Reviewer’s report

Title: Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study

Version: 0 Date: 31 Jan 2019

Reviewer: Licong Cui

Reviewer's report:

This paper presents a user study to evaluate a machine learning (ML) tool (called RobotReviewer) that can facilitate the review and assessment of bias risks in randomized controlled trials (RCTs). This study mainly focuses on the evaluation of the assessment time taken by anonymous participants and the usability of the tool. A total of 41 participants were recruited in this study. Each participant was asked to review four RCT articles, among which two were provided with ML suggestions (i.e., semi-automated) and the other two were without ML suggestions (i.e., manual). Qualitative feedback was also collected for further improvement of the tool. The within-subject design takes into account of various factors that may affect the evaluation including order effect, user fatigue and learning effects. It was concluded that semi-automated assessment was 25% faster than manual assessment, indicating that ML-based semi-automation can improve the efficacy of RCT bias risk assessment. Unlike other ML work that perform retrospective evaluation on existing benchmark datasets, this study adopts a prospective way for the evaluation, which is a challenging task.

The paper is well-written and easy to follow. The reviewer has several comments to further improve the paper.

* The rationales for using two categories of ratings "low" and "high/unclear" seem not convincing. Several participants also suggested separation of "high/unclear" into two categories. More justifications could be provided or the limitation could be discussed.

* Table 2, the last row regarding "changed (no ML)". It is unclear to the reviewer that why there are changed annotations for "no ML". In particular, it is described in the evaluation design subsection that "In the manual condition, participants completed the task without machine assistance, starting from an unannotated PDF shown in the same interface." The reviewer is assuming that "changed (no ML)" refers to the manual assessment in which case only unannotated PDF was provided. More clarification would be appreciated.

* For the result of participants' judgements, although Table 2 presents the basic statistics for participants' assessment results, it is unclear about the degree of agreement among raters (i.e., inter-rater agreement). For each of the RCT articles, since both semi-automated and manual evaluation were involved, is there any significant difference between the inter-rater agreements for the article?
The reviewer understand that the main focus of the evaluation is the assessment time taken and usability of the tool. However, it would be interesting to know the effectiveness of the ML tool based on the human experts' judgements. In another word, if the ML-suggested annotations and manual annotations agreed by the human experts were combined to serve as the reference standard, the precision and recall could be calculated and reported.

The followings are a couple of typos:
- page 10, line 9: "questions which affords" => "questions which afford"
- page 22, line 15: "participate that those" => "participate than those"

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

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.

I declare that I have no competing interests

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