Please see below for a link to the webinar recording for the Trials Methodology Research Partnership:

CONSORT-AI and SPIRIT-AI guidelines

Xiao Lui, University of Birmingham

11 January 2021

On behalf of Health Data Research UK

The slides are also available below.

For any queries, please contact uktmn@nottingham.ac.uk

https://www.youtube.com/watch?v=wTjd3KDpSfc

SPIRIT-AI CONSORT-AI

The SPIRIT-AI and CONSORT-AI initiative is an international collaborative effort to improve the **transparency** and **completeness** of reporting of **clinical trials** evaluating interventions involving artificial intelligence (AI)

Xiao Liu, Alastair Denniston

On behalf of The SPIRIT-AI & CONSORT-AI Working Group

equator network

Is there a problem with reporting in AI?

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu, Livia Faes*, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, Joseph R Ledsam, Martin K Schmid, Konstantinos Balaskas, Eric J Topol, Lucas M Bachmann, Pearse A Keane, Alastair K Denniston*

Summary

Background Deep learning offers considerable promise for medical diagnostics. We aimed to evaluate the diagnostic accuracy of deep learning algorithms versus health-care professionals in classifying diseases using medical imaging.

Methods In this systematic review and meta-analysis, we searched Ovid-MEDLINE, Embase, Science Citation Index, and Conference Proceedings Citation Index for studies published from Jan 1, 2012, to June 6, 2019. Studies comparing the diagnostic performance of deep learning models and health-care professionals based on medical imaging, for any disease, were included. We excluded studies that used medical waveform data graphics material or investigated the accuracy of image segmentation rather than disease classification. We extracted binary diagnostic accuracy data and constructed contingency tables to derive the outcomes of interest: sensitivity and specificity. Studies undertaking an out-of-sample external validation were included in a meta-analysis, using a unified hierarchical model. This study is registered with PROSPERO, CRD42018091176.

Lancet Digital Health 2019; 1: e271–97

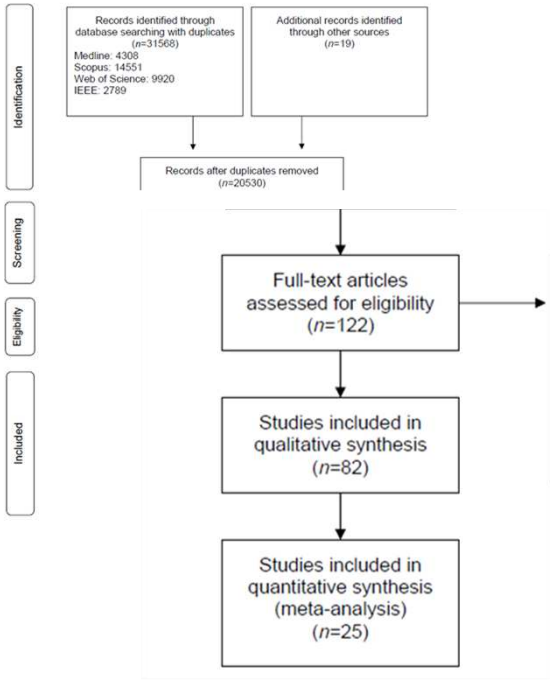
Published Online September 24, 2019 https://doi.org/10.1016/S2589-7500(19)30123-2

This online publication has been corrected. The corrected version first appeared at thelancet.com/digital-health on October 9, 2019

See Comment page e246

*Joint first authors

Department of



Is there a problem with reporting in AI?

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu*, Livia Faes*, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith Shamdas, Christoph Kern, Joseph R Ledsam, Martin K Schmid, Konstantinos Balaskas, Eric J Topol, Lucas M Bachmann, Pearse A Keane, Alastair K Denniston

Summary

Background Deep learning offers considerable promise for medical diagnostics. We aimed to evaluate the diagnostic accuracy of deep learning algorithms versus health-care imaging.

Methods In this systematic review and meta-analysis, we searc and Conference Proceedings Citation Index for studies publish the diagnostic performance of deep learning models and healt disease, were included. We excluded studies that used medica accuracy of image segmentation rather than disease classifica constructed contingency tables to derive the outcomes of inte out-of-sample external validation were included in a meta-anal registered with PROSPERO, CRD42018091176.

Lancet Digital Health 2019;

RESEARCH

OPEN ACCESS

Check for updates

Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

Myura Nagendran,[1] Yang Chen,[2] Christopher A Lovejoy,[3] Anthony C Gordon,[1,4] Matthieu Komorowski,[5] Hugh Harvey,[6] Eric J Topol,[7] John P A Ioannidis,[8] Gary S Collins,[9,10] Mahiben Maruthappu[3]

For numbered affiliations see end of the article.

Correspondence to: M Nagendran, Intensive Care, St Mary's Campus, Imperial College London, Praed Street, London W2 1NY, UK myura.nagendran@imperial.ac.uk (or @MyuraNagendran on Twitter; ORCID 0000-0002-4656-5096)

Additional material is published online only. To view please visit the journal online.

Cite this as: BMJ 2020;368:m689

ABSTRACT

OBJECTIVE

To systematically examine the design, reporting standards, risk of bias, and claims of studies comparing the performance of diagnostic deep learning algorithms for medical imaging with that of expert clinicians.

DESIGN

Systematic review.

DATA SOURCES

Medline, Embase, Cochrane Central Register of

REVIEW METHODS

Adherence to reporting standards was assessed by using CONSORT (consolidated standards of reporting trials) for randomised studies and TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) for non-randomised studies. Risk of bias was assessed by using the Cochrane risk of bias tool for randomised studies and PROBAST (prediction model risk of bias assessment tool) for non-randomised studies.

RESULTS

Inadequate Reporting

- Population characteristics for datasets
- Inclusion/exclusion criteria of participants
- Inclusion/exclusion criteria of images
- Methods for splitting the datasets
- Image preparation and pre-processing
- Procedures for poor quality images
- Provision of the full algorithm
- Instructions on how to use the algorithm
- Decisions made during algorithm training
- Expertise of the human comparator

Randomised Controlled Trials

Randomized Trials of AI Deep Neural Networks in Medicine

Procedure	Detection	Design	N Patients	N Sites	Place	Citation
Colonoscopy	Adenomas	Double-blind, sham control	1046	1	China	Wang P, Lancet Gastro Hep 2020
Colonoscopy	Adenomas	Unmasked	704	1	China	Gong D, Lancet Gastro Hep 2020
Colonoscopy	Adenomas	Unmasked	659	1	China	Su et al, Gastro Endoscopy 2020
Esophagogastro-duodenscopy	Blind spots	Unmasked	324	1	China	Wu L, Gut 2019
Colonoscopy	Adenomas	Unmasked	1058	1	China	Wang P, Gut 2019
Slit-lamp Photography	Childhood Cataracts	Unmasked	350	5	China	Lin H, E Clinical Medicine 2019

MENU

nature medicine

Comment | Published: 24 September 2019

Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed

The CONSORT-AI and SPIRIT-AI Steering Group

Nature Medicine **25**, 1467–1468(2019) | Cite this article

equator network

Enhancing the QUAlity and Transparency Of health Research

CONSORT

The CONSORT-AI Extension: Reporting Guidelines for Artificial Intelligence and Machine Learning Interventions in Randomised Trials (registered on 8th of May, 2019)

Steering Group: Professor Alastair Denniston, Professor Melanie Calvert, Dr Christopher Yau, Professor David Moher, Professor An-Wen Chan, Dr Pearse Keane, Professor Lucas Bachmann, Professor Chris Holmes, Dr Sebastian Vollmer, Dr Xiaoxuan Liu, Dr Livia Faes

Protocol Guidelines for Artificial Intelligence and Machine Learning Interventions in Randomised Trials (SPIRIT-AI Extension) (registered 21 June 2019)

Steering Group: Professor Alastair Denniston, Professor Melanie Calvert, Dr Christopher Yau, Professor David Moher, Professor An-Wen Chan, Dr Pearse Keane, Professor Lucas Bachmann, Professor Chris Holmes, Dr Sebastian Vollmer, Dr Xiaoxuan Liu, Dr Livia Faes

SPIRIT

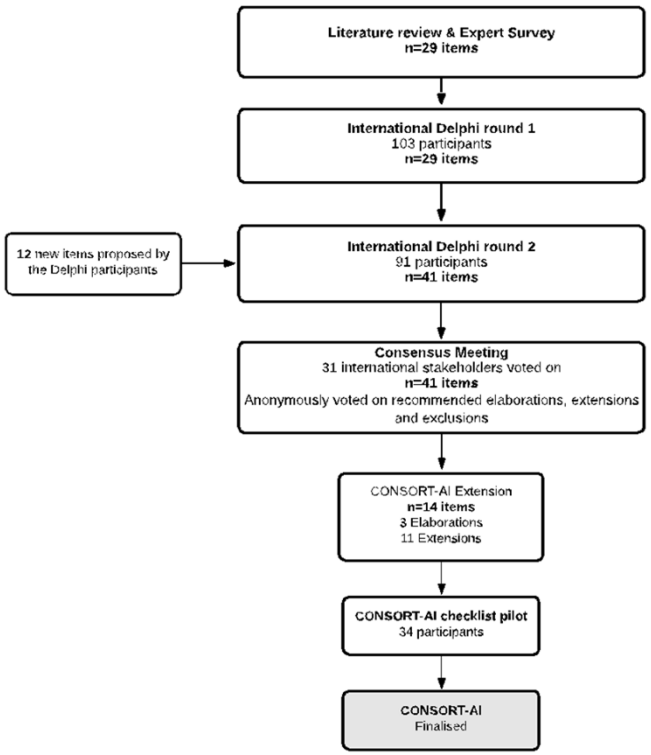
Standard Protocol Items: Recommendations for Interventional Trials

Developing SPIRIT-AI and CONSORT-AI

Review of existing guidance:

- ClinicalTrials.gov search for registered trials
  - 316 Studies found for: "machine learning" OR "deep learning" OR "artificial intelligence" on clinicaltrials.gov
  - **7 completed clinical trials with published results**
  - 1 with a published protocol
- Regulatory bodies and policy
  - **FDA:** "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback" – April 2019
  - **EMA:** none
  - **MHRA**: none
  - **NICE** - Evidence standards framework for digital health technologies
  - Academic literature
    - Kim et al 2019 design characteristics of reporting diagnostic analysis of medical images;
    - England and Cheng 2018, AI for medical image analysis: a guide for authors and reviewers;
    - Park et al 2018 Connecting Technological Innovation in Artificial Intelligence to Real-world Medical Practice through Rigorous Clinical Validation;
    - Park et al 2018 Principles for evaluating the clinical implementation of novel digital healthcare devices;
- Expert survey
- Liu & Faes *et al.* Lancet Digital Health, 2019.

Developing SPIRIT-AI and CONSORT-AI



- 103 international experts took part in the Delphi study
- 31 took part in the 2-day consensus meeting in Birmingham in January 2020.
- Healthcare professionals, methodologists, statisticians, computer scientists, industry representatives, journal editors, policy makers, health informaticists, experts in law and ethics, regulators, patients and funders.

CONSORT-AI

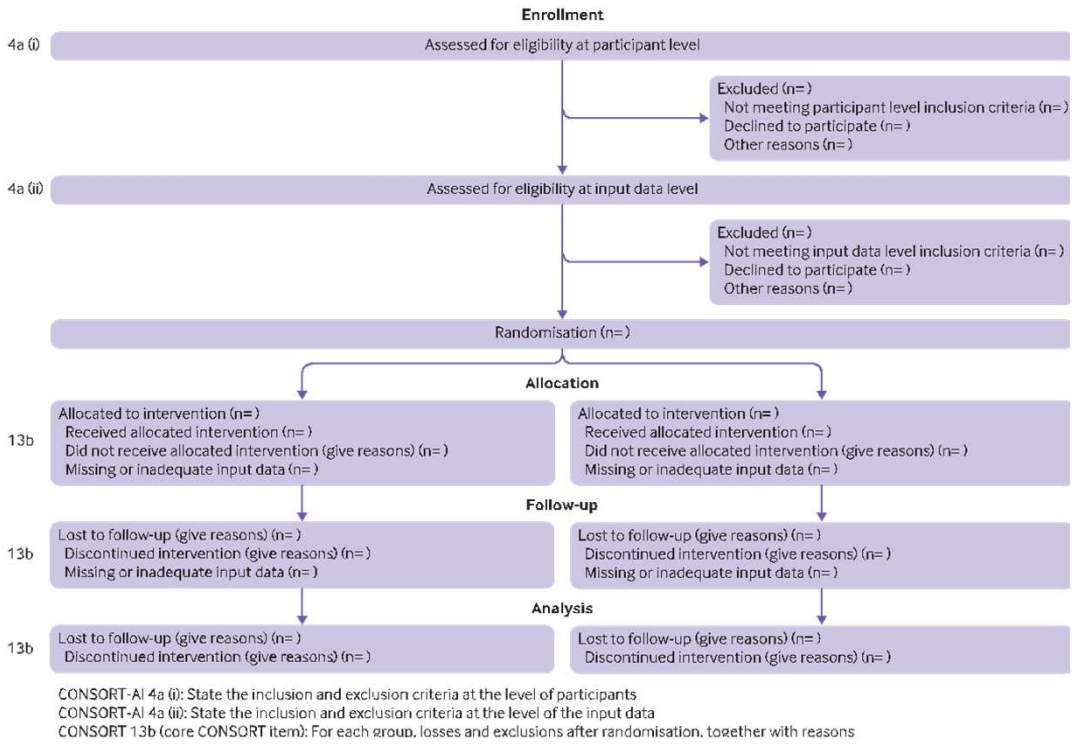
Title and abstract

CONSORT-AI 1a,b (i) Elaboration: Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model.

CONSORT-AI 1a,b (ii) Elaboration: State the intended use of the AI intervention within the trial in the title and/or abstract.

Introduction

CONSORT-AI 2a (i) Extension: Explain the intended use for the AI intervention in the context of the clinical pathway, including its purpose and its intended users (such as healthcare professionals, patients, public).



CONSORT-AI 4a (i): State the inclusion and exclusion criteria at the level of participants
CONSORT-AI 4a (ii): State the inclusion and exclusion criteria at the level of the input data
CONSORT 13b (core CONSORT item): For each group, losses and exclusions after randomisation, together with reasons

CONSORT-AI

Methods

CONSORT-AI 4a (i) Elaboration: State the inclusion and exclusion criteria at the level of participants.

CONSORT-AI 4a (ii) Extension: State the inclusion and exclusion criteria at the level of the input data.

Methods

CONSORT-AI 4b Extension: Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.

CONSORT-AI 5 (i) Extension: State which version of the AI algorithm was used.

CONSORT-AI 5 (ii) Extension: Describe how the input data were acquired and selected for the AI intervention.

CONSORT-AI 5 (iii) Extension: Describe how poor quality or unavailable input data were assessed and handled.

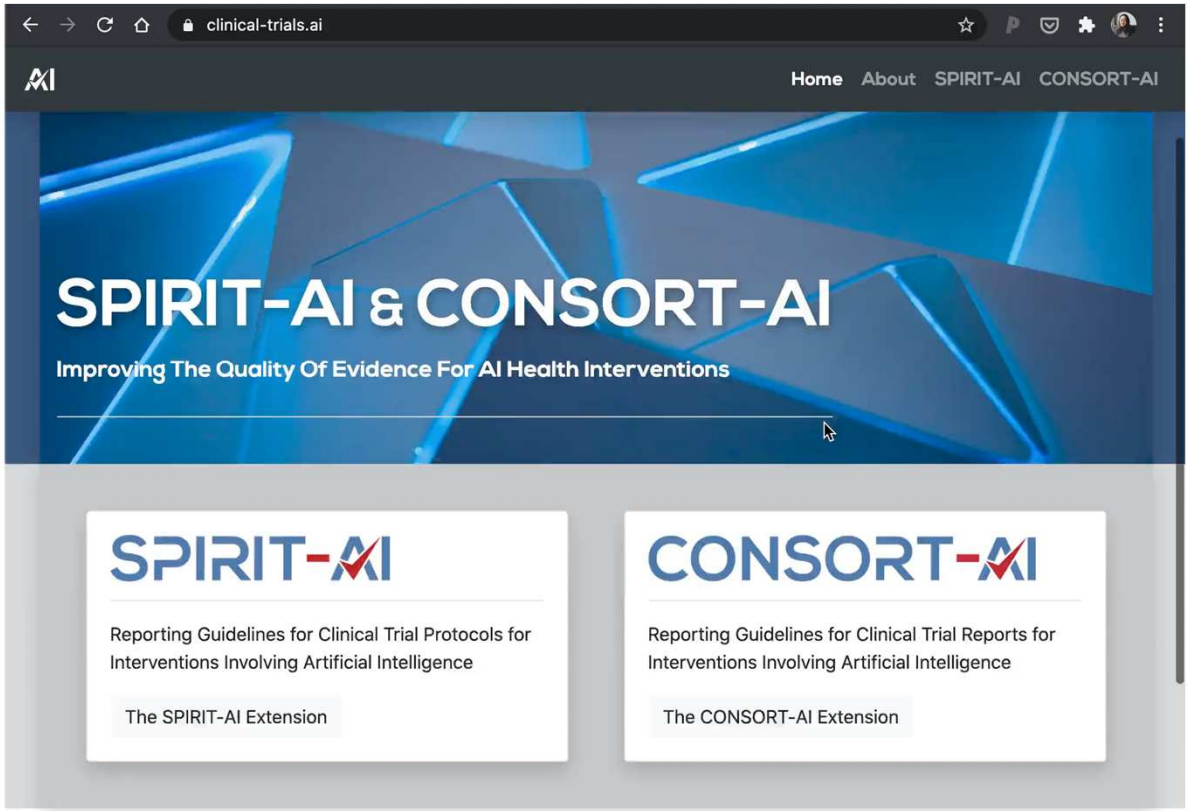
CONSORT-AI 5 (iv) Extension: Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.

Results

CONSORT-AI 19 Extension: Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, explain why not.

Other information

CONSORT-AI 25 Extension: State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.



The SPIRIT-AI and CONSORT-AI initiative is an international collaborative effort to improve the transparency and completeness of reporting of clinical trials evaluating interventions involving artificial intelligence (AI). SPIRIT-AI stands for Standard Protocol Items: Recommendations for Interventional Trials - Artificial Intelligence and CONSORT-AI stands for (Consolidated Standards of Reporting Trials - Artificial Intelligence).

The SPIRIT-AI and CONSORT-AI statements are extensions to the SPIRIT 2013 and CONSORT 2010 reporting guidelines for

nature medicine

THE LANCET
Digital Health

thebmj

September 2020

nature medicine

CONSENSUS STATEM

https://doi.org/10.1038/s41591-

OPEN

Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension

Samantha Cruz Rivera[1,2,3], Xiaoxuan Liu[1,4,5,6,7], An-Wen Chan[8], Alastair K. Denniston[1,3,4,5,6], Melanie J. Calvert[1,2,3,4,5,6,9,10,11], The SPIRIT-AI and CONSORT-AI Working Group*, SPIRIT-AI a
CONSORT-AI Steering Group and SPIRIT-AI and CONSORT-AI Consensus Group

The SPIRIT 2013 statement aims to improve the completeness of clinical trial protocol reporting by providing evid recommendations for the minimum set of items to be addressed. This guidance has been instrumental in promotin cent evaluation of new interventions. More recently, there has been a growing recognition that interventions inve cial intelligence (AI) need to undergo rigorous, prospective evaluation to demonstrate their impact on health out SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence) extensio reporting guideline for clinical trial protocols evaluating interventions with an AI component. It was developed in p its companion statement for trial reports: CONSORT-AI (Consolidated Standards of Reporting Trials-Artificial In Both guidelines were developed through a staged consensus process involving literature review and expert con generate 26 candidate items, which were consulted upon by an international multi-stakeholder group in a two-s survey (103 stakeholders), agreed upon in a consensus meeting (31 stakeholders) and refined through a checklist pi ticipants). The SPIRIT-AI extension includes 15 new items that were considered sufficiently important for clinical tri of AI interventions. These new items should be routinely reported in addition to the core SPIRIT 2013 items. SPIRIT mends that investigators provide clear descriptions of the AI intervention, including instructions and skills required setting in which the AI intervention will be integrated, considerations for the handling of input and output data, th interaction and analysis of error cases. SPIRIT-AI will help promote transparency and completeness for clinical tri for AI interventions. Its use will assist editors and peer reviewers, as well as the general readership, to understan and critically appraise the design and risk of bias for a planned clinical trial.

CONSENSUS STATEMENT

https://doi.org/10.1038/s41591-020-1034-x

nature medicin

OPEN

Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension

Xiaoxuan Liu[1,2,3,4,5], Samantha Cruz Rivera[2,6,7], David Moher[8,9], Melanie J. Calvert[1,2,3,4,5,6,10,11], Alastair K. Denniston[1,2,3,4,5,6,12] and The SPIRIT-AI and CONSORT-AI Working Group*

The CONSORT 2010 statement provides minimum guidelines for reporting randomized trials. Its widespread use has instrumental in ensuring transparency in the evaluation of new interventions. More recently, there has been a growing re tion that interventions involving artificial intelligence (AI) need to undergo rigorous, prospective evaluation to demon impact on health outcomes. The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) exte is a new reporting guideline for clinical trials evaluating interventions with an AI component. It was developed in parall its companion statement for clinical trial protocols: SPIRIT-AI (Standard Protocol Items: Recommendations for Interven Trials–Artificial Intelligence). Both guidelines were developed through a staged consensus process involving literature r and expert consultation to generate 29 candidate items, which were assessed by an international multi-stakeholder gr a two-stage Delphi survey (103 stakeholders), agreed upon in a two-day consensus meeting (31 stakeholders) and r through a checklist pilot (34 participants). The CONSORT-AI extension includes 14 new items that were considered ciently important for AI interventions that they should be routinely reported in addition to the core CONSORT 2010 CONSORT-AI recommends that investigators provide clear descriptions of the AI intervention, including instructions an required for use, the setting in which the AI intervention is integrated, the handling of inputs and outputs of the AI interve the human-AI interaction and provision of an analysis of error cases. CONSORT-AI will help promote transparency an pleteness in reporting clinical trials for AI interventions. It will assist editors and peer reviewers, as well as the general r ship, to understand, interpret and critically appraise the quality of clinical trial design and risk of bias in the reported outc

Randomized controlled trials (RCTs) are considered the gold-standard experimental design for providing evidence of the safety and efficacy of an intervention[1,2]. Trial results, if adequately reported, have the potential to inform regulatory deci-

AI is an area of enormous interest with strong drivers to create new interventions through to publication, implemen and market[3–5]. While AI systems have been researched for time, recent advances in deep learning and neural network

Review

RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension

Samantha Cruz Rivera,[1,2] Xiaoxuan Liu,[1,2,3,4,5,6] An-Wen Chan,[7] Alastair K Denniston,[1,2,3,4,5,8] Melanie J Calvert,[1,2,4,5,9,10,11] On behalf of the SPIRIT-AI and CONSORT-AI Working Group

Correspondence to:
A K Denniston,
Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK,
a.denniston@bham.ac.uk

Cite this as: *BMJ* 2020;370:m3210
http://dx.doi.org/10.1136/bmj.m3210

Accepted: 3 August 2020

The SPIRIT 2013 (The Standard Protocol Items: Recommendations for Interventional Trials) statement aims to improve the completeness of clinical trial protocol reporting, by providing evidence-based recommendations for the minimum set of items to be addressed. This guidance has been instrumental in promoting transparent evaluation of new interventions. More recently, there is a growing recognition that interventions involving artificial intelligence need to undergo rigorous, prospective evaluation to demonstrate their impact on health outcomes.

The SPIRIT-AI extension is a new reporting guideline for clinical trials evaluating interventions with an AI component. It was developed in parallel with its companion statement for clinical trial protocols. SPIRIT-AI extension includes 15 new items that were considered sufficiently important for clinical trials of AI interventions. These new items should be routinely reported in addition to the core SPIRIT 2013 items. SPIRIT-AI recommends that investigators provide clear descriptions of the AI intervention, including instructions and skills required for use, the setting in which the AI intervention will be integrated, considerations for the handling of input and output data, the human-AI interaction and analysis of error cases. SPIRIT-AI will help promote transparency and completeness for clinical trial protocols for AI interventions. Its use will assist editors and peer reviewers, as well as the general readership, to understand, interpret and critically appraise the design and risk of bias for a planned clinical trial.

Introduction

A clinical trial protocol is an essential document produced by study investigators detailing a priori the rationale, proposed methods and plans for how a clinical trial will be conducted.[1,2] This key document is used by external reviewers (funding agencies, regulatory bodies, research and treatment recommendation.[3] However, in most recent cases, the majority of published evidence has consisted of in-silico, early-phase validation. It has been recognised that most recent AI studies are inadequately reported and existing reporting guidelines do not fully cover potential sources of bias specific to AI systems.[4] The welcome

Review

RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension

Xiaoxuan Liu,[1,2,3,4,5] Samantha Cruz Rivera,[3,6] David Moher,[7,8] Melanie J Calvert,[4,5,6,9,10,11] Alastair K Denniston,[1,2,3,4,5,6,12] On behalf of the SPIRIT-AI and CONSORT-AI Working Group

Correspondence to:
A K Denniston, Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK,
a.denniston@bham.ac.uk

Cite this as: *BMJ* 2020;370:m3164
http://dx.doi.org/10.1136/bmj.m3164

Accepted: 4 August 2020

The CONSORT 2010 (Consolidated Standards of Reporting Trials) statement provides minimum guidelines for reporting randomised trials. Its widespread use has been instrumental in ensuring transparency when evaluating new interventions. More recently, there has been a growing recognition that interventions involving artificial intelligence (AI) need to undergo rigorous, prospective evaluation to demonstrate impact on health outcomes.

The CONSORT-AI extension is a new reporting guideline for clinical trials evaluating interventions with an AI component. It was developed in parallel with its companion statement for clinical trial protocols. CONSORT-AI extension includes 14 new items that were considered sufficiently important for AI interventions that they should be routinely reported in addition to the core CONSORT 2010 items. CONSORT-AI recommends that investigators provide clear descriptions of the AI intervention, including instructions and skills required for use, the setting in which the AI intervention is integrated, the handling of inputs and outputs of the AI intervention, the human-AI interaction and provision of an analysis of error cases. CONSORT-AI will help promote transparency and completeness in reporting clinical trials for AI interventions. It will assist editors and peer-reviewers, as well as the general readership, to understand, interpret and critically appraise the quality of clinical trial design and risk of bias in the reported outcomes.

Introduction

The two reporting guidelines for clinical trial protocols and reports were published in September 2020 in ***Nature Medicine, The Lancet Digital Health*** *and* ***The BMJ***.

SPIRIT-AI
CONSORT-AI

SPIRIT-AI CONSORT-AI equator network

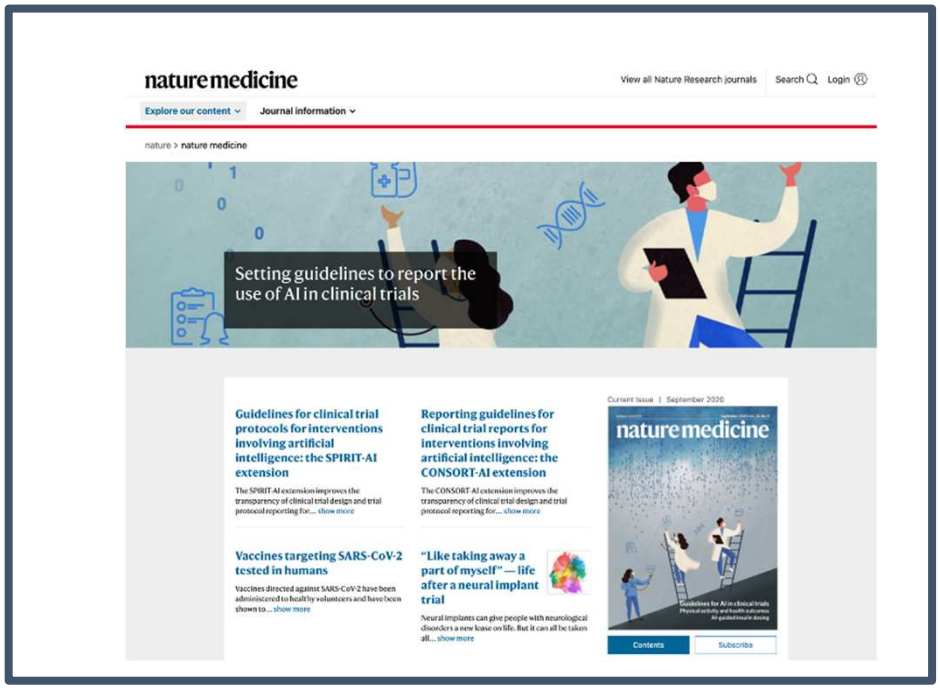
wellcome trust UNIVERSITY OF BIRMINGHAM Birmingham Health Partners HDRUK Health Data Research UK UKRI Research England The Alan Turing Institute

SPIRIT-AI & CONSORT-AI Steering Group:
Alastair K. Denniston, An-Wen Chan, Ara Darzi, Christopher Holmes, Christopher Yau, David Moher, Hutan Ashrafian, Jonathan Deeks, Lavinia Ferrante di Ruffano, Livia Faes, Melanie J. Calvert, Pearse A. Keane, Samantha Cruz Rivera, Sebastian J. Vollmer and Xiaoxuan Liu.

SPIRIT-AI & CONSORT-AI Consensus Group:
Aaron Y. Lee, Adrian Jonas, Andre Esteva, Andrew Beam, An-Wen Chan, Maria Beatrice Panico, Cecilia S. Lee, Charlotte Haug, Christopher Kelly, Christopher Yau, Cynthia Mulrow, Cyrus Espinoza, David Moher, Dina Paltoo, Elaine Manna, Gary Price, Gary S Collins, Hugh Harvey, James Matcham, Joao Monteiro, John Fletcher, M. Khair ElZarrad, Lavinia Ferrante Di Ruffano, Luke Oakden-Rayner, Melanie J. Calvert, Melissa McCradden, Pearse A. Keane, Richard Savage, Robert Golub, Rupa Sarkar and Samuel Rowley.

Impact - will it make a difference?

Endorsed by journals



Welcomed by regulatory experts

FDA

M. Khair ElZarrad - Deputy Director. Office of Medical Policy - CDER. U.S. FDA:

"*Developing a framework that helps facilitate and encourage transparency for the use of AI in clinical trials is important to advancing the field in general, and to establishing trust in AI-based tools and approaches.*"

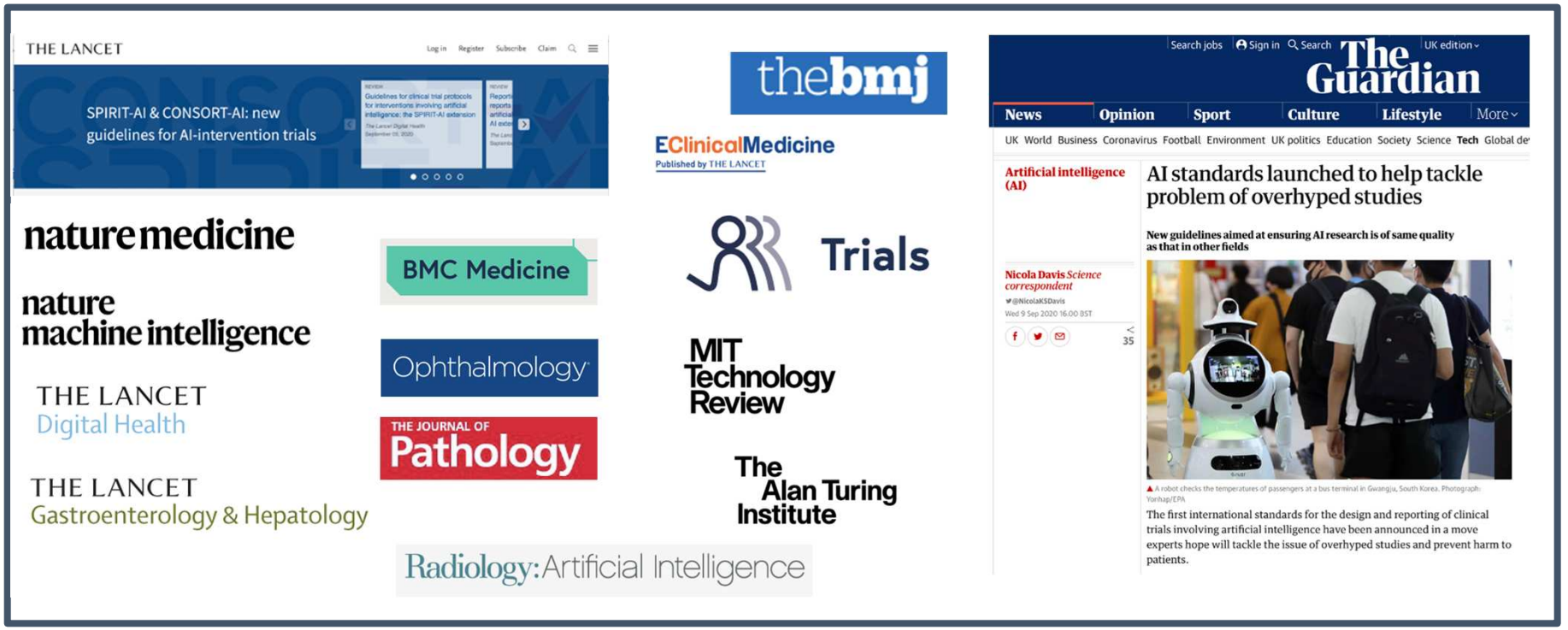
MHRA

Dr **Maria Beatrice Panico**, Medicines and Healthcare products Regulatory Agency (MHRA):

'The SPIRIT(AI) and CONSORT(AI) initiatives will contribute to the safe and scientifically sound development of artificial intelligence in the context of clinical trials'

Impact - will it make a difference?

Widespread coverage - an opportunity to explain why this matters



Impact - future work

Recognising that many studies in the field of AI are not RCTs

Editorial

Table 1. Summary of Guidelines for Artificial Intelligence Studies

Name of Artificial Intelligence Extension	Purpose of Artificial Intelligence System*	Study Design	Phase of Development or Testing of the Artificial Intelligence System	Status
Development and validation phase				
STARD AI[14]	Diagnosis	Diagnostic accuracy study	Testing the diagnostic accuracy of an AI system	In development
TRIPOD ML[15]	Diagnosis or prognosis	Studies developing, validating, or updating a prediction model	Development, validation, or updating of an AI system, or a combination thereof	In development
Testing and regulatory phase				
CONSORT AI[6]	Any health intervention	Randomized trial (report)	Randomized trial report, results for the effectiveness of an AI system	Published online September 9, 2020, in the *British Medical Journal*, *Lancet Digital Health*, and *Nature Medicine*
SPIRIT AI[7]	Any health intervention	Randomized trial (protocol)	Randomized trial protocol for testing the effectiveness of an AI system	Published online September 9, 2020, in the *British Medical Journal*, *Lancet Digital Health*, and *Nature Medicine*

Campbell et al Reporting guidelines for Artificial Intelligence in Medical Research. https://doi.org/10.1016/j.ophtha.2020.09.009