PRISMA AI- Reporting Guidelines for Systematic Reviews and Meta-Analyses on AI in Healthcare

Giovanni E. Cacciamani[1,2,3,4,5*], Timothy N. Chu[1,2,3], Daniel I. Sanford[1,2,3], Andre Abreu[1,2,3,4,5], Vinay Duddalwar[3,6], Assad Oberai[7,8], C-C Jay Kuo[9], Xiaoxuan Liu[10,11,12], Alastair K Denniston[10,11,12,13,14], Baptiste Vasey[15,16], Peter McCulloch[15], Robert F. Wolff[17], Sue Mallett[18], John Mongan[19,20], Charles E. Kahn Jr[21], Viknesh Sounderajah[22], Ara Darzi[22], Philipp Dahm[23], Karel GM Moons[24], Eric Topol[25], Gary S. Collins[26], David Moher[27], Inderbir S. Gill[1,2,3,4,5], Andrew J. Hung[1,2,5]

1. USC Institute of Urology and Catherine and Joseph Aresty Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
2. AI Center at USC Urology, USC Institute of Urology, University of Southern California, Los Angeles, CA, USA.
3. Department of Radiology, University of Southern California, Los Angeles, CA, USA
4. Center for Image-Guided and Focal Therapy for Prostate Cancer, Institute of Urology and Catherine and Joseph Aresty Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
5. Norris Comprehensive Cancer Center Institute of Urology Keck School of Medicine of the University of Southern California Los Angeles California USA
6. USC Radiomics Laboratory, Keck School of Medicine, Department of Radiology, University of Southern California, Los Angeles, CA, USA
7. Department of Aerospace and Mechanical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
8. Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
9. Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA
10. University Hospitals Birmingham NHS Foundation Trust
11. Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, UK
12. Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, UK
13. NIHR Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK
14. Health Data Research, London, UK
15. Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
16. Department of Surgery, Geneva University Hospital, Geneva, Switzerland
17. Kleijnen Systematic Reviews Ltd, Escrick, York, UK
18. Centre for Medical Imaging, University College London, UK
19. Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA
20. Center for Intelligent Imaging, University of California San Francisco, San Francisco, CA, USA
21. Department of Radiology and Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA.
22. Institute of Global Health Innovation, Imperial College London, London, United Kingdom
23. Minneapolis VAMC, Urology Section and University of Minnesota, Department of Urology, Minneapolis, MN, USA
24. Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht University, Utrecht, The Netherlands
25. Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA.
26. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK.
27. Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute.

Wordcount: 998
Figures: 1
Tables: 1
References: 10

***Giovanni E. Cacciamani, MSc, MD, FEBU**
Assistant Professor of Research Urology
Assistant Professor of Research Radiology
Co-Director Artificial Intelligence Center at USC Urology
USC Institute of Urology Catherine and Joseph Aresty Department of Urology
Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
Email: Giovanni.cacciamani@med.usc.edu
Phone: +1 (626) 4911531

Systematic reviews and meta-analyses play a critical role in guiding clinical practice at the point of care, as well as in the formulation of clinical practice guidelines and health policy[1,2]. There are three essential components to an impactful systematic review. Firstly, the design of a study should be based upon a robust research question and search strategy. Secondly, minimization of bias should be enhanced through use of quality assessment tools and study design-specific eligibility criteria. Thirdly, reporting of results should be transparently conducted through adherence to expert-derived reporting items. Thousands of systematic reviews, including meta-analyses, are produced annually with an increasing proportion reporting on artificial intelligence (AI) interventions in healthcare.. With this rapid expansion, there is a need for reporting guidelines tailored to AI [3-7] that will support high-quality, reproducible, and clinically relevant systematic reviews.

AI is rapidly integrating into society and in medicine. A literature search of studies referencing AI in healthcare over the past ten years returned more than 80,000 published studies. Given that interest in AI is reaching an all-time high, there arise new concerns on the quality of these studies, including a lack of: clear explainability of how AI algorithms function; strong evidence of effectiveness in clinical settings; and standardized reporting within primary studies. Efforts have been made to improve understanding of this technology to allow for critical appraisal of AI interventions and to reduce inconsistencies in how studies are structured, as well as reporting of data, methods and results.[3,5,7]. As systematic reviews on AI interventions increase, so does the importance of transparency and reproducibility of the reported data**.**

The most accepted guideline for reporting systematic reviews and meta-analyses is the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. This evidence-based guideline was first published in 2009[2] and served as a set of minimum reporting items that should be addressed when describing a systematic review or meta-analysis. In 2021, updated PRISMA guidelines were published reflecting new evidence and mandates in the years between PRISMA 2009 and its recent update[1]. PRISMA and its extensions are widely accepted by journals and guideline panels. Their direct applicability for particular topics, such as AI interventions, may benefit from inclusion of specific requirements that capture all the nuances particular to these studies.

An ongoing umbrella review (a review of reviews) found that nearly 7000 reviews (systematic and non-systematic) on AI in the category “medicine” have been published. This number is destined to

grow: a search in PROSPERO shows that the number of ongoing systematic reviews on AI is about 1300. Since 2013 there has been an increasing trend in systematic review publications over time with an annual average percent change of +52%(Figure 1a). However, only 20% specified in the title or abstract that they were “systematic” or reported the term “PRISMA” in the abstract. Despite widespread acceptance of PRISMA-guidelines[1], they seem to be under-utilized for AI-based intervention reviews. Nevertheless, review articles reporting “systematic” or “PRISMA” in their title or abstract have been cited more frequently than the other reviews that do not use these terms. (Figure 1b). Furthermore, the most cited review article (over 6000 citations in 5 years) does not cite any reporting guidelines, emphasizing the need for more stringent reporting protocols.

Several reporting guidelines have been updated due to pioneering efforts made by the SPIRIT-AI[7] and CONSORT-AI[3] extensions committees to ensure applicability to clinical studies and reporting involving AI. These efforts are relevant due to the number of active clinical studies involving AI listed on clinicaltrials.gov. In 2020, a search of studies that included “Artificial Intelligence” or “machine learning”, or “deep learning” found around 300 active studies[8]. A search today returns a 7-fold increase with over 2100 active studies on October1st, 2022. The SPIRIT and CONSORT working groups received expert consultation and underwent a Delphi process to provide a consensus-based guideline. These efforts resulted in the CONSORT-AI[3] and SPIRIT-AI[7] extensions which identify additional reporting items that were deemed essential for clinical studies involving AI. Other AI extensions of existing reporting guidelines, de novo guidelines, or bias reporting tools include: Prediction model Risk Of Bias ASsessment Tool (PROBAST-AI) , quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies (QUADAS-AI)[9], checklist for artificial intelligence in medical imaging (CLAIM)[4], and early-stage clinical evaluation of decision support systems driven by artificial intelligence (DECIDE-AI)[6]. Guidance including Standards for Reporting of Diagnostic Accuracy Studies (STARD-AI)[5], and Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD-AI), will shortly be available to cover study designs for diagnostic accuracy studies and clinical prediction model studies, respectively. These efforts highlight the importance of extended frameworks for systematic reviews that are reporting on AI in healthcare.

The PRISMA-AI Steering Committee has begun the process of creating an AI- implementation of PRISMA guidelines and extensions[1] for studies addressing AI-based interventions (Table 1). Our efforts include registering our extension with clinicaltrials.gov (NCT05382455) and EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network as a guideline under development. Clinicians, researchers, statisticians, computer scientists, engineers, methodologists designing clinical trials, systematic reviewers, patients, journal editors, published AI-extensions contributors, and trialists with interest in AI-related to healthcare are being recruited to establish a community collaboration that will create reporting guidelines for systematic reviews and meta-analyses that include AI. The EQUATOR Network's development guidance[10] will be used by the PRISMA-AI team to establish a consensus for the framework of reporting guidelines on AI in systematic reviews and meta-analyses.

The PRISMA-AI implementations will reflect the most pertinent technical details needed for reproducibility, focusing on requirements to critically follow and authenticate the methods used in the systematic reviews and meta-analyses relating to outcomes, risk of bias, and applicability. The PRISMA-AI implementations will assist stakeholders interested in utilizing AI-related information in systematic reviews by creating a framework for reviewers that evaluate the quality of the data reported in publications, deliver a tool for training researchers on AI methodology, and support end-users of systematic reviews such as clinicians, researchers, patients, and policymakers to better estimate the validity and applicability of findings from systematic reviews in their decision-making process.

Table 1 | Process for development of PRISMA-AI

Step	Process	Goal(s)
1. Literature Review	An umbrella review (ongoing) is conducted to examine the quality of reporting systematic reviews with/without meta-analysis, scoping reviews, DTA meta-analysis, protocols	a) Identification of variability in the assessment and reporting. b) Identification of list editable and additional items (first set).

2. Delphi Survey	Delphi Survey among multi-specialty experts (in each medical specialty) who have already published about AI applications in leading medical journals and the lead authors of PRISMA, STARD-AI, CONSORT-AI, SPIRIT-AI, TRIPOD-AI, PROBAST-AI, CLAIM-AI and DECIDE-AI to ensure that the criteria have global applicability in all the disciplines and for each type of study which involves the AI.	c) Identification of editable and adjunctive items (second set).
3. Consensus meeting	Establishment of consensus for approval of the PRISMA-AI implementations	d) creation of the PRISMA-AI implementation criteria Checklist
4. Piloting	Piloting of PRISMA-AI implementation in abroad range of users	
5. Checklist and statement document	Publication of the PRISMA-AI Implementations Checklist and the explanation and elaboration document	
6. Dissemination	Dissemination for the PRISMA-AI Implementation through media campaigns and implementation of the PRISMA-AI Checklist on the PRISMA website	

Competing interests

PD serves as coordinating editor of Cochrane Urology. G.S.C. is the director of the UK EQUATOR Centre. D.M. is the director of the Canadian EQUATOR Centre. XL is an industry fellow (observer) with Hardian Health. AJH is a consultant for Intuitive. ISG is a consultant for STEBA. JM is a consultant for Siemens. CEK receives salary support as Editor of Radiology: Artificial Intelligence. VAD is a consultant for Radmetrix, Inc and Westat Inc. and is an Advisory Board Member to Deeptek, Inc. AD is Chair of the Health Security initiative at Flagship Pioneering UK Ltd. PD serves as coordinating editor of Cochrane Urology. G.S.C. is the director of the UK EQUATOR Centre. D.M. is the director of the Canadian EQUATOR Centre.

REFERENCES

1. Page, M.J.*, et al. Systematic Reviews* **10**, 89 (2021).
2. Moher, D. *et al*. PRISMA statement. *Epidemiology* **22**, 128 (2011).
3. Liu, X.*, et al. Nature Medicine* **26**, 1364-1374 (2020).
4. Mongan, J.,*et al*. *Radiology: Artificial Intelligence* **2**, e200029 (2020).
5. Sounderajah, V.*, et al. Nat Med* **26**, 807-808 (2020).
6. Vasey, B.*, et al. Nat Med* **28**, 924-933 (2022).
7. Cruz Rivera, *et al*. *Nat Med* **26**, 1351-1363 (2020).
8. The CONSORT-AI and SPIRIT-AI Steering Group. *Nat Med* **25**, 1467-1468 (2019).
9. Sounderajah, V.*, et al. Nat Med* **27**, 1663-1665 (2021).
10. Moher, D.,*et al. PLoS Med* **7**, e1000217 (2010).