help researchers improve the design, delivery, reporting, and, ultimately, the quality of their work in the AI era.

Ultimately, the impact of AI-specific reporting guidelines on improving the quality of AI healthcare research is determined largely by the extent to which researchers use them when reporting studies, medical journal editors require authors to use them when submitting studies, and reviewers use them when appraising studies. As a speciality at the forefront of AI healthcare research, ophthalmology can lead by example by adopting and endorsing their use.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Hussein Ibrahim  <https://orcid.org/0000-0001-6182-065X>

Xiaoxuan Liu  <https://orcid.org/0000-0002-1286-0038>

Alastair K. Denniston  <https://orcid.org/0000-0001-7849-0087>

REFERENCES

1. State of AI Report. *State of AI Report 2020*. 2020. <https://www.stateof.ai/>. Accessed April 2021.
2. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577:89-94.
3. Abramoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investig Ophthalmol Vis Sci*. 2016;57:5200-5206.
4. Bellemo V, Lim ZW, Lim G, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digital Health*. 2019;1:e35-e44.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115-118.
6. Nagpal K, Foote D, Liu Y, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med*. 2019;2:48.



7. Huang S-C, Kothari T, Banerjee I, et al. PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. *NPJ Digit Med.* 2020; 3:61.
8. Yim J, Chopra R, Spitz T, et al. Predicting conversion to wet age-related macular degeneration using deep learning. *Nat Med.* 2020;26:892-899.
9. Kim H, Goo JM, Lee KH, Kim YT, Park CM. Preoperative CT-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. *Radiology.* 2020;296: 216-224.
10. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut.* 2019;68:1813-1819.
11. Tyler NS, Mosquera-Lopez CM, Wilson LM, et al. An artificial intelligence decision support system for the management of type 1 diabetes. *Nat Metab.* 2020;2:612-619.
12. Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Med.* 2019;17:143.
13. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health.* 2019;1:e271-e297.
14. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ.* 2020;368: m689.
15. EQUATOR Network. *What is a reporting guideline?* <https://www.equator-network.org/about-us/what-is-a-reporting-guideline/>. Accessed April 2021.
16. Savoy M. IDx-DR for diabetic retinopathy screening. *Am Fam Physician.* 2020;101:307-308.
17. Milea D, Najjar RP, Jiang Z, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N Engl J Med.* 2020;382:1687-1695.
18. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* 2018;2:158-164.
19. EQUATOR Network. EQUATOR Network: what we do and how we are organised. <https://www.equator-network.org/about-us/equator-network-what-we-do-and-how-we-are-organised/>. Accessed April 2021.
20. EQUATOR Network. Toolkits. <https://www.equator-network.org/toolkits/>. Accessed April 2021.
21. EQUATOR Network. Library for health research reporting. <https://www.equator-network.org/library/>. Accessed April 2021.
22. Montgomery P, Grant S, Mayo-Wilson E, et al. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. *Trials.* 2018;19:407.
23. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the SPIRIT-PRO extension. *JAMA.* 2018;319:483-494.
24. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.* 2013;309:814-822.
25. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ.* 2012;345:e5661.
26. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? *BMJ.* 1998;316:201.
27. The CONSORT-AI and SPIRIT-AI Steering Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med.* 2019;25:1467-1468.
28. Chan A-W, Tetzlaff JM, Gøtzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ.* 2013;346:e7586.
29. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.
30. Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med.* 2020;26:1351-1363.
31. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, The SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ.* 2020;370:m3210.
32. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, The SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digital Health.* 2020;2:e549-e560.
33. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, The SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26:1364-1374.
34. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, The SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ.* 2020; 370:m3164.
35. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, The SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digital Health.* 2020;2:e537-e548.
36. Sounderajah V, Ashrafian H, Aggarwal R, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI Steering Group. *Nat Med.* 2020;26:807-808.
37. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
38. Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med.* 2020;26:1320-1324.
39. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10:e1001381.
40. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594.

41. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet*. 2019;393:1577-1579.
42. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc*. 2020;27:2011-2015.
43. The DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med*. 2021;27:186-187.

How to cite this article: Ibrahim H, Liu X, Denniston AK. Reporting guidelines for artificial intelligence in healthcare research. *Clin Experiment Ophthalmol*. 2021;49:470–476. <https://doi.org/10.1111/ceo.13943>