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 uktnm@nottingham.ac.uk

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


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.
Is there a problem with reporting in AI?

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

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

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 Pearlse Keane, Professor Lucas Bachmann, Professor Chris Holmes, Dr Sebastian Vollmer, Dr Xiaosuan Liu, Dr Liiva Fæs

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 Pearlse Keane, Professor Lucas Bachmann, Professor Chris Holmes, Dr Sebastian Vollmer, Dr Xiaosuan Liu, Dr Liiva Fæs
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**

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

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

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 Steering Group:

SPIRIT-AI & CONSORT-AI Consensus Group:
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

![Table 1. Summary of Guidelines for Artificial Intelligence Studies](https://doi.org/10.1016/j.ophtha.2020.09.009)