

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Quality of reporting of randomised controlled trials in artificial intelligence in health care: a systematic review
AUTHORS	Shahzad, Rida; Ayub, Bushra; Siddiqui, M. A. Rehman

VERSION 1 – REVIEW

REVIEWER	Xiaoxuan Liu University of Birmingham
REVIEW RETURNED	15-Apr-2022

GENERAL COMMENTS	The search for papers included in this review is not sufficient for identifying RCTs of AI interventions and is more than 1 year out of date. I do not think it provides sufficiently novel or relevant results to warrant publication in its current form.
-------------------------	---

REVIEWER	Anthony Yeo Western Sydney University
REVIEW RETURNED	21-Apr-2022

GENERAL COMMENTS	The authors have systematically used the criteria present in CONSORT-AI to analyse their studies. They went through the process and then arrived at conclusions. While everything was done correctly, this paper is not very ground breaking in that it represents a survey of the landscape. Thus, I would not recommend this paper as a major research paper. However, it could be published as a minor paper. To this end, they could condense the paper to a shorter length, i.e. make it more concise to convey the same messages.
-------------------------	---

REVIEWER	Susan C Shelmerdine Great Ormond Street Hospital for Children
REVIEW RETURNED	10-May-2022

GENERAL COMMENTS	<p>Thank you for asking me to review this article which aims to address whether or not randomised controlled trials (RCTs) focussing on use of AI in healthcare follow the recently published CONSORT-AI reporting guidelines.</p> <p>Overall Strengths: I think this is an important topic to address and to highlight for future studies, and raise awareness of these reporting guidelines.</p> <p>Overall Limitations: There are a few areas in the text for improvement. I have listed specific comments below for the authors to address.</p> <p>Article strengths and limitations section:</p>
-------------------------	---

- Pg 5/48 Line 8: Please do not state 'first study to evaluate' – we do not know if this will be true by time the study is published or if other preprint articles are already existing on this topic.
- Pg 5/48 Line 13: Please reword this second point in third person (i.e. 'This study...'). I don't know what point you are trying to make here – how does this review help assess applicability? I think you should state that it highlights limitations of currently existing studies?
- Pg 5/48 Line 15: I don't think this is the major limitation – I think the fact there are so many newer articles constantly happening meaning such a study could be easily outdated would be a major limitation.

Introduction:

- In general I think this introduction needs to make a stronger point as to why this study is necessary. It's not so much about disappointment of people (as referred to in line 42) but more inappropriate evidence being published and used to create decisions that impact healthcare negatively. This needs to be emphasized more.

- I think the authors should reference other articles in the literature that have already shown that AI studies are not following guidelines in other aspects, like non adherence to CLAIM checklist for diagnostic accuracy studies and non adherence for external validation in other cases. There are many references they can use to show that other AI studies don't follow guidelines and they want to prove this is also the case for RCTs.

- Page 5/38 – Line 27: Please start with a definition of AI rather than a nebulous statement about potential.

- Page 5/48 Line 31: 'with emphasis on deep learning and neural networks' – I believe deep learning is synonymous with neural networks so don't understand why both have to be mentioned?

- Page 5/48 Line 47: 'RCTs are the highest quality of evidence' – I don't agree – I think meta-analyses and systematic reviews are higher quality of evidence? Please remove or amend accordingly.

- Page 6/48 Line 19: 'RCRs of AI have unique characteristics' – maybe you should list what is so unique about them and what kind of extensions were needed that could not be covered with the usual CONSORT statement that need to be accounted for. This will help educate readers for what is the complexity of these studies that makes them important to address and assess.

Methods:

- Page 6/48 Line 45: Why was the particular timeframe chosen? Why 5 years not 10 years or 20 years? Why ending in December 2020 rather Dec 2021 when we are now in 2022?

- Page 6/48 Line 48: It feels like the search terms are too basic and the full search terminology with number of returns has not been provided in the text. The authors should read other systematic reviews published in high impact journals to get a feel for how search terminologies are performed and presented in a publication.

e.g. Why only search Artificial intelligence, deep learning and machine learning?

Why not also include neural networks, natural language processing,

	<p>support vector, computational intelligence etc? There are many other MeSH terms that can also yield results relating to these types of studies. Why no mention of healthcare if this search was for healthcare/medical related topics? Why no mention to search RCTs in the criteria? This must have yielding a large number of irrelevant studies to look through?</p> <ul style="list-style-type: none"> - Page 6 Line 59: Who were the two investigators who did the search and what is their experience with conducting systematic reviews? - Page 7 Line 38: Who are the two 'trained graders' for the checking of the checklist. What training did they receive to complete this review? - Page 7 Line 52: If the study was registered in PROSPERO please state the ID registration number. <p>Results:</p> <ul style="list-style-type: none"> - I am unable to open Figure 1 so cannot provide comment or review. - I think it is unreasonable to expect readers to know what the different checklist items refer to in the CONSORT AI by memory and it is tedious to refer to supplementary tables. Please state what these items refer to in the text or provide some kind of reference tool in a figure within the main text. - In the text it would be better to put subheadings refer to them as items within the Intro/Methods/Results/Discussion and then elaborate. - In the table please provide % as well as the numbers. Please also list what the items refer to specifically. - You list what the most frequently reported items are in the text on Page 11, Line 18 – but what about the ones that are least reported? Most likely to be missing. In my opinion this is the more important finding to report. <p>Discussion:</p> <ul style="list-style-type: none"> - The discussion is very poorly written and needs revision. It is mostly just repetition of the results, and should belong in the results section than the discussion. - Line 45, Page 12: No need to capitalise 'Review' - The first paragraph should summarise the key findings of the study before going into the specifics. Please amend. - It is odd the authors chose to include % in the discussion but never bothered to make this clear in the results section. Please amend results section, or better still move the discussion to the results and start the discussion again as it is not really a discussion at the moment. - This discussion needs to make reference to the clinical impact and effect of the results from the review. What is the consequence of the most commonly missed items, is it that important or not? Etc. etc. It makes no reference to the prior studies in the literature about non-
--	---

	adherence to reporting standards for other studies and other AI type studies (non-RCTs). It really doesn't place the work in the context of the wider community. I don't think it fully captures items 23b, c, and d in the PRISMA statement.
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comment	Author response	Page and line number of changes
The search for papers included in this review is not sufficient for identifying RCTs of AI interventions and is more than 1 year out of date. I do not think it provides sufficiently novel or relevant results to warrant publication in its current form.	We have updated the study to include 2021	

Reviewer 2 comment	Author response	Page and line number of changes
The authors have systematically used the criteria present in CONSORT-AI to analyse their studies. They went through the process and then arrived at conclusions. While everything was done correctly, this paper is not very ground breaking in that it represents a survey of the landscape. Thus, I would not recommend this paper as a major research paper. However, it could be published as a minor paper. To this end, they could condense the paper to a shorter length, i.e. make it more concise to convey the same messages. [NOTE FROM THE EDITORS: Please note that we do not have a brief report format, so with respect to formatting an length restrictions, some of this reviewer's comments may be rebutted]	Thank you for your comment. We have disregarded it as per journal policy.	

Reviewer 3 comment	Author response	Page and line number of changes
Article strengths and limitations section:		
Pg 5/48 Line 8: Please do not state 'first study to evaluate' – we do not know if this will be true by time the study is published or if other preprint articles are already existing on this topic.	Amended	Pg 4

Pg 5/48 Line 13: Please reword this second point in third person (i.e. 'This study....'). I don't know what point you are trying to make here – how does this review help assess applicability? I think you should state that it highlights limitations of currently existing studies?	Amended	Pg 4
Pg 5/48 Line 15: I don't think this is the major limitation – I think the fact there are so many newer articles constantly happening meaning such a study could be easily outdated would be a major limitation.	We have updated the study.	
Introduction:		
In general I think this introduction needs to make a stronger point as to why this study is necessary. It's not so much about disappointment of people (as referred to in line 42) but more inappropriate evidence being published and used to create decisions that impact healthcare negatively. This needs to be emphasized more.	Amended	Pg 4
I think the authors should reference other articles in the literature that have already shown that AI studies are not following guidelines in other aspects, like non adherence to CLAIM checklist for diagnostic accuracy studies and non adherence for external validation in other cases. There are many references they can use to show that other AI studies don't follow guidelines and they want to prove this is also the case for RCTs.	Added in discussion	Pg 11
Page 5/38 – Line 27: Please start with a definition of AI rather than a nebulous statement about potential.	Amended	Pg 4
Page 5/48 Line 31: 'with emphasis on deep learning and neural networks' – I believe deep learning is synonymous with neural networks so don't understand why both have to be mentioned?	Amended	Pg 4
Page 5/48 Line 47: 'RCTs are the highest quality of evidence' – I don't agree – I think meta-analyses and systematic reviews are higher quality of evidence? Please remove or amend accordingly.	Amended	Pg 5
Page 6/48 Line 19: 'RCTs of AI have unique characteristics' – maybe you should list what is so unique about them and what kind of extensions were needed that could not be covered with the usual CONSORT statement that need to be accounted for. This will help educate readers for what is the complexity of these studies that makes them important to address and assess.	Amended	Pg 5

Methods:		
Page 6/48 Line 45: Why was the particular timeframe chosen? Why 5 years not 10 years or 20 years? Why ending in December 2020 rather Dec 2021 when we are now in 2022?	We have updated the study	
<p>Page 6/48 Line 48: It feels like the search terms are too basic and the full search terminology with number of returns has not been provided in the text. The authors should read other systematic reviews published in high impact journals to get a feel for how search terminologies are performed and presented in a publication.</p> <p>e.g. Why only search Artificial intelligence, deep learning and machine learning?</p> <p>Why not also include neural networks, natural language processing, support vector, computational intelligence etc? There are many other MeSH terms that can also yield results relating to these types of studies. Why no mention of healthcare if this search was for healthcare/medical related topics? Why no mention to search RCTs in the criteria? This must have yielding a large number of irrelevant studies to look through?</p>	Limited search terms were used so that the search results would not be restricted and a maximum number of studies could be included in the final review.	
Page 6 Line 59: Who were the two investigators who did the search and what is their experience with conducting systematic reviews?	Both are experienced and have previously published systematic reviews	
Page 7 Line 38: Who are the two 'trained graders' for the checking of the checklist. What training did they receive to complete this review?	Cochrane systematic review workshop completed by both investigators	
Page 7 Line 52: If the study was registered in PROSPERO please state the ID registration number.	Turned down by PROSPERO due to COVID workload	
Results:		
I am unable to open Figure 1 so cannot provide comment or review.	No other reviewer had this problem	
I think it is unreasonable to expect readers to know what the different checklist items refer to in the CONSORT AI by memory and it is tedious to refer to supplementary tables. Please state what these items refer to in the text or provide some kind of reference tool in a figure within the main text.	As per journal policy, table had to be included as a supplementary table	

In the text it would be better to put subheadings refer to them as items within the Intro/Methods/Results/Discussion and then elaborate.	This would result in too many subheadings in the text	
In the table please provide % as well as the numbers. Please also list what the items refer to specifically.	Numbers are given in the table. Items are specifically given in supplementary table 1	Table 2
You list what the most frequently reported items are in the text on Page 11, Line 18 – but what about the ones that are least reported? Most likely to be missing. In my opinion this is the more important finding to report.	Reported.	Pg 8
Discussion:		
The discussion is very poorly written and needs revision. It is mostly just repetition of the results, and should belong in the results section than the discussion.	The discussion has been revised	Page 5
Line 45, Page 12: No need to capitalise 'Review'	Amended	Pg 11
The first paragraph should summarise the key findings of the study before going into the specifics. Please amend.	Amended	Pg 11
It is odd the authors chose to include % in the discussion but never bothered to make this clear in the results section. Please amend results section, or better still move the discussion to the results and start the discussion again as it is not really a discussion at the moment.	Absolute numbers added in discussion	Pg 11 onwards
This discussion needs to make reference to the clinical impact and effect of the results from the review. What is the consequence of the most commonly missed items, is it that important or not? Etc. etc. It makes no reference to the prior studies in the literature about non-adherence to reporting standards for other studies and other AI type studies (non-RCTs). It really doesn't place the work in the context of the wider community. I don't think it fully captures items 23b, c, and d in the PRISMA statement.	Discussion has been made on commonly missed items and their importance Reference to prior studies showing nonadherence has been added.	Pg 11 onwards