Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed

CONSORT-AI steering group; SPIRIT-AI Steering Group

DOI:
10.1038/s41591-019-0603-3

Reporting Guidelines for Clinical Trials Evaluating Artificial Intelligence Interventions are Needed

As artificial intelligence moves into the realm of clinical trials, consideration is needed on whether the current CONSORT and SPIRIT reporting statements are sufficient to ensure transparency.

The CONSORT-AI and SPIRIT-AI Steering Group

As of June 2019, more than 30 AI algorithms have been approved by the US Food and Drug Administration (including for the detection of diabetic retinopathy, stroke, brain haemorrhage, atrial fibrillation)[1] and over 300 clinical trials have been registered on C*linicalTrials.gov* under the headings 'artificial intelligence', 'machine learning' or 'deep learning'. These algorithms have the potential to transform healthcare, by offering earlier and more accurate diagnoses, lending novel insights into our understanding of diseases, enabling faster and more efficient service delivery and making medical care more available to those who really need it. Optimal reporting is key for evaluating the clinical utility of algorithms, for informing health policy and evidence-based recommendation, and to prevent research waste.[2] Most AI-interventions thus far, particularly diagnostic algorithms, have only been evaluated in the context of diagnostic accuracy. Whilst this initial validation stage is important, a demonstration of good diagnostic accuracy does not necessarily translate to improved patient outcomes. If the ultimate goal of introducing AI into healthcare is to bring about patient benefit, then demonstration of improved patient outcome is needed. This should be done in a prospective clinical trial, where the AI intervention is placed within its intended clinical pathway, with patient outcomes as the primary endpoint, and with demonstrable downstream effects in the broader management strategy.

The CONSORT (Consolidated Standards of Reporting Trials)[3] and SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials)[4] statements are minimum reporting guidelines for randomised trials and trial protocols. Their widespread endorsement has been instrumental in ensuring transparency and completeness for the evaluation of new interventions.

Although this guidance has substantially improved the completeness of clinical trials reporting, there are challenges in trials involving AI interventions that are not addressed by the current guidance. For example, elements which require detailed and specific reporting include the study setting and its ability to administer an ML intervention in real time, the criteria for inclusion at the input-data level as well as at the participant level, the interactions between the human and the algorithm and its potential knock-on effects downstream, and adaptive machine learning technologies (which have the potential to continuously improve in performance). Without complete and transparent reporting, readers cannot assess the validity and generalisability of the findings, which can result in widespread misconception of overstated efficacy and utility. The risk is that an AI-intervention, which might not be effective or feasible in the real world, could be commissioned and implemented.

To address these challenges, the CONSORT-AI and SPIRIT-AI Working Group is preparing international, consensus-based, AI-specific extensions to the CONSORT and SPIRIT statements that will specifically focus on clinical trials in which the intervention includes an ML or other AI component, using the Enhancing Quality and Transparency of Health Research's (EQUATOR) Network methodological framework for guideline development. This initiative will be complementary to the efforts of others working on reporting standards such as the TRIPOD-ML initiative of Collins and Moons, which seeks to improve reporting of ML-driven predictive models development and validation.[6] Our development process includes clinicians, computer scientists, researchers, trialists, industry leaders, regulators, funders, policy makers, journal editors and patient partners with an interest in applications of machine learning in healthcare. Through strong

stakeholder engagement and a modified Delphi consensus process, the AI extensions for CONSORT and SPIRIT guidelines will adequately reflect the concerns of the wider community.

Although many voices are important in this field, we wish to particularly draw attention to the role of medical journal editors. The endorsement of reporting standards by medical journals has been instrumental in ensuring that new interventions are reported transparently and completely to improve decision-making in healthcare. As seen with the original CONSORT and SPIRIT guidelines, the real impact depends on the extent to which journals adopt and require authors to comply with these standards. We wish to acknowledge the profound impact that the ICMJE (International Committee of Medical Journal Editors) community had on the quality of RCT reporting through their adoption of CONSORT. We strongly encourage the community of medical journal editors, particularly those with a focus on digital health, to take part in the Delphi process and support the endorsement of the final guidance in future publications.

A final consensus meeting and the publication of final recommendations will be completed in Spring 2020. Once the new guidance becomes available, we recommend that investigators planning to take an AI intervention through to clinical trials consult the new reporting standards as early as possible. Whilst the guidance will primarily be for ensuring reporting transparency, they can also assist in clinical trial design.

948 words

Competing Interests

Not applicable.

Acknowledgements

This initiative is funded by the Wellcome Trust, Health Data Research UK and the Alan Turing Institute.

Steering Group

Xiaoxuan Liu, Samantha Cruz Rivera, Livia Faes, Lavinia Ferrante di Ruffano, Christopher Yau, Pearse A Keane, Hutan Ashrafian, Sebastian Vollmer, Lucas M Bachmann, Christopher Holmes, Ara Darzi, Jonanthan J Deeks, An Wen Chan, David Moher, Melanie J Calvert, Alastair K Denniston

References

1. FDA Approvals For Smart Algorithms In Medicine In One Giant Infographic - The Medical Futurist. The Medical Futurist. https://medicalfuturist.com/fda-approvals-for-algorithms-in-medicine. Published June 6, 2019. Accessed September 1, 2019.

2. Chalmers I, Glasziou P, Heneghan C, Shepperd S, Loudon K, Bossuyt PM. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-89.

3. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 2011;9(8):672-677.

4. Chan A-W, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. *Lancet*. 2013;381(9861):91-92.

5. Reporting guidelines under development | The EQUATOR Network. http://www.equator-network.org/library/reporting-guidelines-under-development/. Accessed August 4, 2019.6. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *The Lancet*. 2019;393(10181):1577-1579. doi:10.1016/s0140-6736(19)30037-6