Commentary

Implications of the artificial intelligence extensions to the guidelines for consolidated standards of reporting trials and for standard protocol item recommendations for interventional trials (the CONSORT-AI and SPIRIT-AI extensions)

Ameenat Lola Solebo[a,b,c,*], Tasanee Braithwaite[d,e,f]

[a] UCL GOS Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK
[b] Great Ormond Street Hospital for Children NHS Trust, London, UK
[c] National Institute for Health Research Biomedical Research Centre, Great Ormond Street Hospital and UCL GOS ICH, London, UK
[d] The Medical Eye Unit, Guys and St Thomas' NHS Foundation Trust, UK
[e] Centre for Academic Rheumatology, King's College London, UK
[f] Cochrane Eyes and Vision Group, London, UK

Artificial intelligence (AI), the theory and development of computer systems able of performing tasks which normally need the application of human intelligence, holds great promise for improving health outcomes and experiences [1,2]. However, there is some anxiety around the safety and transparency (or 'black box') of AI systems which may impede their integration into healthcare and society more widely. The timely release of the CONSORT and SPIRIT extensions for interventions involving artificial intelligence is welcome, and promises to improve the quality of RCTs involving AI [3,4].

Synthesis of Randomised controlled trials (RCTs) through systematic review and meta-analysis, championed by the Cochrane Collaboration, permits critical overview of the literature, but this 'secondary research' is dependant on the quality of individual studies. In 1996, in recognition of the poor quality of many reported RCTs, a multinational expert working group published the Consolidated Standards of Reporting Trials (CONSORT) Statement, a set of evidence based recommendations for reporting randomized trials [5]. The CONSORT statement and accompanying Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) Statement have evolved and extended alongside changes in trial design, and have been endorsed by over 585 journals [5].

The recent initiative to develop AI-elaboration items for RCTs followed a rigorous consensus process, adopting the EQUATOR Network's methodological framework. Literature review was followed by expert consultation and 2-stage Delphi process involving 103 international stakeholders (journal editors, peer reviewers, pharmaceutical companies, regulatory bodies, academic institutions, funding agencies, clinicians). The CONSORT-AI and SPIRIT-AI elaborations add 14 items under 6 sections, and 15 items under 7 sections, respectively. Specifically, new items within the extensions include: clarity in the manuscript title and background rationale about the intended use of the AI application within a clinical pathway; the requirements for integration into the trial setting; the eligibility criteria both for participants and input data, including how poor quality or unavailable data were assessed and handled and whether there was human-AI interaction in handling the input data; specifying the intervention and protocol for its use and application to decision making or other areas of clinical practice precisely; detailing plans to identify and analyse performance errors; and reporting access to the AI intervention and/or its code. Importantly, commentary reported within the Delphi process revealed that the stakeholder panel appreciated the profound importance of the challenge presented by integration of AI into clinical trials, highlighting unpredictable errors "which are not easily detectable or explainable by human judgement." The stakeholder panel recognised the potential ease with which AI systems could be deployed at scale, and related concerns that, "unintended harmful consequences could be catastrophic." A particularly important addition resulting from this process was CONSORT-AI item 19, recommending analysis of systematic performance errors by the algorithm and their consequences.

Furthermore, a notable exclusion from this initiative was 'continuously evolving' AI systems which are currently in early development, with few tangible examples in healthcare applications. The panel identified the risks inherent in incremental software changes, which could impact safety performance, and highlighted a need for rigorous software version management and post-deployment surveillance.

* Corresponding author at: UCL GOS Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.
E-mail address: a.solebo@ucl.ac.uk (A.L. Solebo).

The CONSORT-AI and SPIRIT-AI elaborations represent a very important and timely advance towards enhancing the quality of study design and reporting for new AI interventions, and supporting the wider community in their transparent evaluation, including consideration of risk of bias. However, it is worth remembering that even with improved quality of RCT design and reporting [6], few of the 20,000 RCT papers published annually translate into clinical benefit for the wider target population [7,8]. In heralding the development of CONSORT-AI and SPIRIT-AI, we must not forget the multiple obstacles to the implementation of RCT findings. These challenges may be generic to RCTs, for example the selection of outcomes which are insufficiently patient centred or precise. They may also be particular to AI, for example the inadvertent propagation and magnification of health care disparities around gender, ethnicity and socioeconomic status [9,10], the heterogeneity of real world health care data maturity preventing widespread deployment of the AI-based intervention [9], or uncertainty over where moral accountability sits with regards to patient harm. Stakeholder involvement, and harmonisation of data estates across health care settings is central to the pathway to impact for AI based interventions. Nevertheless, harnessing the power of AI in order to develop interventions which are then rigorously assessed promises great benefit for patients and for population health.

Declaration of Competing Interest

The authors have no conflicts of interest to disclose.

Funding

References

[1] Lee MR, Sankar V, Hammer A, et al. Using machine learning to classify individuals with alcohol use disorder based on treatment seeking status. EClinicalMedicine 2019;12:70–8. doi: 10.1016/j.eclinm.2019.05.008.

[2] Liu QP, Xu X, Zhu FP, et al. Prediction of prognostic risk factors in hepatocellular carcinoma with transarterial chemoembolization using multi-modal multi-task deep learning. EClinicalMedicine 2020;23:100379. doi: 10.1016/j.eclinm.2020.100379.

[3] 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 Dig Health 2020 In print.

[4] Cruz Rivera S, Liu X, Chan AW, Denniston AK, and the SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Dig Health 2020. In print.

[5] CONSORT. Consolidated standards of reporting trials (CONSORT) statement Available from: http://www.consort-statement.org/, Accessed 18th August 2020

[6] Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surgery 2012;10(1):28–55. doi: 10.1016/j.ijsu.2011.10.001.

[7] National Institute of Health. NCBI PubMed. Available from: https://pubmed.ncbi.nlm.nih.gov/, Accessed 18th August 2020

[8] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials 2017;18(1):122. doi: 10.1186/s13063-017-1870-2.

[9] Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020;395(10236):1579–86. doi: 10.1016/s0140-6736(20)30226-9.

[10] Noseworthy PA, Attia ZI, Brewer LC, et al. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. Circul Arrhythmia Electrophysiolo 2020;13(3):e007988. doi: 10.1161/circep.119.007988.