Reviewer’s report

Title: Key challenges for delivering clinical impact with artificial intelligence

Version: 0 Date: 17 Jul 2019

Reviewer: Seong Ho Park

Reviewer's report:

This manuscript is timely and addresses an important issue regarding AI in healthcare. Although OP/ED articles that overlap with the contents of this manuscript have appeared in some other journal, even including premier journals such as Lancet, JAMA, and Science, this manuscript is more comprehensive. Authors successfully listed virtually all relevant points and explained them with proper citation. The content organization is a bit different from what it would have been from a typical health technology assessment perspective. However, the current organization is also clear enough to follow. Nevertheless, some minor revisions considering the following comments and additional citations would be beneficial.

1. Since the recent boom of deep learning, AI and deep learning are often used almost synonymously. Therefore, it appears that the example studies quoted are deep learning related. Regarding RCT on AI, there is a large RCT named INFANT trial, "The INFANT Collaborative Group. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. Lancet. 2017; 389: 1719-1729." It is not deep learning technology but belongs to the category of AI. To my knowledge, this is the largest RCT of this kind. It is an excellent example of RCT on AI and also serves as an example that high accuracy may not guarantee a better patient outcome.

2. Regarding the quality reporting of AI studies, the mention of TRIPOD is appropriate. However, TRIPOD was made principally considering traditional statistical regression analysis in mind. TRIPOD statement specific to machine learning (TRIPOD-ML) is currently under development (Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019 Apr 20;393:1577-1579). Authors might want to address this briefly to make readers pay attention. 3. Indeed, AUC may not necessarily represent clinical efficacy appropriately. Also, AUC itself does not work by itself for clinical decision making. It has to be accompanied by a threshold that turns the continuous AI output into decision categories. The same factors that limit the generalizability of AI performance would also prohibit the generalization of the threshold. For example, in Hwang et al. JAMA Netw Open. 2019;2:e191095 (authors' reference #5), the same high-sensitivity cutoff yielded a wide range of specificity results (0.566 to 1) across different institutions. 4. The importance of external validation using a representative sample of the intended deployment population in clinical practice is addressed in several locations in the manuscript. However, it would need to be mentioned more directly for emphasis as the lack of proper external validation of AI is currently a critical concern. According to a recent study (Kim et al. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. Korean J Radiol. 2019 Mar;20:405-410. https://doi.org/10.3348/kjr.2019.0025), which analyzed published studies investigating the performance of AI algorithms that analyze medical images to provide diagnostic decisions, only 6% performed external validation. Furthermore, none performed external validation using a prospective dataset representative of the intended clinical cohort. 5. I recommend deleting the mention of economic
incentives including 1) lines 52-53 on page 6, 2) the entire subsection titled "Misaligned incentives for adoption at scale" on page 7, and 3) "economic incentives to promote value-based care" in the last sentence of the conclusion paragraph. Unlike other issues, this issue is dependent upon the social and health system of each country. So, it is not generalizable knowledge. Furthermore, value-based healthcare and cost-effectiveness of a health technology need to take a lot more factors into account than those briefly mentioned in the manuscript. As the authors address in the manuscript, we have yet to investigate the real-world accuracy of AI and health impact of AI. It is yet premature to discuss the cost-effectiveness of AI. The conclusion paragraph should mention the importance of prospective studies and the importance of clinical trials that directly assess the clinical impact of AI beyond technical accuracy, i.e., how AI affects the quality of care, variability between healthcare professionals, efficiency/productivity of clinical practice, and most importantly patient outcomes. These are mentioned in the main body but is not clearly stated in the conclusion paragraph.

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

None exists.

**Statement on potential review bias**
Please complete a statement on potential review bias, considering the following questions:
1. Did you co-author any publication with an author of this manuscript in the last 5 years?
2. Are you currently or recently affiliated at the same institution as an author of this manuscript?
   If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria'. If your reply is yes to any, please give details below.

None exists.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal